Sinopfbd, 2026; 11(1): 321-341 https://doi.org/10.33484/sinopfbd.1814140

Sinop Üniversitesi Fen Bilimleri Dergisi
Sinop Uni J Nat Sci
E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

Optimizing Speech-Based Parkinson's Diagnosis with Hybrid Feature Selection and Machine Learning

Setayesh SAEMİ[1] and Fatma Özge ÖZKÖK[2]

How to cite: Saemi, S. & Özkök, F. Ö. (2026). Optimizing speech-based Parkinson's diagnosis with hybrid feature selection and machine learning. *Sinop Üniversitesi Fen Bilimleri Dergisi*, *11*(1), 321-341. https://doi.org/10.33484/sinopfbd.1814140

Open Access

Research Article

Corresponding Author
Fatma Özge ÖZKÖK
fozgeozkok@erciyes.edu.tr

ORCID of the Authors
S.S: 0009-0003-9125-9738
F.Ö.Ö: 0000-0002-1421-4670

Received: 31.10.2025
Accepted: 17.04.2026

Abstract

Parkinson's disease (PD) is a common neurological disorder that causes cognitive and motor impairments, including speech and gait abnormalities. Medication is often used to treat symptoms, but a cure is still elusive. Effective disease management depends on early detection. This study employs ML approaches, including K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), and Naive Bayes (NB), to distinguish between individuals with PD and healthy controls based on voice signal characteristics. To enhance classification performance, multiple feature selection techniques were applied, including filter-based methods (correlation analysis, Analysis of Variance [ANOVA]), wrapper-based approaches such as Recursive Feature Elimination (RFE), embedded models including Random Forest (RF) and XGBoost importance, and a Genetic Algorithm (GA). Furthermore, a novel hybrid method was proposed by combining GA and XGBoost feature importance to identify the most relevant features. The models were trained and tested using standard preprocessing techniques such as feature scaling and RandomOverSampler to address class imbalance. Experimental results demonstrate that the hybrid feature selection method significantly improved classification accuracy. Using rigorous 10-Fold Stratified Cross-Validation, the proposed method achieved a robust mean accuracy of 92.74% (XGBoost) and a geometric mean of 91.13% (KNN). These findings suggest that integrating evolutionary and model-driven feature selection can significantly enhance the diagnosis of PD, providing a promising approach for voice-based medical decision support systems.

Keywords: Parkinson's disease, speech signal analysis, hybrid feature selection, machine learning, voice-based diagnosis

Hibrit Özellik Seçimi ve Makine Öğrenimi ile Konuşmaya Dayalı Parkinson Tanısının Optimize Edilmesi

Özet

Parkinson hastalığı (PD), konuşma ve yürüme bozuklukları da dahil olmak üzere bilişsel ve motor işlevlerde bozulmalara neden olan yaygın bir nörolojik bozukluktur. Semptomları tedavi etmek için sıklıkla ilaç kullanılır, ancak bir tedavi henüz bulunamamıştır. Etkili hastalık yönetimi, erken teşhise bağlıdır. Bu çalışma, ses sinyali özelliklerine dayanarak Parkinson hastalığı olan bireyleri sağlıklı kontrollerden ayırt etmek için K-En Yakın Komşular (KNN), Çok Katmanlı Algılayıcı (MLP), Aşırı Gradyan Artırma (XGBoost) ve Naive Bayes (NB) dahil olmak üzere makine öğrenimi yaklaşımlarını kullanmaktadır. Sınıflandırma performansını artırmak için, filtre tabanlı yöntemler (korelasyon analizi, Varyans Analizi [ANOVA]), Özyinelemeli Özellik Eleme (RFE) gibi sarmalayıcı tabanlı yaklaşımlar, Rastgele Orman (RF) ve XGBoost önemini içeren gömülü modeller ve Genetik Algoritma (GA) dahil olmak üzere çoklu özellik seçimi teknikleri uygulandı. Ayrıca, en alakalı özellikleri belirlemek için GA ve XGBoost özellik önemini birleştirerek yeni bir hibrit yöntem önerildi. Modeller, sınıf dengesizliğini gidermek için özellik ölçeklendirme ve RandomOverSampler gibi standart ön işleme teknikleri kullanılarak eğitildi ve test edildi. Deneysel sonuçlar, hibrit özellik seçimi yönteminin sınıflandırma doğruluğunu önemli ölçüde iyileştirdiğini göstermektedir; titiz 10 Katlı Katmanlı Çapraz Doğrulama kullanılarak, önerilen yöntem %92.74'lük (XGBoost) sağlam bir Ortalama Doğruluk ve %91.13'lük (KNN) Geometrik Ortalama elde etmiştir. Bu bulgular, evrimsel ve model tabanlı özellik seçiminin entegrasyonunun Parkinson hastalığının teşhisini önemli ölçüde geliştirebileceğini ve ses tabanlı tıbbi karar destek sistemleri için umut vadeden bir yaklaşım sağlayabileceğini göstermektedir.

Anahtar Kelimeler: Parkinson hastalığı, konuşma sinyali analizi, hibrit özellik seçimi, makine öğrenimi, ses tabanlı teşhis

[1] Erciyes University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering , Kayseri, Türkiye

[2] Erciyes University, Engineering Faculty, Department of Computer Engineering, Kayseri, Türkiye

Introduction

Parkinson's disease (PD) is a brain disease that mainly affects movement abilities and results in manifestations such as shaking hands or limb stiffness and slow movement. The early and accurate diagnosis of PD is essential for its effective treatment and management [1, 2]. The traditional diagnosis of PD generally relies on clinical evaluations, which are often subjective and variable. In this regard, Machine Learning (ML) techniques have become increasingly popular to enhance accuracy and efficiency in diagnosis. Between 1996 and 2016, the number of PD cases increased dramatically from 2.5 million to 6.1 million cases worldwide. This fourfold increase is closely associated with population aging and increased life expectancy [3]. In this study, several ML algorithms, including K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Naive Bayes (NB), and Extreme Gradient Boosting (XGBoost), were applied to diagnose PD using vocal biomarkers. The performance of these models was further optimized through preprocessing techniques (such as normalization, standard scaler, and RandomOverSampler) and different filter- and wrapper-based methods for feature selection, including Genetic Algorithm (GA), XGBoost-based importance ranking, Random Forest (RF) technique, analysis of variance (ANOVA), and Recursive Feature Elimination (RFE), and their results were then compared. Additionally, to assess the impact of feature dimensionality, we experimented with multiple sets of selected features, including subsets comprising 5, 12, 13, and 14 features.

The dataset used in this study was obtained by extracting features from voice recordings to capture

variations in frequency, amplitude irregularities, and nonlinear dynamics, which are biomarkers of vocal impairment in patients with PD.
A major challenge in this study is the class imbalance issue with the available dataset, where the number of samples is higher for the Parkinson's positive group compared to the normal healthy controls. We used RandomOverSampler for data balancing to address class imbalance. These methods generate additional samples by resampling minority classes, thereby improving classifier performance.
To address class imbalance, researchers employ oversampling techniques like RandomOverSampler or Synthetic Minority Oversampling Technique (SMOTE). SMOTE's effectiveness stems from generating synthetic examples through interpolation, creating more robust decision boundaries than simple replication. Recent research in 2024, such as [4], explores variants like SVMSMOTE to target samples near the decision boundary, improving the model's ability to distinguish early-stage PD cases. Feature selection and dimensionality reduction are crucial for developing effective ML models. High accuracy and robustness in PD detection are difficult to achieve due to high-dimensional acoustic features and class imbalances in medical datasets. Recent studies address these challenges by integrating data resampling and ensemble learning. For example, [5] enhanced model reliability by using SMOTE to balance data and a hybrid relief feature selection method within a speech-based diagnostic framework. Optimizing a CatBoost algorithm via Grid Search further improved diagnostic accuracy, highlighting the importance of combining class-balancing with hyperparameter tuning for early PD identification. Hyperparameter tuning, using RandomizedSearchCV and GridSearchCV, was performed for KNN and XGBoost to refine the performance of these models further. Optimization techniques systematically explore hyperparameter configurations to identify the best setup for each classifier, ensuring robustness and reliability.
While previous studies have demonstrated the potential of machine learning in PD detection, they often rely on single feature selection methods (like RFE or Filter-based) which may get trapped in local optima or fail to capture complex, non-linear feature interactions. Furthermore, few studies have addressed the trade-off between maximizing classification accuracy and minimizing the feature subset size for clinical efficiency. This study addresses this gap by proposing a novel Hybrid Feature Selection method that integrates the global search capability of a GA with the model-driven feature importance of XGBoost. The proposed hybrid method combines XGBoost feature importance and a Genetic Algorithm (GA). First, XGBoost feature importance scores were used to rank the features, and then GA was applied to identify the optimal feature subset.
The main contributions of this study are as follows:
Evaluation of different machine learning classifiers, including KNN, MLP, XGBoost, and NB, for Parkinson's disease detection using voice features.
Comparison of different feature selection methods, including correlation analysis, ANOVA, RFE, RF feature importance, XGBoost feature importance, and GA.
Development of a hybrid feature selection method based on XGBoost feature importance and GA.
Evaluation of the proposed hybrid method using 10-fold stratified cross-validation, where XGBoost achieved a mean accuracy of 92.74% and KNN achieved a Geometric Mean of 91.13%.
Evaluation of the effects of feature scaling and RandomOverSampler on classification performance.
The application of machine learning techniques for Parkinson's disease detection has attracted increasing attention in recent years. Different classification algorithms and feature selection methods have been proposed to improve diagnostic performance. Some representative studies are summarized below.
In [6], a machine learning-based approach was proposed for Parkinson's disease prediction using a hybrid feature selection method. RFE was applied to reduce the original 22 features to 9 features, and an

accuracy of 97.43% was achieved using XGBoost. The study showed that feature selection can improve classification performance.

According to the method proposed in [7], a multiclass ML approach for PD detection is presented, achieving an unprecedented accuracy of 99.11% using a Feedforward Neural Network (FNN). The study utilized voice recordings from the UCI dataset, using the SMOTE Technique and hyperparameter tuning via RandomizedSearchCV. The FNN model demonstrated exceptional performance with 98.78% accuracy, which outperforms traditional models like Kernel Support Vector Machine (KSVM) (95.89% accuracy).

In [8], an approach was proposed that utilizes acoustic signal analysis combined with RFE for the early detection of PD. By extracting relevant features from voice recordings and employing RFE for feature selection, their method achieved high classification accuracy, demonstrating the potential of non-invasive techniques in the diagnosis of PD.

In a study on PD detection, as shown in [9], a hybrid approach combining filter-based feature selection, GAs, and ensemble learning techniques was proposed to classify patients using voice recordings. They evaluated multiple classifiers, including decision trees, RFs, and XGBoost, in two benchmark datasets. Their results demonstrated that the ensemble methods, especially stacking and bagging, improved classification accuracy, with XGBoost contributing to robust performance with 92.40% accuracy after applying ensemble learning methods and 87.75% after applying GA. The study also highlighted challenges, such as class imbalance, and emphasized the importance of optimized feature selection for improving generalization. In addition, according to [10], they conducted a comparative analysis of three classifiers—RF, KNN, and Support Vector Machine (SVM)—on an imbalanced PD dataset. They applied SMOTE to balance the data and used Correlation-Based Feature Selection (CFS) to identify the most relevant attributes. The KNN model achieved the highest accuracy (94.11%) on the balanced dataset, demonstrating the effectiveness of oversampling in handling class imbalance. While many ML models have achieved high accuracy in diagnosis, few have addressed processing time—a key factor in clinical settings. In a related study [11], authors applied ten predictive models to a biomedical voice dataset, using a GA for feature selection. The best model achieved 97.96% accuracy with only 9 features and a processing time of 1.83 seconds, showing a 16.33% improvement in accuracy after feature selection.

Materials

The dataset used in this study is publicly available through the University of Oxford repository, in collaboration with the National Center for Voice. It was initially prepared to study general speech disorders and contained the voice recordings of 31 persons: PD patients = 23 (16 males and 7 females) and Healthy Controls (HC) = 8 (males = 3 and females = 5). The dataset consists of 195 voice recordings, each made in a sound-treated booth for 36 seconds [12].

A set of biological voice measures, as shown in Figure 1. Each voice measure corresponds to 1 column in Table 1, and every individual vocal recording to a row. Each patient had an average of 6 recordings, though 22 patients had 6 recordings and 9 patients had 7 recordings. The patients are between the ages of 46 and 85 (mean age, 65.8; standard deviation, 9.8), and they have been treated for between 0 and 28 years. Speech recordings were made in a sound-treated booth at the Industrial Acoustic Company, using a microphone placed 8 cm from the lips. In the dataset, it is encoded in such a way that the column "status," where 0 = HC and 1 = those with PD, indicates whether a participant is healthy or with PD [12]. Voice measures and their meaning are shown in Figure 1.

Voice measure	Meaning
Name	ASCII name of subject and recording number (categorical variables).
MDVP:Fo(Hz)	Average vocal fundamental frequency (Numerical variables).
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency (Numerical variables).
MDVP:Flo(Hz)	Minimum vocal fundamental frequency (Numerical variables).
MDVP:Jitter(%)	Several measures of variation in fundamental frequency (Numerical variables).
MDVP:Jitter(Abs)	
MDVP: RAP	
MDVP: PPQ	
Jitter:DDP	
MDVP:Shimmer	Several measures of variation in amplitude (Numerical variables).
MDVP:Shimmer(dB)	
Shimmer: APQ3	
Shimmer: APQ5	
MDVP: APQ	
Shimmer:DDA	
NHR	Measures of the ratio of noise to tonal components in the voice (Numerical variables).
HNR	
status	0 for HC and 1 for PD (Numerical variables).
RPDE	Nonlinear dynamical complexity measures (Numerical variables).
D2	
DFA	Signal fractal scaling exponent (Numerical variables).
spread1	Nonlinear measures of fundamental frequency variation (Numerical variables).
spread2	
PPE	

Figure 1. *Features of UCI Dataset [13]*

Methods

The proposed approach uses Google Colab and Python to determine if a patient has PD. Data preparation, including standard scaler and RandomOverSampler, was performed at different stages for feature selection (ANOVA), Random Forest, XGBoost, RFE, and GA. Hyperparameter tuning, using GridSearchCV and RandomizedSearchCV, was employed for the KNN and XGBoost algorithms, respectively.

Data Preprocessing

A vital aspect of data processing is preprocessing, which eliminates extraneous information and enables the model to learn the data's properties efficiently [14]. The dataset was imported into the Google Colab platform using the Pandas software as a CSV file. Based on the "status" column, we found out that the dataset was unbalanced, with 147 for PD and 48 for HC, or 25% for HC and 75% for PD. There were no duplicates or null values. We divided the dataset into a 70:30 train/test split to prevent both underfitting and overfitting, and then scaled each feature separately. We used several libraries for this work, including

Scikit-learn (Sklearn), NumPy, Pandas, Matplotlib, and Seaborn. The mean and standard deviation are stored and used with the StandardScaler transform on the data. The StandardScaler normalization formula is in (1).

$$\text{Standard scaler} = \frac{x - \mu}{\sigma} \tag{1}$$

Feature selection

To enhance the model's performance and reduce dimensionality, multiple feature selection methods were employed. These include filter-based, wrapper-based, and embedded methods, used individually and in hybrid combinations, with 5, 12, 13, and 14 features selected for comparative analysis. The methods were evaluated using KNN, XGBoost, MLP, and NB, along with StandardScaler for normalization and RandomOverSampler for class imbalance correction.

Filter-Based Selection (ANOVA F-test) with Correlation-Based Filtering

To reduce redundancy among the features, Spearman correlation coefficients were calculated for all feature pairs. Features with an absolute correlation value greater than 0.97 were removed [15]. After this step, ANOVA F-test was applied to determine which features showed significant differences between the target classes. Features with p-values lower than 0.05 were selected and ranked according to their F-scores [16]. The resulting feature subsets were used in the classification experiments.

Wrapper-Based Selection (Recursive Feature Elimination - RFE)

RFE was implemented using a logistic regression estimator. In each iteration, the least important feature was removed until the desired number of features was reached. Different feature subset sizes (5, 12, 13, and 14) were evaluated to investigate their effect on classification performance.

Embedded methods (Random Forest and XGBoost)

Feature importance scores obtained from Random Forest were used to rank and select features. Similarly, XGBoost feature importance scores were used to identify the most relevant features. These selected features were subsequently used in the classification models.

GA

GA was used as a wrapper-based feature selection method to identify relevant features for Parkinson's disease classification. Each chromosome was represented as a binary vector indicating whether a feature was selected or not. During the optimization process, selection, crossover, and mutation operators were applied to generate new feature subsets. The fitness of each chromosome was evaluated using the selected classifier, and the search process continued until the stopping criterion was reached [11].
According to Figure 2, by utilizing genetic operators such as selection, which prefers individuals with higher fitness (i.e., better classification performance), crossover, which recombines parts of two parent chromosomes to form offspring, and mutation, which introduces diversity by flipping bits randomly, GAs promote desirable features and explore new solutions [17]. The fitness function used in this study was defined as `1 - accuracy`, which was calculated by training classifiers on a subset of features

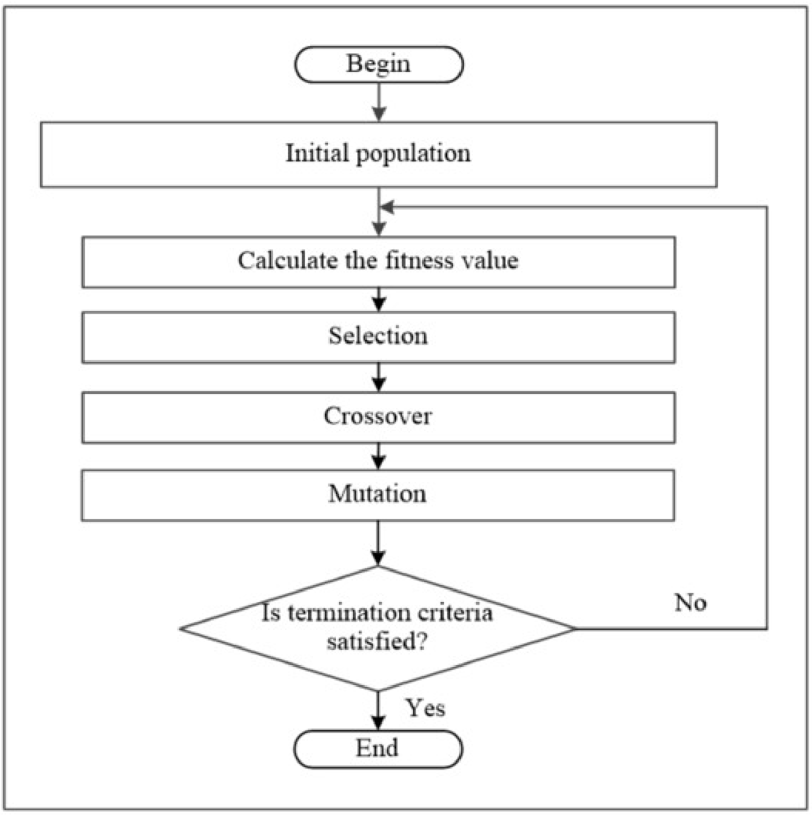


Figure 2. *Flowchart of GA*

represented by each chromosome. By minimizing this function, the GA effectively maximized the classification accuracy. The GA was configured with the following parameters, shown in Table 1:

Table 1. *Hyperparameters for the GA model*

Parameter	Value
Population size	30
Number of generations	50
Mutation probability	0.1
Crossover probability	0.6
Elitism ratio	0.05
Crossover type	Uniform
Maximum iterations without improvement	10

GA performs feature selection through an iterative search process. In this study, GA was used to identify feature subsets that improve classification performance for Parkinson's disease detection.

Experiments were conducted using feature subset sizes of 5, 12, 13, and 14 to evaluate the effect of the number of selected features on classification performance. The selected feature subsets were then evaluated using different machine learning classifiers.

Due to the stochastic nature of GA, slight variations in the selected features were observed across different runs. However, similar performance trends were obtained. The results of GA were also compared with those obtained using the other feature selection methods.

RandomOverSampler

The dataset contains fewer healthy control (HC) samples than Parkinson's disease (PD) samples. To reduce the effect of class imbalance, RandomOverSampler was applied after the train-test split [18]. In the 10-fold cross-validation procedure, oversampling was performed only on the training data of each fold, while the validation data remained unchanged. This approach preserved the original class distribution in the validation set during model evaluation. This strategy effectively balances the classes for training purposes while preventing the model from memorizing synthetic samples during evaluation [19].

Hyperparameter tuning (GridSearchCV - RandomizedSearchCV)

When constructing the ML model, the user often supplies the hyperparameters. For this study, it's better to use GridSearchCV to determine the ideal hyperparameter values, thereby obtaining the best model results. The simplest search method that yields the most accurate outcomes is grid search. Because each test runs separately without relation to time sequence, grid search is easy to do in parallel [20]. It primarily accepts arguments, such as the estimator, param grid, and CV. The following is a description of each argument:
Estimator: the item being utilized as an estimator.
CV: For a K-fold cross-validation, the folds are represented by an integer (cv).
Param grid: A list of parameter values along with their names.
In addition, RandomizedSearchCV is a hyperparameter optimization technique provided by Sklearn that allows for efficient searching for the best hyperparameters of a ML model. It randomly samples a fixed number of combinations from the parameter grid, making it faster and more suitable for large datasets and complex models. By introducing a degree of flexibility, randomness enables the algorithm to dynamically explore the hyperparameter space, which facilitates the iterative improvement of models in a more resource-conscious manner [7].

Classification methods

After preprocessing, the required classifiers were selected and used. KNN, NB, MLP, and XGBoost models are a few of the ML algorithms that were investigated. KNN is a supervised ML algorithm and an easy-to-understand method. The KNN method assumes that related objects are located nearby. Stated differently, comparable objects are near each other [21]. It also provides reasonable results in classifying PD with low error rates. The hyperparameters used are listed below:

1. neighbors = 3
2. distance = manhattan

MLP is a type of feedforward artificial neural network that consists of three layers: an input layer, a hidden layer, and an output layer. The computational engine of the MLP lies in the hidden layers, where neurons transform inputs using non-linear activation functions. Data flows in a forward direction from input to output, and the network is trained using the backpropagation learning technique [22]. By using RandomOverSampler and Sklearn's MLPClassifier, the hyperparameters are mentioned below:

1. Single hidden layer with 100 neurons (hidden layer sizes=100)
2. max iter=500

NB is a probabilistic ML algorithm based on Bayes' theorem with a strong (naive) assumption of independence among features. Despite this simplification, it performs well in many complex real-world problems. Moreover, it is highlighted as a nonparametric classifier, which means it makes fewer assumptions about the form of the underlying data distribution [18].
XGBoost is an optimized and scalable implementation of gradient boosting, recognized as one of the most efficient algorithms in ML [23]. Additionally, it is well-known for its speed and performance. The grid search with RandomizedSearchCV helps optimize the hyperparameters to maximize the model's performance. Hyperparameters are shown in Table 2.

Table 2. *Hyperparameters for the XGBoost Model*

Hyperparameter	Value
min child weight	1
max depth	6
learning rate	0.2
gamma	0.4
colsample_bytree	0.5

Hybrid method

To form the proposed hybrid feature selection approach, we combined XGBoost feature importance with a Genetic Algorithm (GA). This method was designed to leverage the strengths of both embedded and wrapper-based selection techniques—the efficiency and interpretability of XGBoost, and the global search capability of the GA. Rather than running these sequentially, the final hybrid feature subset was constructed by integrating the highest-performing parameters from both independent selection phases while strictly avoiding redundancy. The process was carried out in four key stages: First, XGBoost was trained on the full dataset. The `feature_importances_` attribute of the trained model was used to obtain an initial ranking of features based on their contribution to predictive performance. The top 12 features ranked by XGBoost importance were selected to form the foundational subset (S_{xgb}). Subsequently, the GA was executed independently over the complete feature set. It identified high-performing feature subsets by optimizing a fitness function defined as `1 - accuracy`, where accuracy was calculated from a KNN classifier trained on each selected subset. KNN was chosen as the fitness evaluator due to its low computational cost and high sensitivity to feature relevance during the iterative search process [24, 25]. Finally, we evaluated the optimal feature combinations identified by the GA. From the GA's highest-performing subsets, we isolated and selected exactly 8 distinct features that did not overlap with the foundational XGBoost selection ($S_{ga\_distinct}$). By taking the union of these two distinct groups ($S_{hybrid} = S_{xgb} \cup S_{ga\_distinct}$), we engineered a final, fixed subset of exactly 20 unique features. This deliberate selection process ensures a comprehensive representation of both individual feature strength (via XGBoost) and interactive predictive power (via GA). Figure 3 shows the flowchart of the processing.
Algorithm 1 presents the pseudo-code of the proposed hybrid XGBoost-GA feature selection framework.

Robustness Analysis using K-Fold Cross-Validation

To validate the reliability of the proposed hybrid feature selection method, we implemented a 10-Fold Stratified Cross-Validation scheme. Within each fold, class imbalance was addressed using RandomOverSampler, applied strictly to the training data to prevent data leakage.

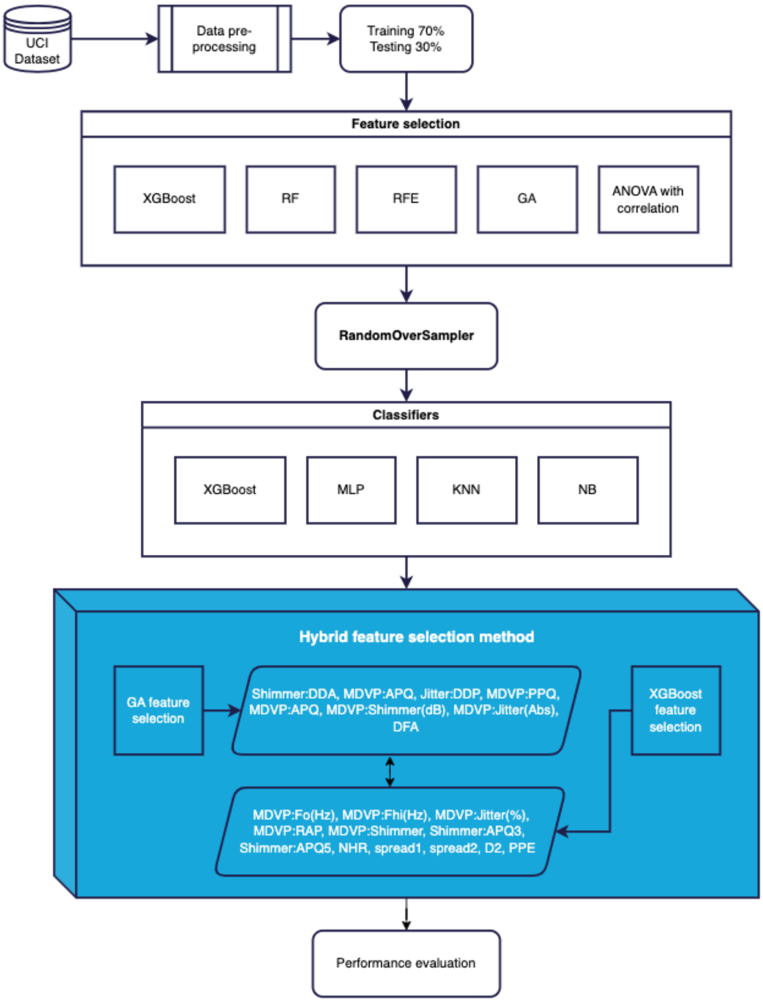


Figure 3. *Overall layout of proposed method*

Algorithm 1 Proposed Hybrid XGBoost-GA Feature Selection Framework (Parallel Strategy)

Require: Dataset $D = \{X, y\}$ with full feature set F
Ensure: Hybrid feature subset S_{hybrid}, Final classification metrics
Parameters: $N_{xgb} = 12$, $N_{ga} = 8$, $Pop = 30$, $Gen = 50$, $P_c = 0.6$, $P_m = 0.1$
Phase 1: Filter-Based Selection (XGBoost)

1. Train XGBoost classifier on full dataset D
2. Compute feature importance scores for all $f \in F$
3. Select top 12 ranked features to obtain foundational subset S_{xgb}

Phase 2: Wrapper-Based Selection (GA)

1. Initialize population P with Pop random binary chromosomes of length $|F|$
2. for generation $g = 1$ to Gen do
   (a) for each chromosome $Chrom_i$ in P do
      i. Decode $Chrom_i$ to obtain selected feature subset $S_i \subseteq F$
      ii. Train KNN classifier using S_i
      iii. Compute accuracy on validation set
      iv. Define fitness as $1 -$ accuracy
   (b) Perform selection, crossover (P_c), and mutation (P_m)
   (c) Update population P
3. Decode best chromosome $Chrom_{best}$ to obtain optimal GA subset S_{ga}
4. Filter S_{ga} to isolate exactly 8 features that do not overlap with S_{xgb}
5. Define this non-overlapping set as $S_{ga\_distinct}$

Phase 3: Hybrid Subset Construction

1. Construct final hybrid feature subset of exactly 20 features:

$$S_{hybrid} = S_{xgb} \cup S_{ga\_distinct}$$

Phase 4: Final Model Evaluation

1. Apply S_{hybrid} to original dataset D
2. Initialize classifiers $C = \{\text{KNN}, \text{XGBoost}, \text{MLP}, \text{NB}\}$
3. Perform 10-fold stratified cross-validation
4. Apply RandomOverSampler within training folds
5. Compute mean Accuracy, F1-score, and Geometric Mean

return Final performance metrics

To enhance the model's performance and reduce dimensionality, multiple feature selection methods were employed. These include filter-based, wrapper-based, and embedded methods, used individually and in hybrid combinations, and selecting 5, 12, 13, and 14 features for comparative analysis. The methods were evaluated using KNN, XGBoost, MLP, and NB, along with StandardScaler for normalization and RandomOverSampler for class imbalance correction. In addition to accuracy and F1-score, Geometric Mean (G-Mean) was also reported since the dataset is imbalanced [26].

Results and Discussion

In this study, different machine learning algorithms were used for Parkinson's disease classification based on voice signal features. The performance of the models was evaluated using different feature selection methods and preprocessing techniques.

The evaluated classifiers included KNN, MLP, XGBoost, and NB. Feature selection was performed using ANOVA with correlation filtering, RFE, XGBoost feature importance, RF feature importance, and a GA-based wrapper method. StandardScaler was applied for feature scaling, and RandomOverSampler was used to reduce the effect of class imbalance.

In addition, the models were evaluated without feature scaling and oversampling to investigate the effect of preprocessing on classification performance. Across all classifiers, performance was consistently lower without preprocessing. For example, KNN and MLP showed a significant drop in accuracy, highlighting the importance of standardized feature values for distance-based and neural network models. The impact of oversampling was especially pronounced in improving recall and sensitivity for the minority class, particularly for NB and KNN.

Feature subsets of sizes 5, 12, 13, and 14 were selected to investigate the effect of dimensionality. In general, larger subsets yielded slightly better performance. Generally, five features showed lower performance compared to the others. Among the tested methods,KNN achieved the highest accuracy when 12 features were selected using XGBoost feature importance with 98.31% accuracy.

Moreover, the GA method produced strong results, especially with MLP (95% accuracy for 13 features), KNN when combined with GridSearchCV and XGBoost. The NB classifier, while computationally efficient, achieved a moderate accuracy of 83% with XGBoost and the GA feature selection method when using 13 selected features with slightly lower recall and precision due to its sensitivity to feature correlations. Tables 3, 4, 5, 6, 7 show the accuracy of ANOVA, GA, XGBoost, RFE, and RF methods on classification algorithms, respectively.

Table 3. *Results of ANOVA with correlation feature selection method*

Model	Acc by 5 features	Acc by 12 features	Acc by 13 features	Acc by 14 features
KNN	94.92	91.53	93.22	93.22
MLP	93.22	86.44	93.22	93.22
XGBoost	86.44	89.93	88.14	89.83
NB	77.97	83.05	83.05	83.05

Preliminary Analysis (Hold-Out Split)

Initial experiments were conducted using a fixed 70:30 train-test split to evaluate the efficacy of the hybrid feature selection method. XGBoost feature importance was combined with GA to form a hybrid feature selection approach. Initially, features were ranked based on their importance scores derived from a trained

Table 4. *Results of GA feature selection method*

Model	Acc by 5 features	Acc by 12 features	Acc by 13 features	Acc by 14 features
KNN	88.13	91.52	93.22	93.22
MLP	89.83	93.22	94.91	93.22
XGBoost	84.74	89.83	88.13	91.52
NB	86.44	81.35	83.05	81.35

Table 5. *Results of XGBoost feature selection method*

Model	Acc by 5 features	Acc by 12 features	Acc by 13 features	Acc by 14 features
KNN	89.83	98.31	96.61	94.92
MLP	89.83	91.52	91.52	94.91
XGBoost	93.22	89.83	89.83	94.91
NB	83.05	83.05	83.05	83.05

Table 6. *Results of RFE feature selection method*

Model	Acc by 5 features	Acc by 12 features	Acc by 13 features	Acc by 14 features
KNN	92.61	92.64	92.61	94.84
MLP	88.13	88.13	88.13	93.22
XGBoost	91.52	88.13	91.52	93.22
NB	83.05	74.57	74.57	79.66

Table 7. *Results of RF feature selection method*

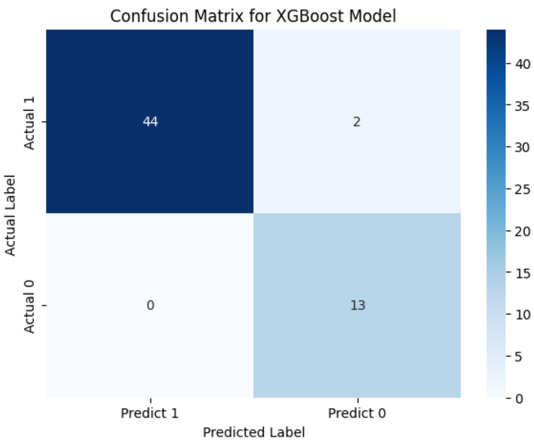
Model	Acc by 5 features	Acc by 12 features	Acc by 13 features	Acc by 14 features
KNN	89.83	96.61	98.31	96.61
MLP	91.53	91.53	91.53	88.14
XGBoost	89.83	88.14	91.53	91.53
NB	77.97	81.36	79.66	79.66

XGBoost model. The GA was then applied to the ranked feature subset obtained from XGBoost feature importance. When the number of selected features was set to 20, the KNN classifier achieved the highest accuracy of 98.31%. Figure 4 presents the confusion matrix obtained for this split and shows that most PD and healthy control samples were correctly classified. The results also indicate that class imbalance affected the classification performance before applying the preprocessing steps. According to Table 8, KNN achieved the best overall performance among all classifiers in terms of accuracy, F1-score, and Geometric Mean. XGBoost also produced competitive results, achieving an accuracy of 96.61% and the highest precision value among the evaluated classifiers. MLP achieved an accuracy of 91.52%, while NB produced lower performance compared with the other methods. Overall, the results show that the proposed hybrid feature selection method can improve classification performance by selecting a compact and informative feature subset. NB, as expected, lagged behind the other models with an accuracy of 72.88%, primarily due to its strong assumptions of feature independence, which are not fully met in this dataset. However, the application of RandomOverSampler improved minority class detection across all models. However, given the limited dataset size we acknowledge that results from a single fixed split may be susceptible to data partitioning bias.

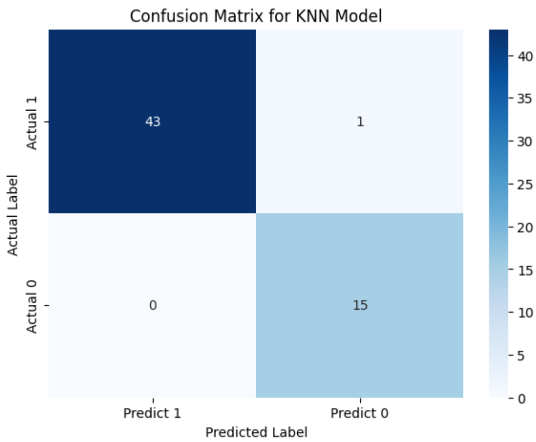
E-ISSN: 2564-7873

Table 8. *Performance of the models with hybrid feature selection approach*

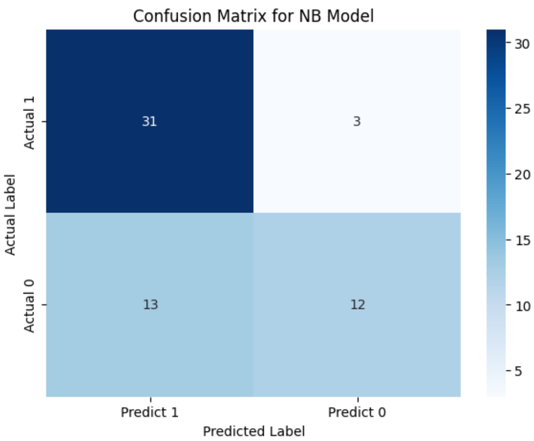
Model	Accuracy	Precision	Recall	F1-Score
KNN	98.31	97	99	98
MLP	91.52	90	88	89
XGBoost	96.61	98	93	95
NB	72.88	70	75	70



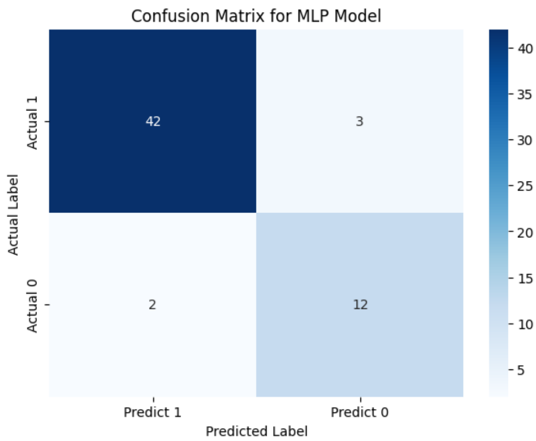
(a) *XGBoost*



(b) *KNN*



(c) *Naïve Bayes*



(d) *MLP*

Figure 4. *Confusion matrices of the proposed models*

This study focused on investigating the application of ML algorithms for diagnosing PD, with an emphasis on enhancing the results through feature selection, preprocessing techniques, and hyperparameter tuning. The comparison with previous studies highlights the potential of combining advanced preprocessing techniques and hyperparameter optimization to enhance predictive accuracy. Furthermore, the results of the models demonstrate the effectiveness of ML approaches in analyzing high-dimensional and nonlinear data.

To assess the importance of preprocessing steps such as feature scaling and class balancing, all models were also tested without applying StandardScaler or RandomOverSampler. As expected, classifiers such as KNN, NB, and MLP demonstrated noticeable reductions in accuracy without preprocessing, as these models are sensitive to differences in feature scale and class distribution.

Interestingly, XGBoost maintained strong performance even without any preprocessing. In some configurations, its accuracy without normalization or oversampling was comparable to, or even higher than, that of the preprocessed setup. This can be attributed to XGBoost's gradient-boosted tree structure, which inherently handles varying feature scales and is more tolerant of imbalanced data due to internal handling of loss gradients and split criteria [27, 28] . For instance, a study [29] reported an accuracy of 84.80% using baseline features without extensive preprocessing. After feature selection, the accuracy slightly improved to 85.60%, indicating that XGBoost can perform well even with minimal data preparation. Our study achieved 95% accuracy without preprocessing. Table 9 shows the results without preprocessing.

Table 9. *Model's performance without preprocessing*

Model	Accuracy	Precision	Recall	f1score
KNN	82.05	70	72	71
MLP	84.61	74	68	70
XGBoost	94.78	97	86	90
NB	69.23	62	70	62

Robustness Analysis (10-Fold Cross-Validation)

To strictly validate the model's generalizability and address the limitations of the single split, we implemented a 10-Fold Stratified Cross-Validation scheme. This approach ensures that every recording is used for both training and testing, providing a statistically robust performance estimate.

Class imbalance was addressed using RandomOverSampler within each fold to prevent data leakage.

Table 10. *Performance of models with hybrid feature selection and 10-Fold CV*

Classifier	Mean Accuracy (± SD)	Mean F1-Score	Mean Geometric Mean (± SD)
KNN	91.26% (± 4.04%)	93.98%	91.13% (± 5.62%)
XGBoost	92.74% (± 7.87%)	95.11%	89.78% (± 9.55%)
MLP	91.32% (± 7.15%)	94.22%	88.03% (± 11.28%)
NB	69.82% (± 10.81%)	74.65%	75.48% (± 9.95%)

Using 10-Fold Stratified Cross-Validation as shown in Table 10, XGBoost achieved the highest mean classification accuracy of 92.74% (± 7.87%). However, KNN exhibited superior stability, reflected by the lowest standard deviation (± 4.04%), and achieved the highest Geometric Mean of 91.13% (± 5.62%). Since the Geometric Mean is a critical metric for evaluating performance on imbalanced datasets, this result indicates that KNN provides the most balanced trade-off between sensitivity and specificity, ensuring reliable identification of both PD and HC cases [26]. In comparison, XGBoost (89.78%) and MLP (88.03%) showed a slight bias toward the majority class, while MLP and NB demonstrated higher variance and substantially lower geometric means (88.03% and 75.48%, respectively), making them less suitable for this clinical classification task.

While XGBoost yielded the highest raw accuracy, the proposed KNN-based hybrid framework offers the best combination of high accuracy, low variance, and balanced detection capability. In contrast, NB performed poorly (69.82%), confirming its unsuitability for this high-dimensional feature space.

These results confirm that the proposed hybrid feature selection method yields robust and generalizable performance, effectively distinguishing between PD and HC subjects even under rigorous statistical validation. As shown in Table 11, several recent studies [6, 7] report high accuracies (up to 98.78%) using simple train-test splits. However, given the small size of the UCI dataset (195 samples), such results are highly susceptible to data partitioning bias. In contrast, our proposed method achieves a competitive 92.74% accuracy under 10-Fold Stratified Cross-Validation, which is a far more rigorous standard for generalization. Furthermore, our KNN model achieves a geometric mean of 91.13%, indicating superior stability in distinguishing between PD and HC subjects compared to methods that maximize accuracy alone.

Table 11. *Comparison of Proposed Method with Existing Literature*

Study	Dataset	Feature Selection	Classifier	Validation Method	Accuracy
Wasif et al. [6]	UCI Voice	RFE	XGBoost	Train-Test Split	97.43
Mohapatra et al. [5]	UCI Voice	RReliefF + SMOTE	CatBoost	Cross-Validation	92.61
Srinivasan et al. [7]	UCI Voice	None	FNN	Train-Test Split	98.78
Ali et al. [9]	Multiple	GA + Filter	Ensemble	Cross-Validation	92.40
Proposed Method	UCI Voice	Hybrid Method	KNN	10-Fold CV	91.26
Proposed Method	UCI Voice	Hybrid Method	XGBoost	10-Fold CV	92.74

In this study, we investigated the effectiveness of GA and XGBoost feature importance for feature selection in classifying PD using various ML models. Our hybrid approach demonstrated superior performance across multiple classifiers, particularly improving classifier performance, as evidenced by notable enhancements in KNN and XGBoost, indicating its effectiveness in enhancing model accuracy. It leverages the strengths of both methods: GA's global search capability and XGBoost's ability to capture feature relevance. This synergy enhances classifier performance by enabling the selection of a feature subset, leading to improved accuracy. By efficiently providing initial diagnostics, the suggested models can help lower treatment costs. Additionally, this can serve as a soft diagnostic tool for doctors and a teaching tool for medical students. Furthermore, several potential ways exist to enhance the scalability and accuracy of this prediction model.

Our findings align with results from previous research while offering superior statistical validation. For instance, while prior studies reported KNN accuracies around 96% [30] and XGBoost betweeen 91% and 94% [30, 31] using standard splits, our method achieved a mean cross-validation accuracy of 92.74% for XGBoost and 91.26% for KNN. Although these values are slightly more conservative than the 98% observed in our preliminary single-split analysis, they represent a more reliable estimate of generalizability derived from 10-Fold Stratified Cross-Validation. Consistent with the literature, NB yielded the lowest performance (69.82%), mirroring earlier findings of 71-74% [32, 33]. These results confirm that the proposed hybrid feature selection strategy, combined with RandomOverSampler for class balancing, effectively captures vital vocal patterns. The high Geometric Mean achieved by KNN (91.13%) further indicates that selected features generalize well to unseen data, presenting a robust solution for clinical decision support systems.

The proposed hybrid diagnostic framework demonstrates significant potential for integration into Clinical Decision Support Systems (CDSS) for the early screening of PD. By utilizing voice recordings—a non-invasive, low-cost, and easily obtainable biomarker—this method offers a practical alternative to expensive and invasive traditional diagnostic tools.

1. Telemedicine and Remote Monitoring: The proposed method uses only 20 selected features and may be suitable for integration into telemedicine systems. Voice samples can be collected remotely, enabling the monitoring of individuals without requiring frequent hospital visits. Such systems may be particularly useful for elderly individuals and patients living in areas with limited access to healthcare services.
2. Clinical Decision Support: The KNN classifier achieved a Geometric Mean of 91.13%, indicating balanced classification performance. Therefore, the proposed approach may be used as a decision support tool to assist clinicians in the preliminary assessment of Parkinson's disease. Individuals identified as high risk can be referred for further clinical evaluation.
3. Computational Efficiency: The proposed method is based on machine learning algorithms such as XGBoost and KNN and does not require specialized hardware during the prediction stage. As a result, the method can be implemented on standard computer systems and may be suitable for practical applications.

Limitation of the Study

The UCI Parkinson's dataset contains 195 voice recordings obtained from 31 subjects, including 23 individuals with Parkinson's disease and 8 healthy controls. Due to the limited number of subjects, particularly in the healthy control group, subject-wise 10-fold cross-validation may lead to unstable validation folds and unreliable performance estimates. Therefore, a record-wise stratified 10-fold cross-validation strategy was adopted to preserve class distribution across folds and ensure stable model evaluation. Although this approach has been widely used in similar studies, it may introduce a risk of learning subject-specific characteristics. Consequently, the reported results should be interpreted within the scope of the available dataset. Future studies should evaluate the proposed approach on larger and independent datasets using subject-wise data partitioning to further assess its generalization capability.

Computational Complexity and Resource Analysis

The hybrid framework employs parallel feature selection, using XGBoost and GA independently on the complete feature set, followed by hybrid feature union and classifier evaluation.

Initially, XGBoost was trained on the 206-sample training set to calculate feature importance. This computationally inexpensive step selected the top 12 features.

The GA-based wrapper method constitutes the primary computational load. With a population of 30 and 50 generations (1500 fitness evaluations), each chromosome (feature subset) was evaluated using a KNN classifier trained on the training set. Although KNN's fitness evaluation scales quadratically with the number of training samples, the small training set (206 instances) kept this cost manageable. The GA phase took 1–3 minutes per run, dominating execution time.

Following GA convergence, the top 8 GA-selected features were combined with the 12 XGBoost-selected features, creating a hybrid subset of up to 20 features.
Finally, four classifiers (KNN, XGBoost, MLP, and NB) were trained and evaluated on this hybrid subset using 10-fold stratified cross-validation on the training set, then independently tested on the 59-sample test set. This reduced-dimensionality stage (maximum 20 features) was computationally efficient, taking less than one minute.

Hardware and Resource Consumption

All experiments were conducted on a system equipped with an Intel Core i7 processor and 16 GB RAM, without GPU acceleration.
Total execution time per experiment: approximately 2–4 minutes
GA phase contribution: 70% of total runtime
Peak RAM usage: below 2 GB
CPU utilization: moderate to high during GA optimization, low during other phases
The framework demonstrated stable execution without memory constraints. The available 16 GB RAM was substantially more than required for this dataset size.
Given the small dataset size, the proposed hybrid method is computationally efficient and well-suited for biomedical datasets of similar scale. Although the GA fitness evaluation exhibits quadratic behavior, the limited sample size ensures practical runtime feasibility on standard desktop hardware.

Conclusion

Parkinson's disease is an important neurological disorder, and early diagnosis can support disease management. In this study, a hybrid feature selection method based on XGBoost feature importance and a Genetic Algorithm (GA) was proposed. Different feature selection methods and machine learning classifiers were evaluated using voice signal features. The results showed that the proposed hybrid method achieved the best overall performance among the evaluated approaches. Using 10-fold stratified cross-validation, the proposed method achieved a mean accuracy of 92.74% with XGBoost and a Geometric Mean of 91.13% with KNN.
The findings indicate that combining model-based feature importance with an evolutionary search strategy can improve feature selection and classification performance for Parkinson's disease detection. The proposed approach can contribute to voice-based decision support systems by providing an effective framework for disease classification.
Although promising results were obtained, the study was conducted on a single dataset with a limited number of subjects. Therefore, further validation on larger and independent datasets is needed. Future studies will focus on evaluating the proposed method on different datasets and improving computational efficiency.

Declarations

Acknowledgments:

Funding/Financial Disclosure: The authors declare that no financial support was received for the research, authorship, or publication of this study.

Ethics Committee Approval and Permissions: The authors declare that this study does not require any ethics committee approval or special permission.

Conflict of Interests: The authors stated that there are no conflict of interest in this article.

Authors Contribution: Authors contributed equally to the study.

Artificial Intelligence (AI): The authors declare that AI tools (such as ChatGPT and Gemini) were used solely for language correction, translation, and readability enhancement during the preparation of this article. The authors assume full responsibility for the originality, accuracy, and integrity of the entire text, including all AI-assisted language outputs.

References

[1] Iyer, A., Di Giovanni, M., Clark, D., Hagenaars, S. P., Ritchie, S. J., Muñoz-Sandoval, D., Zhang, X., Leng, J., Waddell, J., Carlile, S., & Simonyan, K. (2023). A machine learning method to process voice samples for identification of Parkinson's disease. *Scientific Reports*, *13*(1), 20615. https://doi.org/10.1038/s41598-023-47568-w

[2] Zesiewicz, T. A. (2019). Parkinson disease. *Continuum*, *25*(4), 896–918. https://doi.org/10.1212/con.0000000000000764

[3] Mathur, A., Dwivedi, R. K., & Rastogi, R. (2024). A survey of machine learning-based approaches for alzheimer's disease prediction. *Educational Administration Theory and Practices*. https://doi.org/10.53555/kuey.v30i1.5984

[4] Shyamala, K., & Navamani, T. M. (2024). Design of an efficient prediction model for early Parkinson's disease diagnosis. *IEEE Access*, *12*, 137295–137309. https://doi.org/10.1109/access.2024.3421302

[5] Mohapatra, S., Swain, B. K., & Mishra, M. (2025). Early Parkinson's disease identification via hybrid feature selection from multi-feature subsets and optimized catboost with smote. *Systems Science & Control Engineering*, *13*(1). https://doi.org/10.1080/21642583.2025.2498909

[6] Wasif, T., Hossain, M. I. U., & Mahmud, A. (2021). Parkinson disease prediction using feature selection technique in machine learning. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 1–5. https://doi.org/10.1109/icccnt51525.2021.9580151

[7] Srinivasan, S., Ramadass, P., Mathivanan, S. K., Panneer Selvam, K., Shivahare, B. D., & Shah, M. A. (2024). Detection of Parkinson disease using multiclass machine learning approach. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-64004-9

[8] Alalayah, K. M., Senan, E. M., Atlam, H. F., Ahmed, I. A., & Shatnawi, H. S. A. (2023). Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method. *Diagnostics*, *13*(11), 1924. https://doi.org/10.3390/diagnostics13111924

[9] Ali, A. M., Salim, F., & Saeed, F. (2023). Parkinson's disease detection using filter feature selection and a genetic algorithm with ensemble learning. *Diagnostics*, *13*(17), 2816. https://doi.org/10.3390/diagnostics13172816

[10] Majed, R. J., Al-Heddi, R. M., & Zeki, A. M. (2023). Parkinson's disease detection using data mining models: A comparative study. In 2023 4th International Conference on Data Analytics for Business and Industry (ICDABI), IEEE, 290–294. https://doi.org/10.1109/icdabi60145.2023.10629302

[11] Bunterngchit, C., & Bunterngchit, Y. (2022). A comparative study of machine learning models for Parkinson's disease detection. In 2022 International Conference on Decision Aid Sciences and Applications (DASA), IEEE, 465–469. https://doi.org/10.1109/dasa54658.2022.9765159

[12] Little, M., McSharry, P., Hunter, E., Spielman, J., & Ramig, L. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nature Precedings*. https://doi.org/10.1038/npre.2008.2298.1

[13] Guyon, I., & Elisseeff, A. (2006). An Introduction to Feature Extraction. Springer Berlin Heidelberg, 1–25. https://doi.org/10.1007/978-3-540-35488-8_1

[14] Dogra, V., Singh, A., Verma, S., Kavita, Jhanjhi, N. Z., & Talib, M. N. (2021), Understanding of Data Preprocessing for Dimensionality Reduction Using Feature Selection Techniques in Text Classification. Springer Nature Singapore, 455–464. https://doi.org/10.1007/978-981-16-3153-5_48

[15] Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

[16] Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.

[17] Albadr, M. A., Tiun, S., Ayob, M., & AL-Dhief, F. (2020). Genetic algorithm based on natural selection theory for optimization problems. *Symmetry*, *12*(11), 1758. https://doi.org/10.3390/sym12111758

[18] Sharanyaa, S., Renjith, P. N., & Ramesh, K. (2020). Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 437–442. https://doi.org/10.1109/iciss49785.2020.9316078

[19] Brownlee, J. (2020). *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-sensitive Learning*. Machine Learning Mastery.

[20] Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications.

[21] Hossain, E., Hossain, M. F., & Rahaman, M. A. (2019). A color and texture based approach for the detection and classification of plant leaf disease using knn classifier. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, 1–6. https://doi.org/10.1109/ecace.2019.8679247

[22] Abirami, S., & Chitra, P. (2020). Chapter fourteen - energy-efficient edge based real-time healthcare support system. In P. Raj & P. Evangeline, editors, The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases, volume 117 of *Advances in Computers*, Elsevier, 339–368. https://doi.org/10.1016/bs.adcom.2019.09.007

[23] Zoralioğlu, Y., Gül, M. F., Azizoğlu, F., Azizoğlu, G., & Toprak, A. N. (2023). Predicting academic performance of students using machine learning techniques. In 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 1–6. https://doi.org/10.1109/asyu58738.2023.10296648

[24] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/tit.1967.1053964

[25] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. volume 4. Springer.

[26] Ramyachitra, D., & Manikandan, P. (2014). Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, *5*(4), 1–29.

[27] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16, ACM, 785–794. https://doi.org/10.1145/2939672.2939785

[28] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00021

[29] Abdurrahman, G., & Sintawati, M. (2020). Implementation of xgboost for classification of Parkinson's disease. *Journal of Physics: Conference Series*, *1538*(1), 012024. https://doi.org/10.1088/1742-6596/1538/1/012024

[30] Khanum, D. A., Akhtar, A., Raza, S. A., Ansari, M. I., & Shaikh, H. (2022). Parkinson's disease detection using machine learning algorithms. *International Journal for Research in Applied Science and Engineering Technology*, *10*(10), 786–790. https://doi.org/10.22214/IJRASET.2022.46272

[31] Hani, A. A., Abdulwahab, A. A., Nasser, M., Al-Ali, S., & Almohammad, A. (2024). Comparative analysis of state-of-the-art classifiers for Parkinson's disease diagnosis. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*, 409–423. https://doi.org/10.22437/jiituj.v8i2.32771

[32] Mittal, C., & Sharma, A. (2022). Parkinson's disease detection using different machine learning algorithms. *International Journal of Scientific and Research Publications*, *12*(2), 23. https://doi.org/10.29322/ijsrp.12.02.2022.p12205

[33] Elen, A., & Avuclu, E. (2020). A comparison of classification methods for diagnosis of Parkinson's. *International Journal of Intelligent Systems and Applications in Engineering*, *8*(4), 164–170. https://doi.org/10.18201/ijisae.2020466309