Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-025-10358-9

RESEARCH ARTICLE-ELECTRICAL ENGINEERING

# Automated Early Prediction of Parkinson's Disease Based on Artificial Intelligent Techniques

**Shereen A. Bakry**[1] · **Nourelhoda M. Mahmoud**[1]

Received: 31 January 2025 / Accepted: 24 May 2025


**Abstract**
Parkinson's disease (PD) is one of the most prevalent neurodegenerative disorders, with a sharp increase predicted. Classifying and predicting PD at an early stage is crucial. Application of artificial intelligence (AI) is a significant factor in the diagnosis of various disorders. Based on patient data, machine learning (ML) and deep learning (DL) can automatically predict PD. This research aims to develop an automated approach for early PD prediction based on vocal symptoms and AI techniques. To forecast PD, specific AI models have been implemented. Extreme gradient boosting (XGB or XGBoost), artificial neural networks (ANN), Naive Bayes (NB), K-nearest neighbor (KNN), multilayer perceptron (MLP), logistic regression (LR), support vector machine (SVM), and ridge classifier with cross-validation (RidgeCV) were among the AI models used. The dataset was subjected to various data preprocessing approaches, such as Min–Max scaling and synthetic minority over-sampling technique (SMOTE). Sensitivity, accuracy, F1-score, precision, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC) were among the evaluation measures used to assess the effectiveness of the implemented AI system. The results demonstrated that, with 98% accuracy, 97% precision, 100% sensitivity, 98% F1-score, 97% specificity, and 100% AUC, the XGB model utilizing SMOTE approach achieved the best results. With the proposed approach, patients can forecast their PD early. The proposed work contributes significantly to the field of neurodegenerative disease research by demonstrating the effectiveness of AI techniques in early PD prediction, which can have profound implications for patient care and treatment strategies.

**Keywords** Parkinson's disease · Artificial intelligence · Biomedical applications · Synthetic minority over-sampling technique · Extreme gradient boosting

## 1 Introduction

PD is called "paralysis agitans" [1]. PD has been described as a complex neurodegenerative illness with a range of symptoms brought on by the degeneration of dopaminergic neuronal cells in substantia nigra (SN) [2]. While Alzheimer's disease is considered the first prevalent neurodegenerative disorder, the PD is considered the second one. It significantly impairs individuals and diminishes their quality of life. The appearance of PD is higher in developed countries and increases with age. Affecting 1–2% of the global population over 65 and 4–5% of those over 85, PD is extremely rare in individuals under 40 [3]. In the previous 25 years, the incidence of PD has doubled. In 2019, an estimated 8.5 million people globally were living with PD. Current estimates indicate that PD caused 329,000 deaths in 2019, a more than 100% increase since 2000. Additionally, disability-adjusted life years (DALYs) due to Parkinson's disease (PD) increased by 81% from 2000 to 5.8 million in 2019 [4]. In the coming years, it is expected that there will be an expansion in efforts to find effective solutions, which will require the use of appropriate tests and frameworks. Further research in this area is essential because taking the correct steps is critical for both diagnosing and treating the condition. As part of a commitment to models, the ideal characteristics of datasets are communicated, and predictions are generated. The substantia nigra, a cluster in the midbrain that produces dopamine, should be visible as depicted in Fig. 1[5].

Many researchers believe that environmental and genetic factors caused PD, including environmental toxins, head trauma, drinking water, living in rural areas and pesticide

✉ Nourelhoda M. Mahmoud
nourelhoda.mahmoud@mu.edu.eg

[1] Biomedical Engineering Department, Faculty of Engineering, Minia University, Minia, Egypt

Published online: 15 June 2025

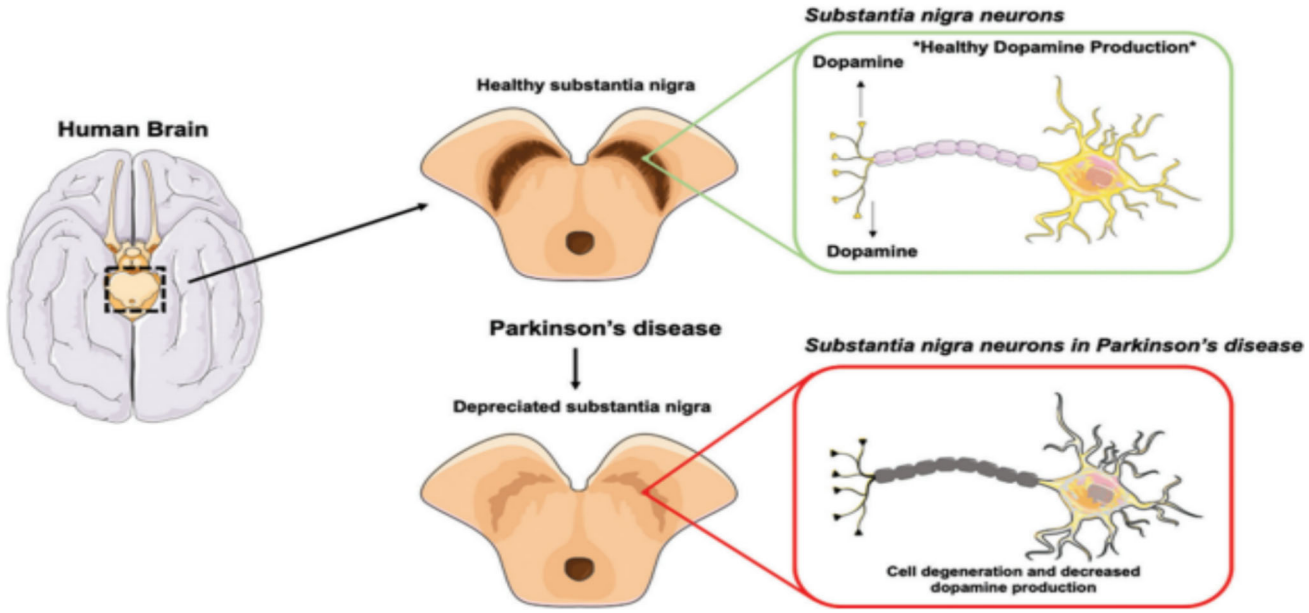


**Fig. 1** SN neurons in PD [5]

exposure. The mentioned factors can vary from person to person. Additionally, each individual with PD may experience specific symptoms in a unique way. PD is divided into various stages, with the first stage being the mildest. At this stage, PD patients experience minimal interference with daily activities, and symptoms like tremors are usually limited to one side of the body. The second one is considered moderate, where symptoms such as resting tremors, stiffness, and trembling can be felt on both sides of the body, and facial expressions change. The third one is the mid-stage, where significant changes including balance loss and decreased flexibility. Occupational therapy and medication may help reduce symptoms. The fourth stage is a progressive stage, where the condition worsens, and patients may require assistive devices like a walker to move. The fifth stage is the most advanced and debilitating stage of PD, where stiffness in the legs cause freezing when standing, and patients may be unable to stand without falling. They may also experience hallucinations and occasional delusions [6].

Idiopathic Parkinson's disease (IPD), also known as Parkinson's, vascular parkinsonism, and drug-induced parkinsonism are the three main forms of PD. The majority of people with parkinsonism have IPD, which is the most common kind. Idiopathic means that the cause is unknown. IPD is a multisystemic synucleinopathy of the nervous system affecting the nigrostriatal system and maybe useful for extra-neurological diagnostic purposes. Examples of inclusion body diseases that occur intracerebrally in a small number of susceptible nerve cell types are Lewy plaques, Lewy bodies, and Lewy neurites. Melano-neurons and projection cells that generate long, sparsely or non-myelinated axons are particularly vulnerable [7]. PD symptoms fall into two categories: non-motor and motor. Bradykinesia, tremor, stiffness, and postural instability are examples of motor symptoms. Gastrointestinal issues, anorexia, anxiety, dyspepsia, pain, constipation, dysphagia, fatigue, psychosis, hallucinations, cognitive decline, issues with impulse control, and dementia are examples of non-motor symptoms [8]. Figure 2 displays PD non-motor symptoms, while Fig. 3 displays PD motor symptoms.

Most people with PD experience dysphonia or voice impairment. Voice-related issues are one dysphonia-related measure that can be used to assess individuals at different stages of the condition. Ninety percent of speech or voice data is thought to be useful for making a diagnosis and determining the presence of a disease. The two main speech problems that people with PD encounter are dysarthria and hypophonia. A person suffering from dysarthria will speak slowly and hardly at all, while a person with hypophonia will speak very softly and weakly. Damage to the central nervous system causes these illnesses. As a result, most PD physicians identify dysarthria in their patients and work to restore voice intensity based on tailored therapy [9]. AI has an essential function in Numerous fields[10, 11]; it helps clinicians with a variety of patient care and intelligent health systems [12, 13]. Healthcare uses AI extensively to diagnose illnesses, find medications, and assess patient risks. These approaches range from ML to DL. AI techniques are reliably identifying illnesses, therefore a range of medical data sources such as computed tomography scans, genomics, ultrasound, magnetic resonance imaging, mammography, and others must be


**Fig. 2** Non-motor symptoms


**Fig. 3** Motor symptoms

used. Additionally, AI improved hospital stays and expedited patients' readiness to resume their therapy at home [14].

The proposed paper aims to construct an intelligent system for the early prediction of PD based on vocal symptoms. A variety of AI models, including RidgeCV, XGB, MLP, LR, ANN, KNN, NB, and SVM. The dataset underwent data preparation using methods like SMOTE and Min–Max scaling. The system efficiency was assessed using the assessment metrics, which included specificity, F1-score, accuracy, precision, sensitivity, and AUC. The remaining sections are structured as: Sect. 2 outlines the relevant works, including an in-depth analysis of the dataset, methods, and techniques put forth by various authors. Section 3 delineates the comprehensive analysis of the proposed methodology and materials, encompassing data acquisition, data preparation and processing, data splitting, data augmentation, the proposed models, and metrics employed for assessment. Section 4 demonstrates the results, discusses the performance achieved by the proposed AI models and compares the previous studies with the most accurate proposed model. Finally, Sect. 5 concludes the proposed work.

## 2 Literature Review

This section provides illustrations of the in-depth study of the literature review. A review of the literature is provided, together with information about the datasets, AI models, and assessment criteria used, along with their names and sources.

Govindu, A. et al. [15] proposed ML models for the early detection of PD. The audio data from the University of California, Irvine (UCI) about vocal modulations of PD patients were the dataset used. LR, KNN, SVM, and random forest (RF) were the classification models that were employed. The outcomes demonstrated that the RF classifier was the most effective ML method for PD detection. AUC of 70%, sensitivity of 95%, precision of 86%, and detection accuracy of 91.83% were all attained by the RF classifier model. The limitation of this study was limited comparison with DL models, while traditional ML models were compared, DL techniques with proven efficacy in feature extraction for Parkinson's detection were not extensively explored. Another limitation was imbalanced dataset, the original dataset had a significant imbalance (109 records of Parkinson's patients and 40 healthy individuals), which could bias the models.

Radha et al. [16] demonstrated classification models to distinguish between PD patient samples and healthy subjects.

The dataset was collected from King's College London Hospital in September 2017. The dataset comprised 37 voice samples from the same participants who were invited to have a spontaneous conversation with the test executor, in addition to 37 sound recordings from those who read the paragraph aloud to the executor. Models such as the hidden Markov model (HMM), ANN, and convolutional neural network (CNN) were used. ANN-based Parkinson detection outperformed CNN and HMM-based Parkinson detection systems in terms of performance. The ANN classifier model achieved a detection accuracy of 96%. The limitation of this study is comparison across classifiers, while ANN, CNN, and HMM models were evaluated, the study lacked a robust exploration of ensemble methods or modern architectures such as transformers that might improve accuracy.

Asmae et al. [17] proposed ML models to distinguish between PD patients and healthy individuals. The Parkinson's UCI dataset was utilized. Thirteen male and female patients, 23 of whom were diagnosed with PD, contributed 195 sustained vocal phonations to the ML database. The age range of the patients is 46–85 years old. ANN and KNN were the classification models employed. Regarding accuracy, the ANN classifier performed better than the KNN classifier. The detection accuracy of the ANN classifier model was 96.7%. The limitation was overemphasis on ANN and KNN, the study only compared two classifiers (ANN and KNN), limiting the exploration of other ML or DL methods that might perform better.

Nishat et al. [18] proposed various boosting techniques for the early evaluation of PD, to distinguish between healthy and PD, four classification models were used including gradient boosting, light gradient-boosting machine (LGBM), XGBoost, and adaptive boosting (AdaBoost). The UCI ML Repository provided the data. The PD categorization dataset was collected in the faculty of medicine at Istanbul university. 107 men and 81 women, totaling 188 PD patients aged from 33 to 87, were examined. The testing demonstrated that LGBM (Tuned) performed best with accuracy 93.39%, sensitivity 96.6%, precision 95%, and F1-score 95.8%. The limitation was an imbalanced dataset.

Rana, A. et al. [19] illustrated supervised classification algorithms for the early prediction of PD. SVM, NB, KNN, and ANN were the models that were employed. The data, which included recorded speech sounds, were acquired by Max Little from the University of Oxford. The dataset comprised a range of acoustic speech collected from 195 individuals, 147 of whom had PD. Based on the testing data, the ANN was shown to have the maximum accuracy. With a detection accuracy of 96.7%, an F1-score of 87.01%, a Matthews's correlation coefficient (MCC) of 70.11%, a sensitivity of 92.42%, and a specificity of 91.25%, the ANN classifier model performed admirably. The limitation was validation was not applied.

Wroge et al. [20] proposed supervised classification methods for accurate diagnosis of PD. Decision Trees (DT), Extra Trees, Gradient Boosted Decision Trees, ANN, RF, and SVM were the models used for classification. The information utilized came from mPower (Sage Bionetworks' clinical observational study), which collect data and biomarkers from healthy individuals and others with PD. The Gradient Boosted Decision Tree classifier model had 86% detection accuracy, 79% F1-score, 85% precision, and 86% sensitivity. The limitation was an imbalanced dataset.

Lahmiri et al. [21] proposed eight alternative patterns ranking that evaluated when combined with a nonlinear SVM to discriminated between Parkinson's sufferers and control individuals who were in optimal health. The 195 voices from 147 patients with PD and 48 healthy volunteers were recorded. Each vocal phonations were measured using a set of 22 vocal patterns, and the SVM was optimized using the Bayesian optimization (BO) technique, yielding accuracy of 92.13%, sensitivity of 82.79%, and specificity of 95.27%. The limitation was limited in comparison with DL models, focusing on using only one type of ML algorithm (SVM).

Pahuja, G. et al. [6] proposed comparison between classifiers for the early diagnosis of PD. Three algorithms, MLP, SVM, and KNN were employed. The UCI repository provided the dataset. ANN performed as the best classifier using the Levenberg–Marquardt algorithm, which had a geometric mean of 95.16%, sensitivity of 93.75%, specificity of 96.59%, and classification accuracy of 95.89%. The limitations were imbalanced dataset and comparing a few algorithms.

Berus, L. et al. [22] proposed multiple feedforward ANNs with varied topologies for early PD detection. The UCI ML repository provided the dataset. Twenty healthy individuals and twenty PD patients from faculty of Medicine, Istanbul university provided the data. Healthy individuals span in age from 43 to 77, while PD patients range in age from 45 to 83. The best results were obtained by ANN (fine-tuned), which had the highest accuracy (86.47%), sensitivity (88.91%), specificity (84.02%), and MCC (73.21%). The limitation was that the study focused on using only one type of DL algorithm (ANN).

Ahmed et al. [5] proposed six distinct ML algorithms for early PD prediction. Stochastic gradient descent (SGD), DT, RF, KNN, LR and XGB were the models used for classification. The Physio net Gait Analysis database provided the data that were used. It included a variety of biological tone voice evaluations from 31 individuals, 23 of them had PD. According to the findings, RF had the best accuracy (97%), AUC (91%), and F1-score (96%). The limitation was limited comparison with DL models, while traditional ML models were compared, DL techniques with proven efficacy in feature extraction for Parkinson's detection were not extensively explored.

**Table 1** Limitations of the previous studies and how to overcome them in the proposed work

| References | Limitations | Overcomes the limitations of the previous studies |
|---|---|---|
| Govindu et al. [15] | Limited Comparison with DL Models: while traditional ML models were compared, DL techniques with proven efficacy in feature extraction for Parkinson's detection were not extensively explored<br>Imbalanced Dataset: the original dataset had a significant imbalance (109 records of Parkinson's patients vs. 40 healthy individuals), which could bias the models | Explore DL Models: investigate convolutional and recurrent neural networks for feature extraction and classification<br>Balanced Dataset: use more sophisticated balancing techniques such as SMOTE |
| Radha et al. [16] | Comparison Across Classifiers:<br>While ANN, CNN, and HMM models were evaluated, the study lacked a robust exploration of ensemble methods or modern architecture such as transformers that might improve accuracy | Use Advanced ML Models: Experiment with ensemble techniques like RF and XGBoost |
| Asmae et al. [17] | Overemphasis on ANN and KNN:<br>The study only compares two classifiers (ANN and KNN), limiting the exploration of other ML or DL methods that might perform better | Evaluate More Classifiers:<br>Compare the performance of ANN and KNN with other advanced algorithms such as RF, Gradient Boosting Machines (e.g., XGBoost), or DL models like MLP |
| Mirza et al. [18] | Imbalanced Dataset | Balanced Dataset: use more sophisticated balancing techniques such as SMOTE |
| Rana et al., 2022 [19] | Validation is not used | Improved models' accuracy by using validation |
| Wroge et al. [20] | Imbalanced Dataset | Balanced Dataset: use more sophisticated balancing techniques such as SMOTE |
| Shmuel et al. [21] | Limited Comparison with DL Models: focused on using only one type of ML algorithm (SVM) | Explore DL and more ML models |
| Pahuja, G. et al. [6] | Imbalanced Dataset<br>Compared and used a few algorithms | Balanced Dataset: use more sophisticated balancing techniques such as SMOTE<br>Explore ML Models and more DL Models |
| Berus et al. [22] | Focused on using only one type of DL algorithm (ANN) | Explore more ML Models and more DL Models |
| Ahmed et al. [5] | Limited Comparison with DL Models: While traditional ML models were compared, DL techniques with proven efficacy in feature extraction for Parkinson's detection were not extensively explored | Explore DL Models: Investigate convolutional and recurrent neural networks for feature extraction and classification |
| Francisco Santos. [23] | Focused on using only one type of ML algorithm (XGB) | Explore more ML Models and more DL Models |

Francisco [23] proposed XGBoost algorithm for early PD detection. A large amount of patient-specific information, such as audio analysis and vocal frequency measures, and crucial demographic data, were painstakingly added to the collection. The binary "status" feature of the dataset assigns a value of 0 to people without PD and a value of 1 to those who have it. The accuracy of the XGBoost model was 93.33%. The model's sensitivity is 95.65%, its F1-score is 95.65%, and its precision is 91.67%. The limitation was that this study focused on using only one type of ML algorithm (XGB).

Table 1 shows the limitations of the previous studies and how to overcome them in the proposed work.

## 3 Methodology

The procedures, techniques, and algorithms are explained in detail in this section. Figure 4 displays the proposed automated system's block diagram. To detect PD at an early stage, it examines sound waves from a dataset. Examine anomalies in the dataset and replace any missing values to make it better. SMOTE is the approach used to balance the dataset. Min–Max scaling is used for feature normalization. Eight AI classification models were implemented including SVM, KNN, LR, ANN, MLP, RidgeCV, XGB, and NB. The proposed work primarily emphasizes performance comparison, data balancing (using SMOTE), and preprocessing.

The proposed system was implemented using the Python programming language within the Anaconda development environment, which is widely used for AI and data analysis applications. The experiments were conducted on a system

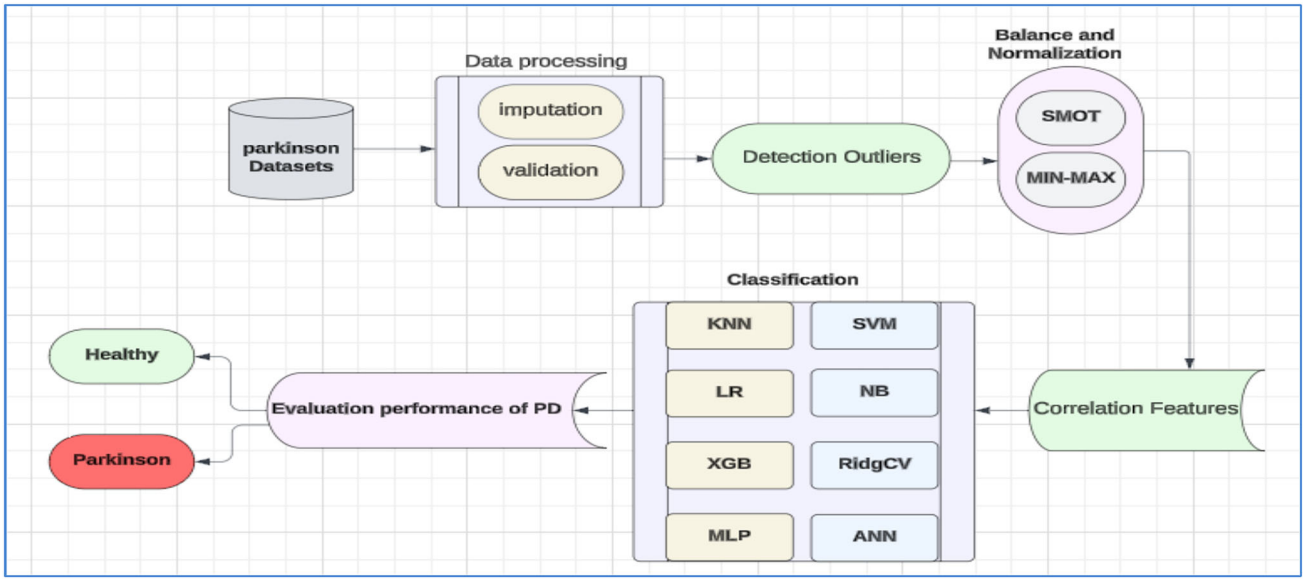


**Fig. 4** Proposed system for PD early prediction

equipped with an Intel Core i7 processor (2.8 GHz) and 8 GB RAM. The implementation relied on several widely adopted Python libraries, including, scikit-learn, XGBoost, imblearn, pandas, numpy, matplotlib and seaborn. Scikit-learn is used for building classification models such as SVM, KNN, LR, and RidgeCV, as well as for computing performance metrics. XGBoost is employed to implement the XGBoost model. Imblearn is utilized for handling class imbalances through SMOTE. Pandas and numpy are used for data manipulation and numerical operations. Matplotlib and seaborn are used for data visualization, including ROC curves and confusion matrices. These tools provided a robust and flexible framework for developing, evaluating, and visualizing machine learning models for early Parkinson's disease prediction based on vocal features.

### 3.1 Data Acquisition

The database used for the proposed system is based on speech recordings from the UCI ML Repository by Max Little from Oxford University, which is publicly available [24]. The voice collected from 31 patients, 23 from them with PD were included in this dataset. A specific voice measure was indicated by each column in the table and each row denoting one of the 195 voice recordings made by these individuals (the "name" column). The "status" column, which is set to 0 for healthy and 1 for PD, was primarily used to distinguish between people with PD and those who were not. Each row in the dataset represented a single voice recording occurrence. Each patient had approximately 6 recordings (some had more, some less). Twenty-three features, each describing the voice measure and its interpretation, were extracted from speech signals and included in the dataset.

### 3.2 Features Description

The dataset consists of 23 features extracted from voice signals, each of which describes the voice measure and its interpretation. Twenty-two of them are regarded as input features, while status (0 for healthy and 1 for PD) is considered an output feature. The input features are shown in Table 2, and each one describes the voice measure and its interpretation.

### 3.3 Data Preparation and Processing

The conversion of unprocessed data into a format that can be used and understood is known as data processing. One of the most important factors in ensuring the success of the next projects is data analytics. Two phases make up data processing: (1) imputation, which replaces missing values, eliminates outliers, and removes duplicate entries; and (2) validation, which confirms consistency and completeness [26]. Fivefold cross-validation ($k = 5$) was applied to mitigate overfitting and ensure robustness. It was applied during the training process to evaluate the performance of each model. This involved dividing the dataset into five parts, training the model on four folds, and validating on the fifth, iterating across all partitions. This approach ensured robust performance evaluation and minimized the bias associated with single train-test splits.

**Table 2** Voice measure interpretation for the PD dataset

| No | Features | Description |
|---|---|---|
| 1 | MDVP: Fo | The average frequency of fundamental speech [5] |
| 2 | MDVP: Fhi (Hz) | The maximum possible vocal basic frequency [5] |
| 3 | MDVP: Flo (Hz) | The minimum possible vocal basic frequency [5] |
| 4 | MDVP: Jitter (%) | Different multi-dimensional voice program (MDVP) metrics from Kay Pentax. MDVP is a conventional method of measuring the frequency of vocal fold vibrations at pitch period to vibrations at the beginning of the following cycle, known as pitch mark [15] |
| 5 | MDVP: Jitter (Abs) | Different multi-dimensional voice program (MDVP) metrics from Kay Pentax. MDVP is a conventional method of measuring the frequency of vocal fold vibrations at pitch period to vibrations at the beginning of the following cycle, known as pitch mark [15] |
| 6 | MDVP: RAP | Different multi-dimensional voice program (MDVP) metrics from Kay Pentax. MDVP is a conventional method of measuring the frequency of vocal fold vibrations at pitch period to vibrations at the beginning of the following cycle, known as pitch mark [15] |
| 7 | MDVP: PPQ | Different multi-dimensional voice program (MDVP) metrics from Kay Pentax. MDVP is a conventional method of measuring the frequency of vocal fold vibrations at pitch period to vibrations at the beginning of the following cycle, known as pitch mark [15] |
| 8 | Jitter: DDP | Different multi-dimensional voice program (MDVP) metrics from Kay Pentax. MDVP is a conventional method of measuring the frequency of vocal fold vibrations at pitch period to vibrations at the beginning of the following cycle, known as pitch mark [15] |
| 9 | MDVP: Shimmer | Shimmer Local amplitude perturbation [17] |
| 10 | MDVP: Shimmer (dB) | Local disturbance of amplitude (decibels) [17] |
| 11 | Shimmer: APQ3 | 3-point Amplitude Perturbation Quotient [17] |
| 12 | Shimmer: APQ5 | 5-point Amplitude Perturbation Quotient [17] |
| 13 | MDVP: APQ | 11-point Amplitude Perturbation Quotient [17] |
| 14 | Shimmer: DDA | The mean absolute difference between the successive periods' amplitudes [17] |
| 15 | NHR | First measurements of the voice's noise to tone component ratio [25] |
| 16 | HNR | Second measurements of the voice's noise to tonal component ratio [25] |
| 17 | RPDE | First measures of nonlinear dynamical complexity [25] |
| 18 | D2 | Second measure of nonlinear dynamical complexity [25] |
| 19 | spread1 | First nonlinear measurement of the underlying frequency fluctuation [25] |
| 20 | spread2 | Second nonlinear indicator of the underlying frequency fluctuation [25] |
| 21 | PPE | The third nonlinear indicator of the underlying frequency fluctuation [25] |
| 22 | DFA | Detrended Fluctuation Analysis measures the extent of stochastic self-similarity of noise in speech signals [15] |

Duplicate values are absent from the dataset since the number of rows in the dataset is equal to the number of unique column values. "Numerical variables" are continuous features, with the exception of the binary categorical "status" feature. Thus, the qualities need to be transformed into an object type. If disparities are identified during the data processing, the suitable action is taken, contingent upon the severity of the abnormalities. It could entail investigating and addressing problems with data validity, handling outliers, inputting missing figures, and deleting duplicates. The intention is to safeguard the data's quality and integrity so that accurate and trustworthy analysis may be carried out.

After analysis, the 195 audio signals in the dataset are split into two imbalanced categories: Parkinson's and healthy. 30.25 percent of this dataset is tested, and the remaining 69.75% is used for training. The dataset contained 48 healthy records (24.62%) and 147 (75.38%) PD records prior to balancing. The dataset has equal records and is balanced during the training phase.

**Table 3** The dataset's skewness value

| Feature No | Input features | Skewness value |
|---|---|---|
| 1 | MDVP: Fo (Hz) | 0.5917 |
| 2 | MDVP: Fhi (Hz) | 2.5421 |
| 3 | MDVP: Flo (Hz) | 1.217 |
| 4 | MDVP: Jitter (%) | 3.0849 |
| 5 | MDVP: Jitter (Abs) | 2.6490 |
| 6 | MDVP: RAP | 3.3607 |
| 7 | MDVP: PPQ | 3.0738 |
| 8 | Jitter: DDP | 3.3620 |
| 9 | MDVP: Shimmer | 1.6664 |
| 10 | MDVP: Shimmer (dB) | 1.999 |
| 11 | Shimmer: APQ3 | 1.5805 |
| 12 | Shimmer: APQ5 | 1.7986 |
| 13 | MDVP: APQ | 2.6180 |
| 14 | Shimmer: DDA | 1.5806 |
| 15 | NHR | 4.2207 |
| 16 | HNR | - 0.5104 |
| 17 | RPDE | - 0.1434 |
| 18 | DFA | - 0.033 |
| 19 | spread1 | 0.4321 |
| 20 | spread2 | 0.1444 |
| 21 | D2 | 0.4303 |
| 22 | PPE | 0.7974 |

Recursive feature elimination (RFE) was employed to reduce dimensionality and improve classification performance.

### 3.3.1 Detecting Outliers

The skewness technique of statistical evaluation is employed to ascertain the dataset's symmetry of distribution. Values for skew features are those that deviate from a median to the left or right. The data are said to be symmetrical when the median, mean and mode all appear at the same location. When the distribution of the tail to the right is longer or fatter, it is said to have positive skewness, indicating that the median and mean are greater than the mode [27]. Table 3 displays the skewness values for each feature.

### 3.3.2 Balance of Dataset

PD dataset has 195 voices that are categorized into two imbalanced classes: Class 1 (which has a rate of 75%) and Class 0 (which is healthy, with a percentage of 25%), Fig.5 illustrates imbalanced dataset. The dataset is therefore out of balance. The minority class will be neglected while the majority class is attended by the diagnostic procedure. As a result, the

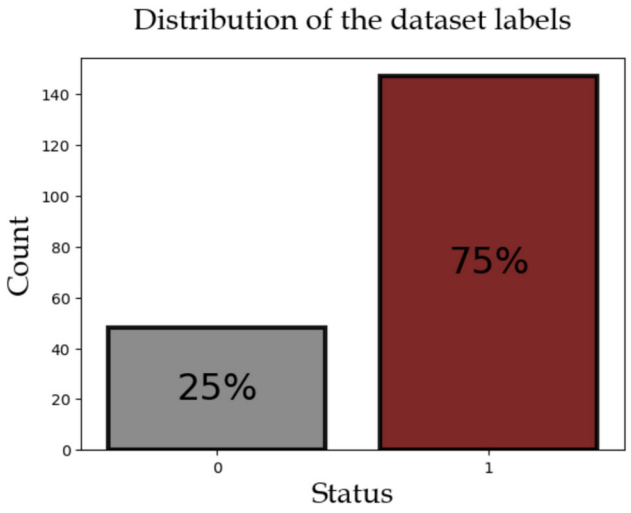


**Fig. 5** Imbalanced dataset

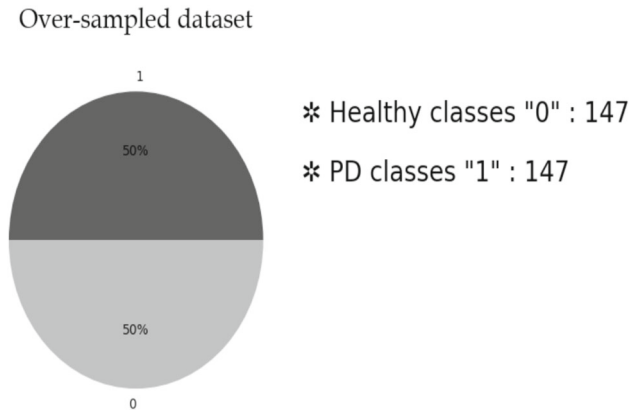


**Fig. 6** The dataset after the use of SMOTE

dataset has to be balanced. The bulk of classes in the dataset could lose essential details, if up sampling approaches are used. The oversampling approach thus gets around this problem. As a result, there are more samples for the minority classes.

SMOTE was suggested and realized to solve this problem, it operates by introducing new samples into minority groups only during the training period [28]. This method looks for minority class samples and determines the closest to every place in order to produce new samples. This technique is repeated until balanced and obtained equal classes [29]. Figure 6 depicts the dataset after the usage of SMOTE. The Healthy classes have nearly equaled the PD classes.

### 3.3.3 Normalizing and Scaling Dataset

To guarantee that the dataset is properly scaled and prepared for the task at hand, painstaking data pretreatment methods

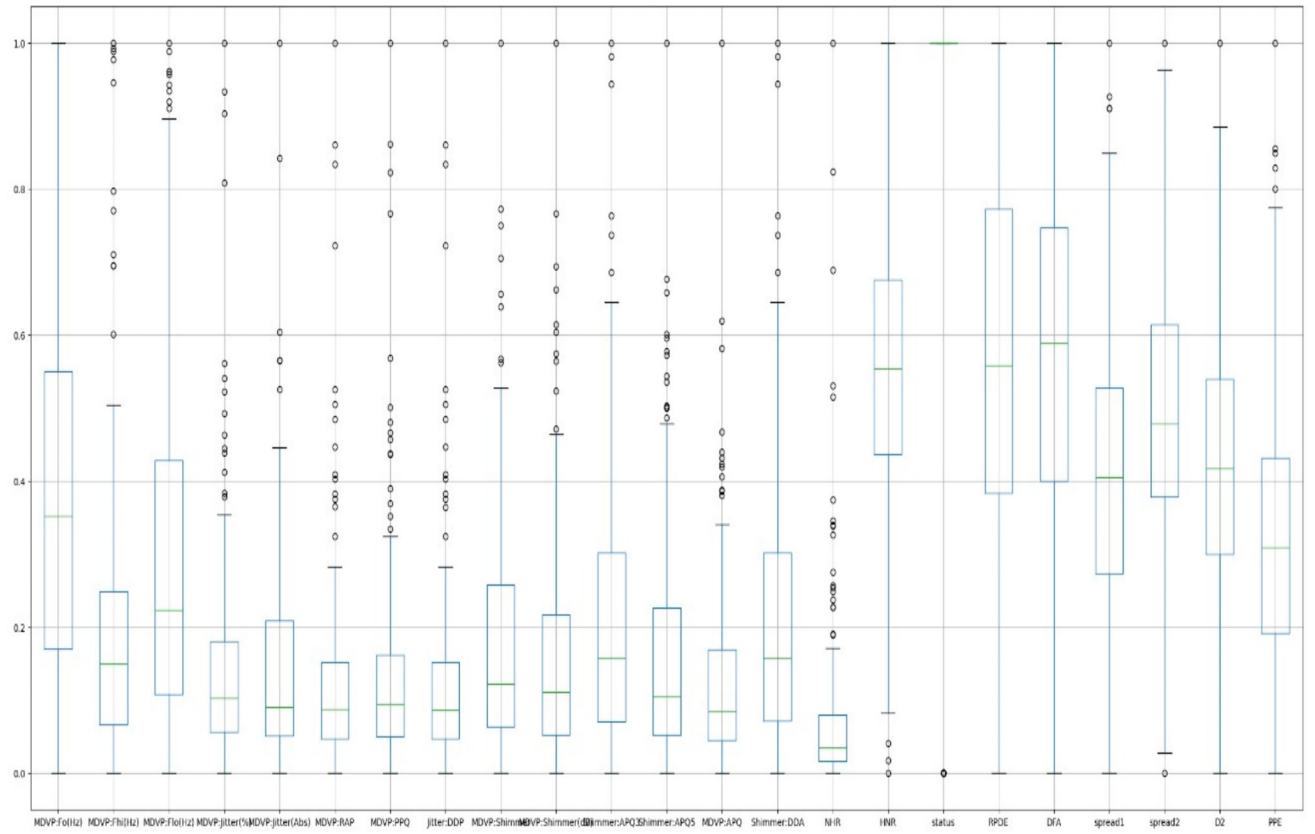


**Fig. 7** Boxplots of the input features that are scaled

are utilized. Feature normalization is performed using Min—Max scaling to align the dataset's varied features within a consistent range. This standardization is critical for accelerating model convergence and optimizing performance [23]. Furthermore, the dataset is meticulously divided into separate training and testing sets. The training set lays the groundwork for the model's learning, whereas the testing set allows for an objective assessment of the model's performance.

Equation (1) illustrates normalizing the balanced data in which (X) is a specific attribute of the dataset that is represented by a column, and (xi) is the value of that column, where (i) is the column's number of components. (Xmin) denotes the column's minimal value, while (Xmax) denotes the column's maximum value. The input features scaled before plotting their boxplots are shown in Fig. 7, where in Fig. 7, X-axis demonstrates the names of the features (MDVP: Fo(Hz), MDVP: Fhi(Hz), MDVP: Flo(Hz), MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP, MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA, NHR, HNR, status, RPDE, DFA, spread1, spread2, D2, PPE), respectively. Y-axis is the scaled values of each feature (ranging from 0 to 1) after Min–Max normalization.

$$\text{Min} - \text{Max} = \frac{xi - x\text{min}}{x\text{max} - x\text{min}} \tag{1}$$

Outliers are data points that go outside the data's interquartile range (IQR) and are sometimes referred to as extreme values. Outliers might be considered as abnormalities or errors, but they can also represent real-world variability and provide useful insights into data. Proposed case and dataset distribution indicate that outliers represent unusual events or extreme cases that should be included in the research. Outliers in finance can reveal crucial information about the data and its underlying distribution [30].

### 3.3.4 Correlation Features

The raw data are analyzed and interpreted using statistical methods. The correlation coefficient, which depicts the relationship between each feature and the next. Correlation coefficient runs from − 1 to + 1, suggesting a positive relationship between two features when the value of one feature rises or falls while the values of the other features rise or fall simultaneously. A negative relationship exists when the value of one attribute increases while the value of another decreases, or vice versa. When one attribute has no effect on the other, the correlation is zero [31]. Equation (2), which

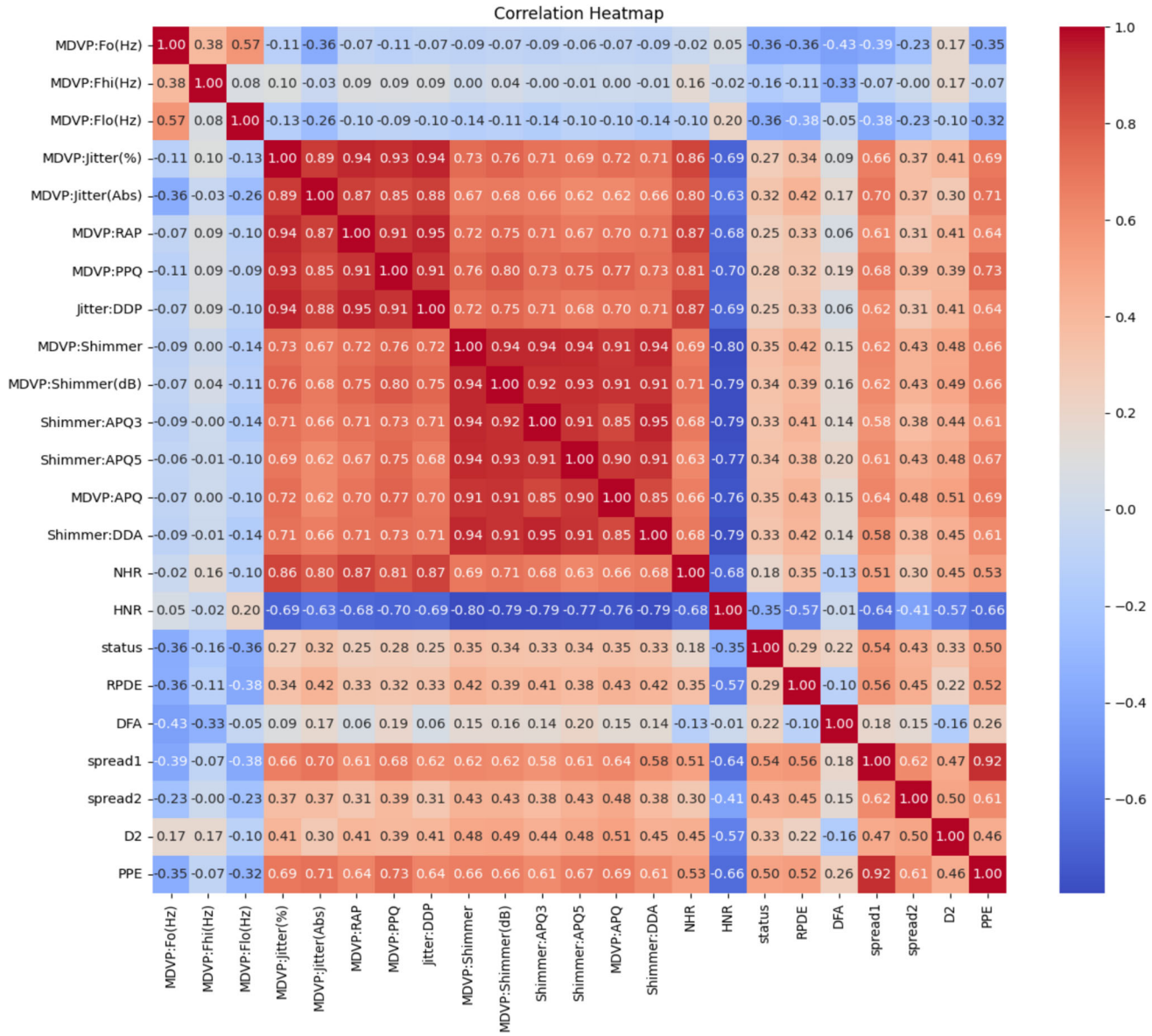


**Fig. 8** Correlation between features

gives the coefficient for adjusting all the attributes, modifies the Pearson's coefficient for analyzing the association between every feature.

$$\rho(\mathrm{X},\ \mathrm{Y}) = \frac{\mathrm{covariance}(X,\ Y)}{\mathrm{Std}(X) \cdot \mathrm{Std}(Y)} \tag{2}$$

For positive and negative correlations, the range of the correlation coefficient is 0.80 to -0.80, correspondingly. Positive feature-to-feature correlations are favorable. Before the dataset can be classified, the highly related features must be removed. The high correlation is due to the dataset's imbalance. To balance the dataset, oversampling was performed. Next, the correlation coefficient between the dataset's features was calculated again. Figure 8 shows a correlation between the dataset's attributes.

### 3.3.5 Evaluation Metrics:

Five statistical criteria were used to evaluate the classification algorithms' performance on the PD dataset: specificity, accuracy, precision, sensitivity, and F1-score. The best method for assessing the effectiveness of categorization models is to use these assessment measures. Equations (3) to (7) show the equations that are used for evaluation [32, 33], where TP is the true positive and TN is the true negative that corresponds to correctly categorized cases, while FP is the false positive

and FN is false positive that indicates wrongly occurrences [34].

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100 \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{4}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{5}$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{6}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \tag{7}$$

where TN indicates correctly classified normal instances, FP indicates correctly classed normal cases as Parkinson's, TP indicates correctly identified Parkinson's instances, and FN indicates correctly classified PD instances as normal. These experimental metrics are all too little to assess students in an unbalanced dataset. For the majority class, accuracy is a deceptive evaluation metric that hardly ever accurately forecasts the parameters of the minority class [35]. In proposed investigations, it is employed the AUC since it is resistant to imbalanced data distributions.

## 3.4 Classification Algorithms

Implementing a model to categorize and distinguish between various kinds of data is the aim of classification. Classifying involves grouping the data into distinct classes based on certain characteristics [36]. Guidance derived from analyzing the "classification model" created in the training set. Then, using test data to evaluate the quality of the produced model based on testing.

The most popular eight categories of classification algorithms in the literature were utilized in this work to aid in the early detection of PD.

### 3.4.1 Support Vector Machine (SVM)

The SVM algorithm identified a decision boundary represented by this hyperplane. The data space is divided into 2 half, healthy and non-healthy cases, by the decision boundary. The distance between the closest data point and the decision border is shown by the geometric mean edge. The geometric edge is positive when the training data can be split linearly, and the resolution limits are separated (hyperplane). It aims to maximize the margin by finding hyperplane [21].

### 3.4.2 K-Nearest Neighbor (KNN)

The KNN technique employed supervised ML that solves regression and classification problems. In KNN, there is no knowledge step. This method looks for similarities between newly collected and pre-existing data pieces. The similarities which are computed using the Euclidean distance show how close or far from the training one that was previously recorded. Stated otherwise, a new data point's association with this class gets stronger the closer it is to a training point [17]. The procedures that used to apply the KNN algorithm are set variable $K$; calculate the distance between the test and training data points; use Euclidean distance, arrange the group in ascending order; and add labels for $K$-values.

### 3.4.3 Extreme Gradient Boosting (XGB)

Gradient-boosting decision trees are used by XGB, a potent ML tool, to aid in decision-making and data comprehension. It has been used by academics and data scientists worldwide to enhance their ML algorithms. is designed for speed, ease of use, and performance on large datasets. It can be used immediately after installation without requiring any additional setup because it does not require tuning or parameter optimization [37].

### 3.4.4 Naive Bayes (NB)

NB classifier is a single technique, it is essentially a family of algorithms founded on the idea that every pair of features that is being classed stands alone. NB is used in supervised learning methods. The method uses posterior probability computing to classify new items [38].

### 3.4.5 Artificial Neural Networks (ANN)

ANNs draw inspiration from the organic neural networks found in human brains. With the neural network serving as a scaffolding, numerous ML algorithms can work together to analyze complex inputs. Usually, this type of system learns by application and does not have task-explicit rules in its programming [25].

### 3.4.6 Multilayer Perceptron (MLP)

An extension of the perceptron designed to handle nonlinear, separable problems is the MLP. The output layer of the perceptron shows the outcomes, while the hidden layers solve the problem and receive input data from the input layer.

The elements comprise the MLP architecture are signals, such text or graphics, are received by the input layer and sent to the subsequent layer; output layer shows diagnostic findings and network replies; hidden layers utilize input data

**Table 4** Parkinson's diagnosis results before and after applying SMOTE technique

| | Model | Accuracy | Precision | Sensitivity | F1-score | Specificity | AUC |
|---|---|---|---|---|---|---|---|
| Before SMOTE | SVM | 90% | 89% | 100% | 94% | 43% | 80% |
| | KNN | 95% | 94% | 100% | 97% | 71% | 98% |
| | ANN | 87% | 89% | 97% | 93% | 43% | 83% |
| | RidgeCV | 92% | 91% | 100% | 96% | 57% | 82% |
| | LR | 90% | 89% | 100% | 94% | 43% | 85% |
| | NB | 72% | 89% | 75% | 81% | 57% | 77% |
| | XGB | 95% | 94% | 100% | 97% | 71% | 94% |
| | MLP | 90% | 89% | 100% | 94% | 43% | 86% |
| After SMOTE | SVM | 95% | 100% | 90% | 95% | 100% | 100% |
| | KNN | 97% | 97% | 97% | 97% | 97% | 100% |
| | ANN | 95% | 100% | 90% | 95% | 100% | 98% |
| | RidgeCV | 90% | 90% | 90% | 90% | 90% | 95% |
| | LR | 85% | 86% | 83% | 84% | 87% | 95% |
| | NB | 86% | 100% | 72% | 84% | 100% | 97% |
| | XGB | 98% | 97% | 100% | 98% | 97% | 100% |
| | MLP | 97% | 100% | 93% | 96% | 100% | 99% |

to address issues. considering the circumstances; data can be transferred only in the forward direction using feedforward networks; and neurons in the subsequent layer that are linked to every other neuron in the layer.

#### 3.4.7 Logistic Regression (LR)

LR is a simple and efficient approach for categorical problems involving binary and linear classification (when the goal is categorical). An LR uses a logistic function to model a binary output variable. It is not required for inputs and outputs to have linear relationships because LR employs a nonlinear log transformation. As a loss function in LR, maximum likelihood estimation (MLE) is a conditional probability function. Probabilities in class 0 are those that are higher than 0.5. If not, class 1 is applied to them [39].

#### 3.4.8 Ridge Classifier with Cross-Validation (RidgeCV)

In large-variable scenarios, RidgeCV is useful since it aims to balance bias and variance while preventing overfitting. It performs well with high-dimensional datasets and is frequently applied to problems requiring accurate predictions and interpretability. Based on the Ridge Regression theory, RidgeCV is an extension of the ridge classifier. It is a linear regression extension that uses cross-validation to determine the optimal regularization value. Using this approach, the dataset is divided into smaller groups, and the model's performance is evaluated using different regularization parameter (alpha) values. Cross-validation, or CV, is a technique that aids in determining which alpha provides the highest generalization performance [40].

## 4 Results and Discussion

Fivefold cross-validation ($k = 5$) was applied during the training process to mitigate overfitting and ensure robustness. Additionally, SMOTE was applied during training to balance the dataset, which enhanced classification performance, particularly for the minority class (healthy individuals).

Specifically, the findings indicate that the SMOTE technique improved the accuracy of the minority class predictions and reduced the bias toward the majority class, resulting in a more accurate and balanced model overall. Table 4 illustrates how the SMOTE technique reduced bias toward the majority class and increased the accuracy of minority class predictions, creating a more accurate and balanced model overall. The results showed that based on the proposed analysis and evaluation, XGB classifier after applying SMOTE outperformed all other models, with 98% accuracy, 97% precision, 100% sensitivity, 98% F1-score, 97% specificity and 100% AUC. Confusion matrix represents actual class instances as rows, while predicted class occurrences are represented as columns. The four possible outcomes: TP, TN, FP and FN. Figure 9 illustrates confusion matrices of proposed models: SVM, KNN, LR, RidgeCV, XGB, NB, ANN and MLP, respectively. Figure 10 illustrates the ROC curves of the proposed models: SVM, KNN, LR, RidgeCV, XGB, NB, ANN and MLP which are 100%, 100%, 95%, 95%, 100%, 97%, 98%, and 99%, respectively.

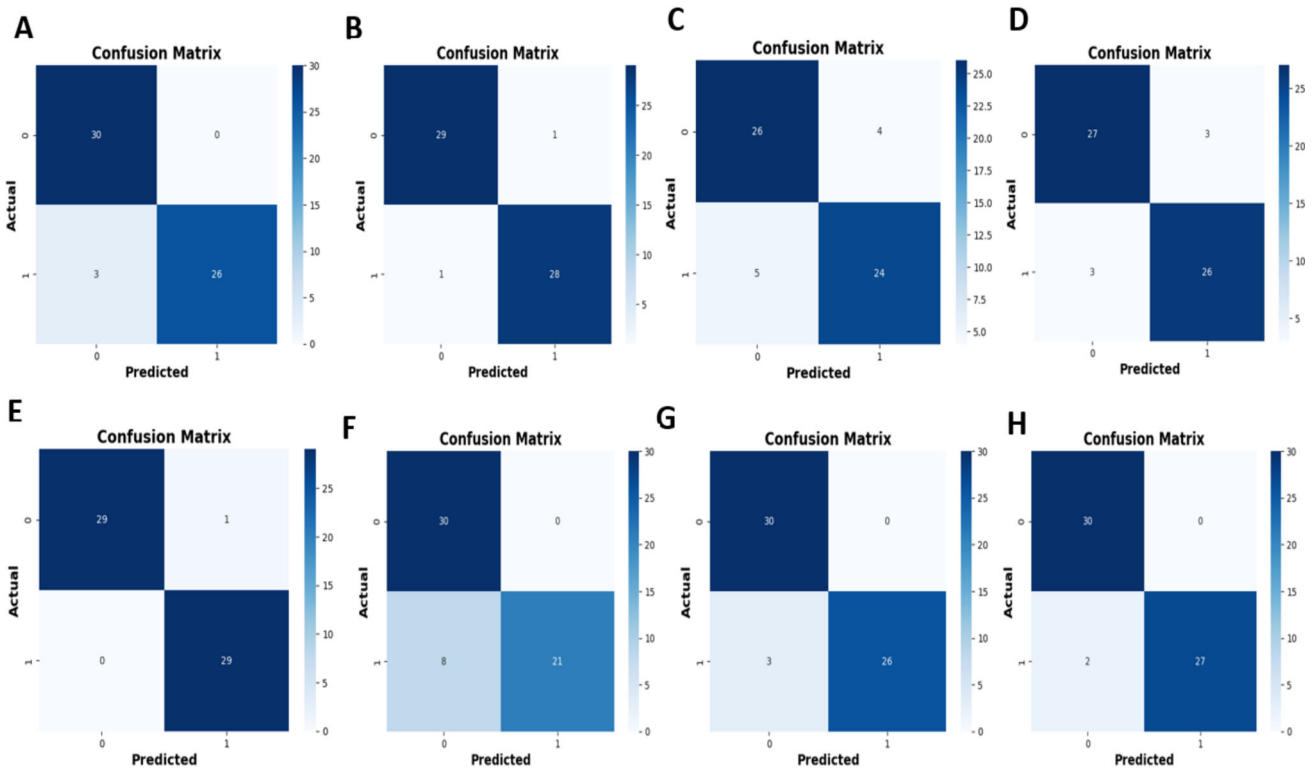


**Fig. 9** Confusion matrices of proposed models: SVM, KNN, LR, RidgeCV, XGB, NB, ANN, and MLP from A to H, respectively

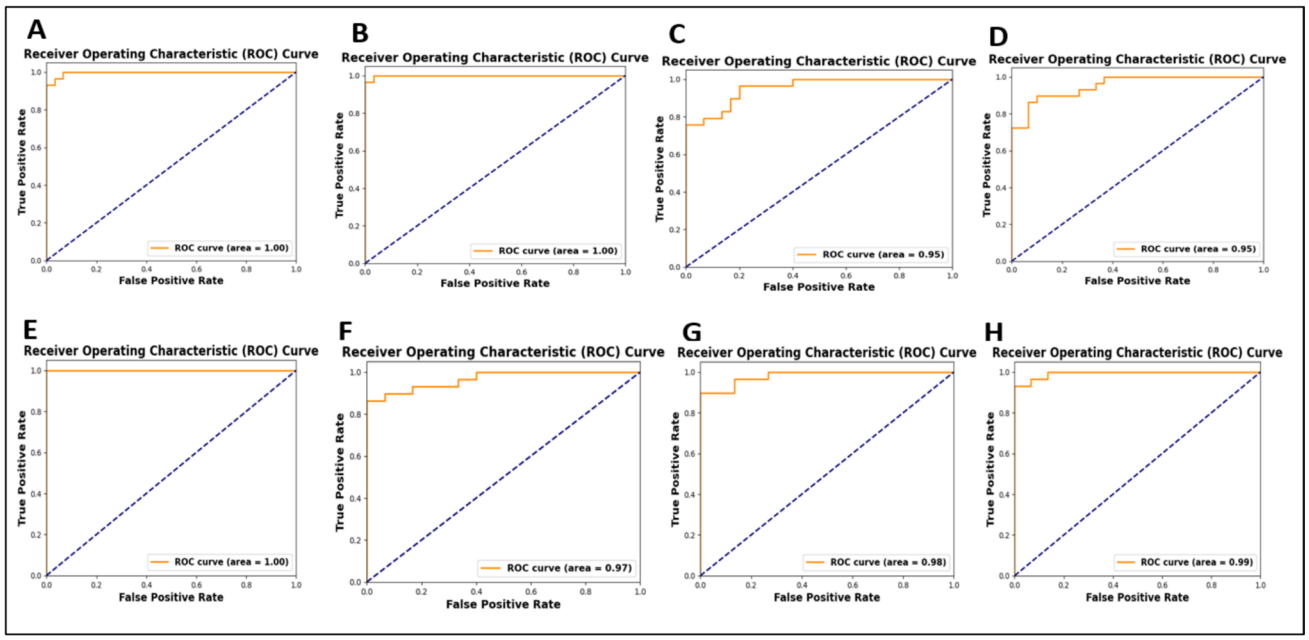


**Fig. 10** ROC curves of proposed models: SVM, KNN, LR, RidgeCV, XGB, NB, ANN, and MLP from A to H, respectively

**Table 5** Comparing the proposed system's effectiveness to the most recent research findings from the literature

| References | Dataset | Methodology | Best model | Performance for the best model |
|---|---|---|---|---|
| Govindu et al. [15] | PPMI and UCI ML | KNN, SVM, RF, LR | RF | Accuracy = 91.83%<br>Precision = 86%<br>Specificity = 95%<br>AUC = 70% |
| Radha et al. [16] | King's College London Hospital in Denmark Hill | ANN, CNN and HMM | ANN | Accuracy = 96% |
| Asmae et al. [17] | UCI ML | KNN and ANN | ANN | Accuracy = 96.7% |
| Mirza et al. [18] | UCI ML | Gradient Boosting, Light GBM, XGB and AdaBoost | Light GBM | Accuracy = 93.39%<br>Precision = 95%<br>Sensitivity = 96.6%<br>F1-score = 95.8% |
| Rana et al. [19] | Max Little of the University of Oxford | SVM, NB, KNN, and ANN | ANN | Accuracy = 96.7%<br>Sensitivity = 92.42%<br>F1-score = 87.01%<br>Specificity = 91.25%<br>MCC = 70.11% |
| Wroge et al. [20] | MPower | DT, Extra Trees, Gradient Boosted Decision Tree, ANN, RF and SVM | Gradient Boosted Decision Tree | Accuracy = 86%<br>Precision = 85%<br>Sensitivity = 86%<br>F1-score = 79% |
| Shmuel et al. [21] | Collected data | SVM + BO | SVM + BO | Accuracy = 92.13%<br>Sensitivity = 82.79%<br>Specificity = 95.27% |
| Pahuja et al. [6] | From UI ML | MLP, ANN, KNN and SVM | ANN with Levenberg–Marquardt | Accuracy = 95.89%<br>Sensitivity = 93.75%<br>Specificity = 96.59%<br>Geometric mean = 95.16% |
| Berus et al. [22] | From UCI ML | ANNs | ANN (fine-tuned) | Accuracy = 86.47%<br>Sensitivity = 88.91%<br>Specificity = 84.02%<br>MCC = 73.21% |
| Ahmed et al. [5] | From the PhysioNet Gait Analysis database | DT, SGD, RF, LR, XGB and KNN | RF | Accuracy = 97%<br>F1-score = 96%<br>AUC = 91% |
| Francisco Santos [23] | Collected dataset | XGB | XGB | Accuracy = 93.33%<br>Precision = 91.67%<br>Sensitivity = 95.65%<br>F1-score = 95.65% |
| Proposed models | UCI ML | SVM, KNN, RidgeCV, LR, NB, XGB, MLP and ANN | XGB | Accuracy = 98%<br>Precision = 97%<br>Sensitivity = 100%<br>F1-score = 98%<br>Specificity = 97%<br>AUC = 100% |

Incorporating feature importance analysis, particularly from the XGBoost model, can provide insight into which vocal biomarkers (e.g., jitter, shimmer, Harmonics-to-noise ratio (HNR)) are most indicative of PD. Such analysis would enhance the explainability of the model and offer valuable guidance for clinicians. This can be achieved using XGBoost's built-in feature_importances_ attribute or SHAP (SHapley Additive exPlanations) values to visualize and quantify the contribution of each acoustic feature to the model's predictions**.**

Table 5 presents a performance comparison between the best proposed classification model and other models that have been explored in the literature. The proposed model outperforms previous research in terms of outcomes; the XGB classifier's accuracy was 98% using the proposed methods. The suggested approach has great efficiency, is dependable, simple to use, takes a lot of time, and can predict PD early.

## 5 Conclusions

PD, which impairs daily activities among older people, is caused by a deficiency of dopamine. This disease is hard to diagnose since its signs include vague and can be confused with other illnesses. A lot of research has been done in medicine and science to identify PD early. AI models that analyze vocal symptoms have helped in early diagnosis. In order to assist in the early diagnosis of PD, the proposed work offers insightful information and a speech dataset with 22 input features and 1 output feature. Since these characteristics seemed to be closely related, a high-level diagnosis was not appropriate for them. The dataset is prepared using SMOTE method and minimum–maximum scaling. There were several classifiers used: SVM, KNN, LR, XGB, NB, RidgeCV, ANN, and MLP. All classifiers achieved better outcomes for both PD diagnosis and normal cases. With 98% accuracy, 97% precision, 100% sensitivity, 98% F1-score, 97% specificity, and 100% AUC, the XGB model utilizing the SMOTE approach produced the best results. Acoustic signals have been demonstrated to be an automatic and early means of detecting PD. It is significant because it led to early diagnosis and treatment plan, therefore it would enhance the standard of living for PD patients. The proposed study advances the understanding of PD detection.

The limitations of the proposed study are that it relies solely on a single dataset and more methods need to be utilized. Also, the proposed study did not address the real-time applicability of the proposed method. The model in this study achieved high classification performance, when trained and evaluated on the UCI ML Parkinson's dataset. However, using a single dataset for both training and testing may not sufficiently demonstrate the model's generalization capabilities. An independent validation on a separate external dataset would offer a more rigorous test of robustness. While the results are promising, future work should focus on validating the proposed models across diverse datasets collected from different populations and clinical settings to ensure their broader applicability. Also, for future work, nested cross-validation is recommended employing or evaluating models on an external independent dataset. Dimensionality reduction is recommended using techniques such as PCA or L1 regularization to reduce model complexity and overfitting risk. Other methods such as RF feature importance, T-tests, and mutual information scores can be utilized for further enhance the identification of the most discriminative vocal features.

**Authors Contributions** Shereen A. Bakry was involved in the conceptualization, formal analysis, methodology, software, validation, and writing—original draft. Nourelhoda M. Mahmoud contributed to the conceptualization, formal analysis, methodology, software, supervision, validation, and writing—review and editing.

**Funding** Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). There is no fund for this work.

**Data Availability** The dataset used in this paper is from the UCI ML Repository by Max Little from Oxford University. This dataset is publicly available from the website: https://archive.ics.uci.edu/dataset/174/parkinsons.

## Declarations

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



## References

1. McDonald, C.; Gordon, G.; Hand, A.; Walker, R.W.; Fisher, J.M.: 200 Years of Parkinson's disease: What have we learnt from James Parkinson? Age Ageing **47**(2), 209–214 (2018). https://doi.org/10.1093/ageing/afx196
2. Jellinger, K.A.: Cell death mechanisms in Parkinson's disease. J. Neural Transm. **107**(1), 1–2 (2000). https://doi.org/10.1007/s007020050001
3. Váradi, C.: Clinical features of Parkinson's disease: the evolution of critical symptoms. Biology **9**(5), 103 (2020). https://doi.org/10.3390/biology9050103

4. World health organization. Assessment and PD burden. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease.(accessed October 1, 2024)
5. Ahmed, I.; Aljahdali, S.; Khan, M.S.; Kaddoura, S.: Classification of parkinson disease based on patient's voice signal using machine learning. Intel. Autom. Soft Comput. **32**(2), 705–722 (2022). https://doi.org/10.32604/iasc.2022.022037
6. Pahuja, G.; Nagabhushan, T.N.: A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE J. Res. **67**(1), 4–14 (2021). https://doi.org/10.1080/03772063.2018.1531730
7. Foltynie, T.; Brayne, C.; Barker, R.A.: The heterogeneity of idiopathic Parkinson's disease. J. Neurol. **249**, 138–145 (2002). https://doi.org/10.1212/WNL.35.4.522
8. Mahesh, T.R.; Vinoth, K.V.; Bhardwaj, R.; Khan, S.B.; Alkhaldi, N.A.; Victor, N.; Verma, A.: An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's disease. Decis. Anal. J. **10**, 100381 (2024). https://doi.org/10.1016/j.dajour.2023.100381
9. Chnachal, M.M.; Vishnu, K.M.: A review on Parkinson's disease diagnosis using machine learning techniques. i-manager's J. Image Process. **10**(1), 21 (2023). https://doi.org/10.26634/jip.10.1.19381
10. Hassanin, D.A.; Khaled Abd-Ellah, M.; Khalaf, A.A.M.; Gharieb, R.R.: Automatic localization of common carotid artery in ultrasound images using deep learning. J. Adv. Eng. Trends **40**, 127 (2021). https://doi.org/10.21608/jaet.2020.41138.1040
11. Aboelmagd, M.R.; Diab, A.A.Z.; Dousoky, G.M.: Failure analysis in photovoltaic power systems using artificial neural network. J. Adv. Eng. Trends **41**, 205 (2022). https://doi.org/10.21608/jaet.2021.49408.1069
12. Alsiddiky, A.; Fouad, H.; Soliman, A.M.; Altinawi, A.; Mahmoud, N.M.: Vertebral tumor detection and segmentation using analytical transform assisted statistical characteristic decomposition model. IEEE **8**, 145278–145289 (2020). https://doi.org/10.1109/ACCESS.2020.3012719
13. Mahmoud, N.M.; Fouad, H.; Soliman, A.M.: Smart healthcare solutions using the internet of medical things for hand gesture recognition system. Complex Intell. Syst **7**, 1253–1264 (2021)
14. Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F.: Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J. Ambient. Intell. Humaniz. Comput. **14**(7), 8459–8486 (2023). https://doi.org/10.1007/s12652-021-03612-z
15. Govindu, A.; Palwe, S.: Early detection of Parkinson's disease using machine learning. Procedia Comput Sci **218**, 249–261 (2023). https://doi.org/10.1016/j.procs.2023.01.007
16. Radha, N.; Sachin Madhavan, R.M.: Parkinson's disease detection using machine learning techniques. Int. J. Early Childhood Special Educ. **30**(2), 543–552 (2021). https://doi.org/10.24205/03276716.2020.4055
17. Asmae, O.; Abdelhadi, R.; Bouchaib, C.; Sara, S.; Tajeddine, K.: Parkinson's disease identification using KNN and ANN algorithms based on voice disorder. IEEE (2020). https://doi.org/10.1109/IRASET48871.2020.9092228
18. Nishat, M.M.; Hasan, T.; Nasrullah, S.M.; Faisal, F.; Asif, M.; and Hoque, M.A.: Detection of Parkinson's disease by employing boosting algorithms. In: 2021 Joint 10th International Conference on Informatics, Electronics and Vision, ICIEV 2021 and 2021 5th International Conference on Imaging, Vision and Pattern Recognition, icIVPR 2021, Institute of Electrical and Electronics Engineers Inc., (2021). https://doi.org/10.1109/ICIEVICIVPR52578.2021.9564108
19. Rana, A.; Dumka, A.; Singh, R.; Rashid, M.; Ahmad, N.; Panda, M.K.: An efficient machine learning approach for diagnosing parkinson's disease by utilizing voice features. Electronics **11**(22), 3782 (2022). https://doi.org/10.3390/electronics11223782
20. Wroge, T.J.; Yasin, Y.; Yasin¨ozkanca, ¨.; Demiroglu C.; Si, D.; Atkins, D.C.; and Ghomi, R.H.: Parkinson's disease diagnosis using machine learning and voice. In: IEEE Signal Processing in Medicine and Biology Symposium (SPMB), (2018). https://doi.org/10.1109/SPMB.2018.8615607
21. Lahmiri, S.; Shmuel, A.: Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed. Signal Process. Control **49**, 427–433 (2019). https://doi.org/10.1016/j.bspc.2018.08.029
22. Berus, L.; Klancnik, S.; Brezocnik, M.; Ficko, M.: Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors **19**(1), 16 (2018). https://doi.org/10.3390/s19010016
23. Francisco Santos D.: Parkinson's disease detection using XGBoost and machine learning, medRxiv, pp. 2010–2023, (2023). https://doi.org/10.1101/2023.10.23.23297369
24. https://archive.ics.uci.edu/dataset/174/parkinsons.(accessed October 1, 2024).
25. Sadek R.M. et al.: ANN for Parkinson's disease prediction. In: 2019. International Journal of Academic Health and Medical Research (IJAHMR), Vol. 3 Issue 1, (2019)
26. Alalayah, K.M.; Senan, E.M.; Atlam, H.F.; Ahmed, I.A.; Shatnawi, H.S.A.: 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 (2023). https://doi.org/10.3390/diagnostics13111924
27. Hatem, G.; Zeidan, J.; Goossens, M.; Moreira, C.: Normality testing methods and the importance of skewness and kurtosis in statistical analysis. BAU J. Sci. Technol. (2022). https://doi.org/10.54729/KTPE9512
28. Tesfahun, A.; Bhaskari, D.L.: Intrusion detection using random forests classifier with SMOTE and feature reduction. IEEE (2013). https://doi.org/10.1109/CUBE.2013.31
29. GhoshRoy, D.; Alvi, P.A.; Santosh, K.: Explainable AI to predict male fertility using extreme gradient boosting algorithm with SMOTE. Electronics **12**(1), 15 (2022). https://doi.org/10.3390/electronics12010015
30. Fitrianto, A.: Comparing outlier detection methods using boxplot generalized extreme studentized deviate and sequential fences. Aceh Int. J. Sci. Technol (2022). https://doi.org/10.13170/aijst.11.1.23809
31. Wotzka, D.; Sikorski, W.; Szymczak, C.: Investigating the capability of PD-type recognition based on UHF signals recorded with different antennas using supervised machine learning. Energies **15**(9), 3167 (2022). https://doi.org/10.3390/en15093167
32. Mahmoud, N.M.; Fouad, H.; Alsadon, O.; Soliman, A.M.: Detecting dental problem related brain disease using intelligent bacterial optimized associative deep neural network. Clust. Comput. **23**, 1647–1657 (2020)
33. Mahmoud, N.M.; Soliman, A.M.: Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. Sci. Rep. (2024). https://doi.org/10.1038/s41598-024-59783-0
34. Mohammed, B.A., et al.: Multi-method analysis of medical records and MRI images for early diagnosis of dementia and Alzheimer's disease based on deep learning and hybrid methods. Electronics **10**(22), 2860 (2021). https://doi.org/10.3390/electronics10222860
35. Zisselman, E.; Adler, A.; Elad, M.: Compressed learning for image classification: a deep neural network approach. In: Processing, analyzing and learning of images, shapes, and forms: Part 1, pp. 3–17. Elsevier, Amsterdam (2018). https://doi.org/10.1016/bs.hna.2018.08.002
36. Senan, E.M.; Abunadi, I.; Jadhav, M.E.; Fati, S.M.: Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms. Comput. Math. Methods Med. **2021**, 1–16 (2021). https://doi.org/10.1155/2021/8500314

37. He, J.; Zhang, J.: Early diagnosis of Parkinson's disease based on XGBoost algorithm and logistic regression. Highlights Sci Eng Technol **107**, 14–22 (2024). https://doi.org/10.54097/r3rttj57
38. Rana, A.; Dumka, A.; Singh, R.; Rashid, M.; Ahmad, N.; Panda, M.K.: An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. Electronics (2022). https://doi.org/10.3390/electronics11223782
39. Hossain, M.A.; Amenta, F.: Machine learning-based classification of Parkinson's disease patients using speech biomarkers. J. Parkinsons Dis. **14**(1), 95–109 (2024). https://doi.org/10.3233/JPD-230002
40. Dissanayake, O.; McPherson, S.E.; Allyndree, J.; Kennedy, E.; Cunningham, P.; and Riaboff, L.: Evaluating ROCKET and Catch22 features for calf behaviour classification from accelerometer data using Machine Learning models, (Apr 2024). arXiv preprint arXiv:2404.18159