IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 March 2024, accepted 10 June 2024, date of publication 1 July 2024, date of current version 1 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3421302

RESEARCH ARTICLE

Design of an Efficient Prediction Model for Early Parkinson's Disease Diagnosis

K. SHYAMALA AND T. M. NAVAMANI

School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India

Corresponding author: T. M. Navamani (navamani.tm@vit.ac.in)

ABSTRACT Parkinson's Disease (PD) is a long-lasting and progressive brain disorder that disrupts the body's nervous system pathways. This disruption leads to various issues with movement and control, leading to various symptoms, including tremors, stiffness, and difficulty with movement and coordination. In the early stages of this condition, the patients struggle to speak and also speak slowly. Dysphonia, a speech impairment or alteration in speech, is experienced by 70 to 90 percent of Parkinson's patients and is an early indication of the disease. Hence, speech can be a vital modality for an early stage of PD diagnosis. In literature, various Machine Learning models are implemented for PD diagnosis based on speech data. However, issues like class imbalance, feature selection, and interpretable prediction analysis are not addressed effectively. Moreover, the accuracy and trustworthiness of the prediction results are essential for providing better healthcare services. Thus, we propose an Interpretable Feature Ranking XGBoost (IFRX) model that effectively addresses the above-mentioned issues. The proposed model has a sequence of processes, such as data preprocessing, class balancing, feature selection, classification, and eXplainable Artificial Intelligence (XAI). We trained the IFRX model based on speech data, which ranks the relevant features, builds an XGBoost classifier and ranked the features according to their relevance in diagnosing PD using Shapley Additive exPlanations (SHAP). Using the proposed model, we implemented eight Machine Learning classifiers for PD diagnosis based on speech data. Among these classifiers, the XGBoost approach shows better prediction performance with an accuracy of 96.61%.

INDEX TERMS Parkinson's disease, feature selection, XGBoost, explainable artificial intelligence, machine learning.

I. INTRODUCTION

A neurodegenerative condition that affects the dopamine-producing cells in a specific area of the brain is called Parkinson's Disease (PD). Following Alzheimer's Disease, PD is the second most widespread neurodegenerative chronic condition [1]. It is a condition that moves slowly over time and is extremely difficult to diagnose at an early stage. Based on the data from the World Health Organisation, around 8.5 million individuals globally suffered from Parkinson's disease in 2019. Despite significant medical advancements, Parkinson's disease remains a global challenge. According to the current estimates, PD caused 329,000 deaths in 2019, an increase of more than 100 percent from 2000 [2]. Specific neurons in the substantia nigra, responsible for dopamine production, become impaired, leading to the characteristic dopamine deficiency in Parkinson's disease. The chemical in charge of regulating movement in the body is dopamine. It is unknown what causes these dopamine-producing neurons to die [3]. Scientists believe that a person's work, family history, brain injury, genetics, and environmental factors are significant for the lack of dopamine deficiency.

Each person with PD experiences the condition differently, and symptoms vary from person to person. Over time, the condition deteriorates despite the first minor symptoms. Both motor and non-motor symptoms fall within this category [4]. Motor symptoms include things like tremors (shaking), slowness when walking, dysphonia (variations in speech or voice), rigidity, and gait (shorter steps or difficulty walking). Non-motor symptoms include things like illusion, memory

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey.

loss, sleep disturbance, constipation, and loss of smell. PD significantly impacts the patient's daily routine; in some instances, the patient cannot shower, dress, or carry out other everyday tasks [5]. Levodopa and syndopa are two medications used to treat PD; the dosage of each drug is based on the severity of the condition [6]. Parkinson's disease can be managed with medicines and exercise, but surgery is recommended for severe motor complaints. Therefore, Early diagnosis is crucial in maximizing a patient's potential for maintaining a good quality of life [7].

Recently, Machine Learning (ML) has become increasingly popular across various industries because of its capacity for pattern detection [8], [9]. Making better judgments is aided by developing decision-support systems based on machine learning [10], [11]. From 2013 to 2023, researchers unveiled many computer-aided PD detection tools, analyzing not just traditional medical scans (Magnetic Resonance Imaging (MRI) [12], Single-Photon-Emission Computed Tomography (SPECT) but also speech patterns [13], handwriting movements [7], Gait signals [14], Electroencephalography (EEG) signals [15], Electromyography (EMG) signals [16] and Freezing of Gait (FoG) signals [17]. Researchers leverage deep learning and machine learning approaches for more accurate PD identification. Seventy to ninety percent of Parkinson's patients experience dysphonia, or speech disorder or change in speech, which is an early sign. Due to their low tone, rapid bursts, shaking voices, and extended pauses, PD sufferers' voices are challenging to comprehend [18]. While striving for increased accuracy, researchers have successfully employed artificial intelligence and data science tools to create a range of speech-based algorithms for PD detection.

As a result of technical improvement, ML models are growing in popularity in the healthcare industry. Early detection of PD plays a crucial role in disease prediction and requires a method that extracts essential features for predicting PD. In literature, ML techniques are used for PD prediction. However, issues like class imbalance, feature selection, and explainable artificial intelligence are not addressed effectively. Inspired by the devastating consequences of misdiagnosis, we set out to design and develop an effective prediction model for early-stage PD diagnosis based on the voice dataset. The proposed work demonstrates how specific features can empower health practitioners to detect PD disease early, significantly improving patient outcomes. Using RFE with XGBoost classifier, our suggested technique adopts a unique strategy. This approach could help with early illness diagnosis and offer a more thorough knowledge of PD prediction.

The significant contributions of this work are as follows:
1. Proposed a novel Interpretable Feature Ranking XGBoost (IFRX) model based on the features of voice, which has higher stability, robustness, and efficiency of achieving the maximum classification performance compared to other machine learning methods.

2. To enhance the prediction accuracy of Parkinson's disease, the speech dataset is balanced using the Synthetic Minority Oversampling Technique with Support Vector Machines (SVMSMOTE) method.

3. Provided better model interpretability and extracted the most critical features for accurate PD prediction using SHAP analysis.

The remaining sections are organized as follows: Section II thoroughly analyzes existing research on PD prediction models using diverse ML methods. Section III delves into the speech dataset used to train and test data, outlining the evaluation metrics chosen to assess its effectiveness. Section IV details the outcomes of feature extraction, feature selection, model training, and evaluation. Section V showcases the outcomes of the IFRX model and the results of the performance analysis. Finally, Section VI wraps up the key findings and suggests the concluding remarks on the effectiveness of our proposed PD prediction model.

II. RELATED WORKS

Several research works have leveraged speech signals for Parkinson's disease diagnosis using various features and techniques. This work explores the current progress in utilizing Recursive Feature Elimination (RFE) to address the challenges in diagnosing PD through speech analysis. It provides useful information on enhanced classification accuracy and interpretability. Machine learning techniques have become beneficial resources for analyzing voice recordings and extracting relevant features for diagnosing and monitoring Parkinson's disease.

Solana-Lavalle et al. [19] proposed a wrapper feature selection technique to reduce the number of selected voice features in PD detection and increase accuracy using a large public dataset. Only four classifiers are used for training, such as Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and SVM-based Multi-Layer Perceptron (MLP). Tuncer et al. [20] proposed a novel approach to detect PD using vowels automatically. A hybrid of the Minimal Average Maximum (MAMa) tree and the Singular Value Decomposition (SVD) technique were applied to generate the pertinent feature subset. While the MAMa tree and SVD are used for feature extraction, other advanced techniques might yield better discriminative features. Moreover, the hybrid model lacks explainability.

Liu et al. [21] suggested an integrated ensemble approach for PD detection that is weighted in Local discriminant Preservation Projection (LPP). The proposed approach preserves the neighborhood structure of PD speech samples while concurrently increasing the inter-class variance and reducing the intra-class variance of the samples in a preferred manner. Unbalanced PD samples were well handled. To circumvent the curse of dimensionality, feature selection should have been used, which can also decrease dimensionality. Sarkar et al. [22] comprehensively evaluated signal processing methods for PD classification based on audio recordings and for feature extraction Tunable Q-factor Wavelet Transform (TQWT) used from patients' voices, which offers superior frequency resolution compared to the traditional discrete wavelet transform. However, validating PD telemedicine systems' effectiveness and potential biases in real-world clinical settings is crucial for reliable assessment. Goyal et al. [18] compared several classification approaches to demonstrate each classifier's capabilities. To decrease the cardinality of the speech dataset without significantly compromising accuracy, three distinct feature importance approaches Genetic Algorithm (GA), Minimum Redundancy Maximum Relevance (mRMR), and Principal Component Analysis (PCA) are also investigated. Two types of datasets were employed. Better accuracy was attained with the XGBoost classifier without the need for a class imbalance technique. The use of speech features in the diagnosis of Parkinson's disease is investigated as a potential use of feature selection techniques by Goyal et al. [23]. The advantages of both GA and Support Vector Machine-Recursive Feature Elimination approaches are combined into a two-stage feature selection process and provide better performance.

Bchir included the integration of relevant feature weighting into the Gaussian mixture model to tackle the problem of high dimensionality. Further, integrating a clustering technique that determines the relevant feature weights with the Gaussian Mixture model classifier can be considered to improve the accuracy [24]. Ashour et al. [25] suggested a new two-step feature selection method using cubic-SVM and weighted hybridization (Eigenvector Centrality Feature Selection (ECFS) and Principal Component Analysis (PCA)) for detecting speech loss in Parkinson's disease. However, this approach makes it difficult to understand how individual features contribute to the final prediction. Therefore, using interpretable models in conjunction with SVM could provide valuable insights into the role of specific features, ultimately leading to a better understanding of the disease.

Polat and Nour [26] introduced a new data sampling technique inspired by the One-Against-All (OGA) approach to classify Parkinson's disease using acoustic features from speech signals. However, the authors have not applied a feature engineering process to select the features. The primary accomplishment of Sowmaya et al. [27] work is the development of an accurate Parkinson's disease prediction system. The authors used an improved technique and evaluated the evolutionary algorithm optimization approach with speech acoustic and decomposition features in addition to the SVM classifier. Hoq et al. [28] combined two models based on a Support Vector Machine (SVM) integrating with a Principal Component Analysis (PCA) and a Sparse Autoencoder (SAE) to identify PD patients based on their vocal features. The PD speech dataset was balanced based on the SMOTE approach and achieved an accuracy of 94.4%. Though PCA and SAE are used in the study to select the features, there exists a lack of effectiveness in identifying the relevant features for Parkinson's disease diagnosis. Chawla et al. [29] proposed a technique based on the Zebra Optimization Algorithm (ZOA) and Recursive Feature Elimination Cross-Validation (RFECV) within Nature Inspired Feature Selection (NIFS) to identify PD. This approach holds promise for diagnosing other diseases, provided the datasets have many attributes. Lamba et al. [13] introduced a novel MIRFE feature selection approach depending on the Mutual Information gain and Recursive Feature Elimination method, which achieved an accuracy of 93.88%. SMOTE is employed as a preprocessing approach for class balancing. However, there are alternative sampling techniques that may be used, to improve accuracy. By choosing the features and fine-tuning the hyperparameters of ML algorithms, Alalayah et al. [30] presented a novel method to optimize the strategies for detecting PD in its early stages using speech. Biases may occur if specific features are inadvertently left out or overrepresented.

It can be inferred from the above research works that many studies have not effectively addressed the issue of class imbalance, which can lead to biased model performance that favors the majority class, thereby compromising the reliability of the diagnosis. Having more features leads to a downturn in the model precision owing to many irrelevant and correlated feature subsets. Thus, the dataset must be optimized to include the most relevant features. It is observed from the existing works that the feature selection technique played a vital role in selecting the most pertinent features, which significantly boosts accuracy for various ML classification algorithms. Various methods like Wrappers, SVD, and mRMR proved effective performance, though the optimal choice depends on the data and the chosen classifier. Popular classifiers like Random Forest, SVM, and KNN perform well, but Wrappers and ensemble learning techniques have achieved the highest accuracy. Existing models often lack transparency and interpretability, which are essential for understanding the decision-making process of machine learning models, especially in medical diagnostics where explainability can significantly impact clinical acceptance and trust. Addressing the selection of the most relevant features, class imbalance, and avoiding overfitting requires careful consideration for predicting the disease at an early stage. To enhance the accuracy of PD disease prediction early, relevant and essential features must be chosen from the available feature set. Here, we propose an Interpretable Feature Ranking XGBoost (IFRX) model to select the essential features from the voice and to fix the class imbalance issue for early diagnosis of PD.

III. DATASET DESCRIPTION

Data collection is the primary step in research, and this task also delves into the pivotal role of the dataset and feature information. The University of California Irvine (UCI) Machine Learning Repository holds a dataset containing voice measurements for studying Parkinson's disease. This data, collected by Max Little and colleagues, includes recordings from 31 individuals, 23 of whom have Parkinson's [31]. Table 1 provides a detailed description of the speech dataset.

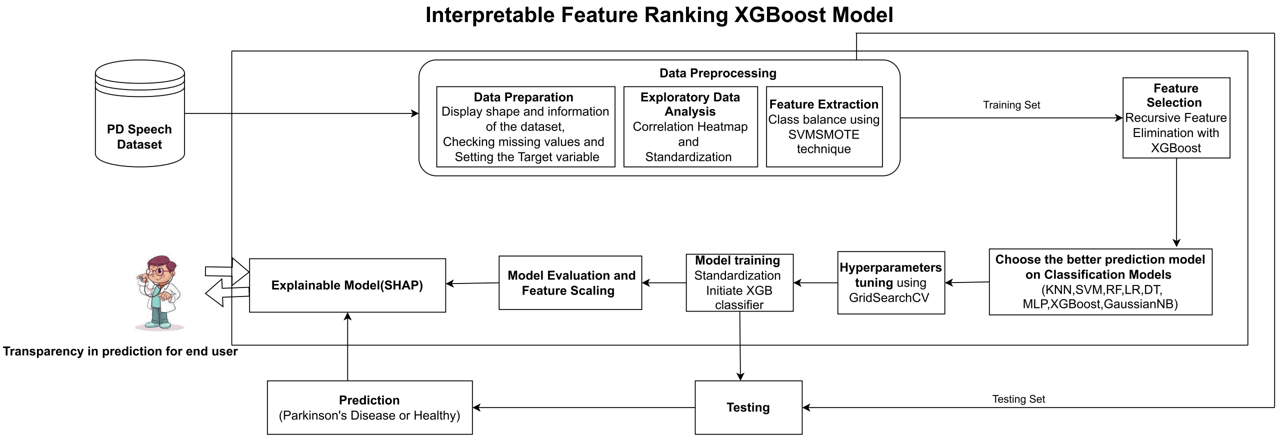

FIGURE 1. Architectural diagram of the proposed system.

TABLE 1. Attributes of feature set.

S No.	Features	Technical Specification	Description	Type
1	Name	Subject Name	ASCII subject name and recording number	object
2	MDVP:Fo(Hz)	Average Vocal Fundamental Frequency	Calculated as the mean of all fundamental frequency (Fo) values in the voice sample	float64
3	MDVP:Fhi(Hz)	Maximum Vocal Fundamental Frequency	The highest Fo value observed in the voice sample	float64
4	MDVP:Flo(Hz)	Minimum Vocal Fundamental Frequency	The lowest Fo value observed in the voice sample	float64
5	MDVP: Jitter (%)	Jitter measures (Several measures of variation in fundamental frequency)	Percentage of the average absolute difference between consecutive periods divided by the average period	flaot64
6	MDVP: Jitter (Abs)	Jitter measures (Several measures of variation in fundamental frequency)	The average absolute difference between consecutive periods	float64
7	MDVP: RAP	Jitter measures (Several measures of variation in fundamental frequency)	MDVP Relative Amplitude Perturbation	int
8	MDVP: PPQ	Jitter measures (Several measures of variation in fundamental frequency)	MDVP Point Perturbation Quotient	flaot64
9	Jitter: DDP	Jitter measures (Several measures of variation in fundamental frequency)	The average difference between jitter cycles	flaot64
10	MDVP-Shimmer	Shimmer measures	Measures of variation in amplitude	flaot64
11	MDVP-Shimmer (dB)	Shimmer measures	Measures of variation in amplitude in decibels	flaot64
12	Shimmer: APQ3	Shimmer measures	Three-point perturbation quotient	flaot64
13	Shimmer: APQ5	Shimmer measures	Five-point perturbation quotient	float64
14	MDVP: APQ11	Shimmer measures	MDVP 11-point perturbation quotient	flaot64
15	Shimmer: DDA	Shimmer measures	Average differences between the amplitudes	float64
16	NHR	Noise-to-Harmonics Ratio	Two measures of the ratio of noise to tonal components in the voice	flaot
17	HNR	Noise-to-Harmonics Ratio	Two measures of the ratio of noise to tonal components	float64
18	RPDE	Nonlinear Dynamical Complexity Measures	Recurrence Period Density Entropy	float64
19	D2	Nonlinear Dynamical Complexity Measures	Correlation dimension	float64
20	DFA	Signal fractal scaling exponent	Detrended Fluctuation Analysis exponent	flaot64
21	spread 1	Nonlinear Fundamental Frequency Variation Measures	Three nonlinear measures of fundamental frequency variation	flaot64
22	spread 2	Nonlinear Fundamental Frequency Variation Measures	Three nonlinear measures of fundamental frequency variation	flaot64
23	PPE	Nonlinear Fundamental Frequency Variation Measures	Pitch Period Entropy	flaot64
24	Status	0: Healthy 1: Parkinson's Disease	Healthy or Not	flaot64

Table 1 is associated with the voice data of individuals. Each row in the table represents a specific feature extracted from the voice recordings, along with its description. The dataset used in the study is to train and test the ML models to distinguish between healthy and Parkinson's disease voices based on these features. We may learn the importance of each characteristic and its role in the speech data analysis from this data. To make better decisions and to conduct a more thorough analysis, the technical specification column delves into the procedures and computations utilized to extract features from the speech data. Feature analysis might reveal specific vocal patterns that strongly correlate with Parkinson's, potentially aiding in diagnosis and monitoring disease progression.

IV. METHODOLOGY

The proposed IFRX model is presented with a sequence of processes such as Data Preparation, Exploratory data analysis, Class balancing, Feature selection, Classification, Hyperparameter Tuning, Model Training, and Explainable Artificial Intelligence.

Figure 1 shows the general architecture of the proposed system. It depicts the process of training the IFRX model to predict PD from speech data. Initially, Preprocessing is done for the input speech dataset. It includes defining a target variable, standardizing, and verifying null values. Secondly, feature extraction is performed to streamline the model by selecting essential features and eliminating those that are insignificant. During model training, a pipeline of classification algorithms is built, data is oversampled to correct for class imbalance and model assessment is assessed using cross-validation. Explainable Artificial Intelligence (XAI) techniques, such as SHAP, demonstrate the importance of global features categorized by class.

A. PRE-PROCESSING

It is started by importing the dataset in the form of numeric data. The dataset is loaded and explored to understand its structure and features. The data preparation step involves checking the data shape and information of the dataset. Figure 2 depicts the histogram representation of features. The diagram shows the distribution of 23 variables. The x-axis of each plot shows the values of the variable and the y-axis shows the frequency of each value. The distribution of each variable is different. Some variables are normally distributed, while others are skewed.

The presence of null values is checked in the dataset and a heatmap to visualize correlations between features,

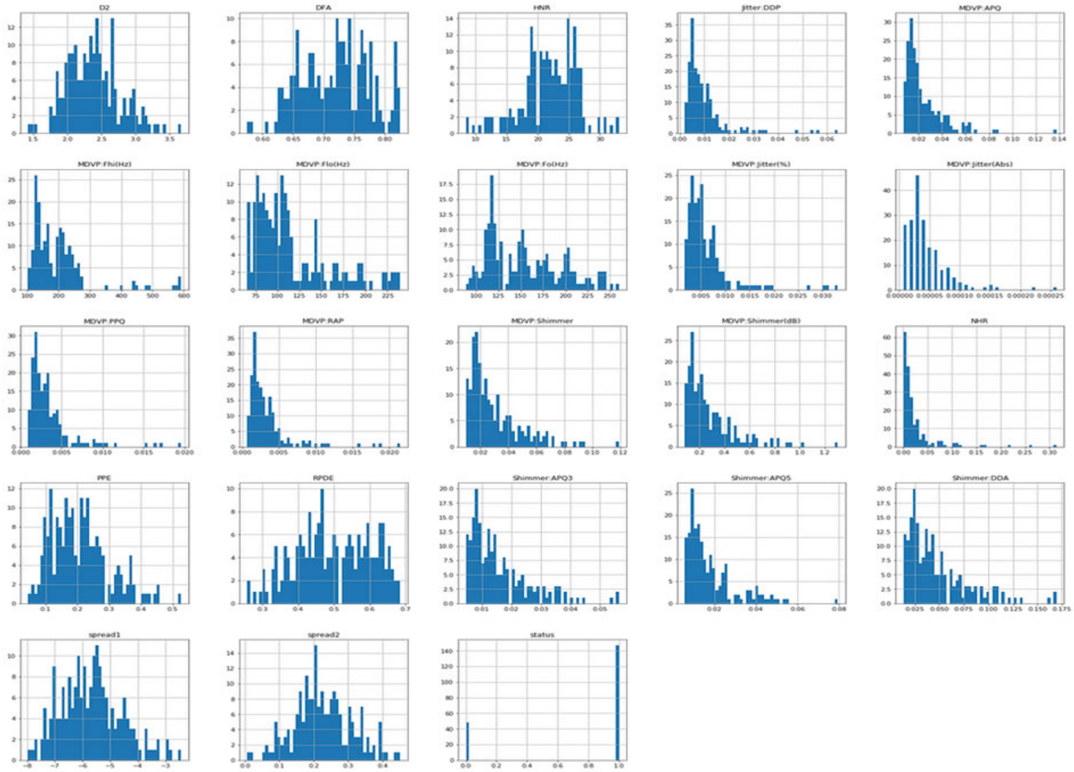

FIGURE 2. **Distribution of features.**

as shown in Figure 3. Exploratory Data Analysis is to understand the data and relationship between variables more deeply and helps in selecting the most relevant features for model training. The associations between various features are evaluated using a correlation heatmap. Standardization is applied to ensure that all features are on a consistent scale and numerical features are scaled with a mean of 0 and a standard deviation of 1. This process improves overall data quality, increases model precision, and eliminates data bias.

The process began by examining the distributions and correlations of individual features. Next, the data was standardized and balanced to resolve class imbalance problems.

B. FEATURE EXTRACTION

Feature extraction is the most important component of preprocessing since it reduces the dimensionality of the data, improving model performance using pertinent data and enabling the model to represent underlying patterns more accurately. It is essential in tasks involving class balancing since it converts the original feature data into a new set of features that are more distinct and informative. This transition can aid in alleviating the impact of class imbalance by enhancing the distinctiveness of the minority class relative to the dominant class. Models trained on imbalanced data tend to overfit the majority class and may not generalize well to new data. Class imbalance is addressed using the SVMSMOTE technique shown in Figure 4, enhancing the model's ability to learn from the minority class instances. It increases the size of the minority class without introducing bias using Support Vector Machines to determine the decision boundaries inside the class and then creating synthetic data points within those limits. Moreover, SVMSMOTE increases the accuracy of machine learning models, hence enhancing their performance.

SVMSMOTE helps to ensure a more balanced representation of the minority class (Parkinson's Disease patients) in the dataset. The purpose of SVMSMOTE is to make sure that the features can be compared and interpreted. Following the class balancing, data was partitioned and standardized. Here, 70-30% split up is applied to divide the data into a training set and a testing set.

C. FEATURE SELECTION

After extracting the features, the relevant features have to be chosen using a suitable feature selection method. Feature selection is the process of choosing the most essential features from the available feature set. It is advantageous because it uses fewer subsets of features, which lowers computing time and boosts PD prediction performance [32]. In this work, we introduce Recursive Feature Elimination with the XGBoost classifier (RFE with XGBoost) technique to select the 15 most important features from the voice dataset of PD patients and perform a grid search to find better hyperparameters for the XGBoost classifiers.

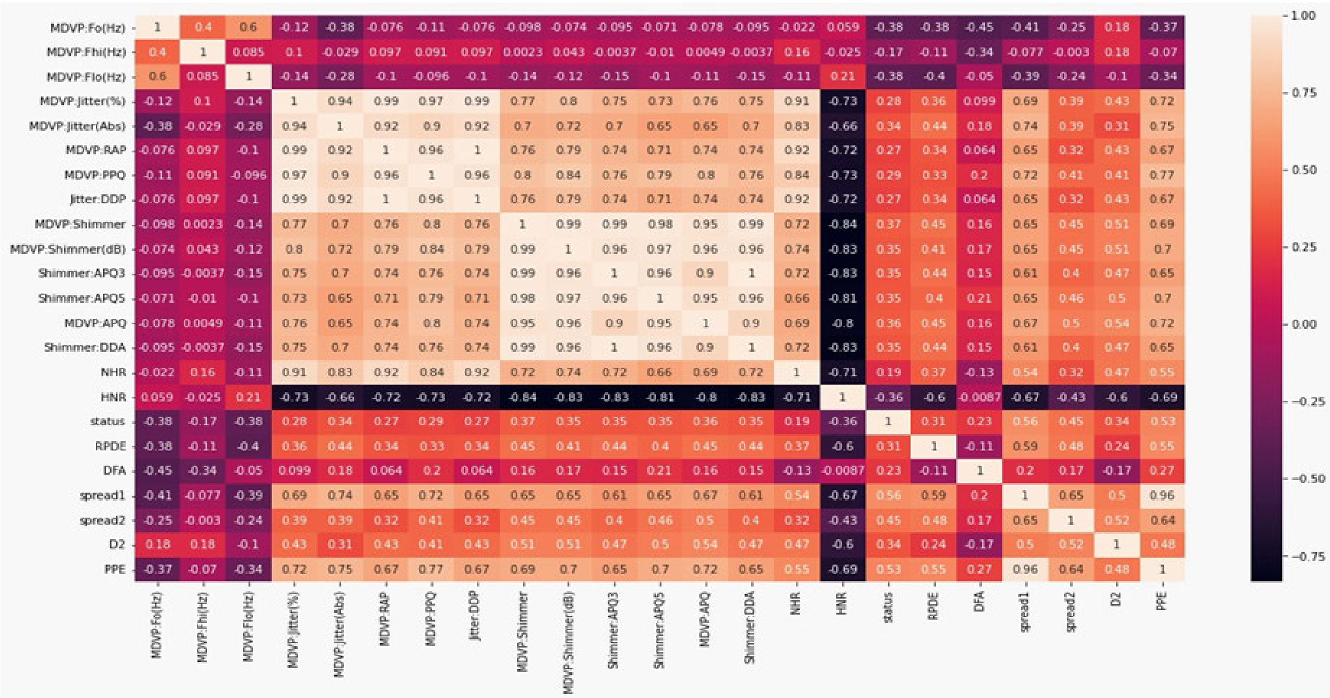

FIGURE 3. Correlation between features.

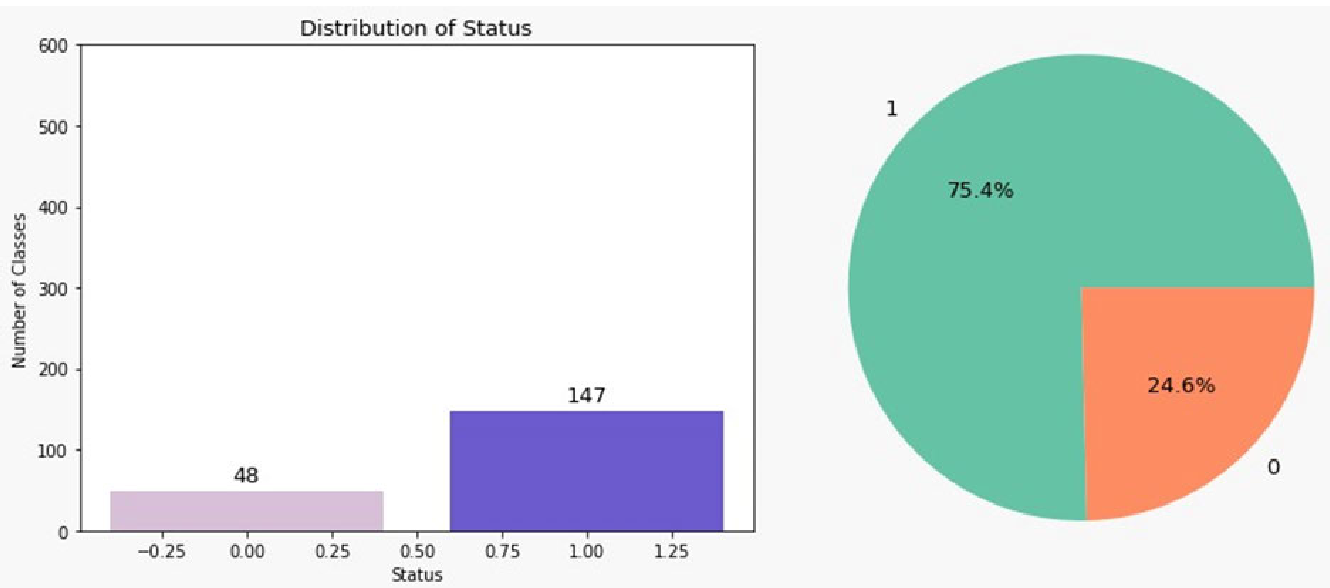

FIGURE 4. Representation of class imbalance of the dataset.

1) RECURSIVE FEATURE ELIMINATION (RFE) WITH XGBoost CLASSIFIER

RFE involves iteratively training a model with all the features and eliminating the feature with the least significant score [33] to acquire a new set of eligible features. The feature significance is determined using the XGBoost Classifier [34], [35]. The XGBoost Classifier is an ensemble algorithm that maximizes model performance using gradient boosting utilizing tree-based techniques and can be computed as follows. The objective function L_t in XGBoost is defined as:

$$L_t = \sum_{i=1}^{J} \left[G_{tj}\omega_{tj} + \frac{1}{2}\left(H_{tj} + \lambda\right)\omega_{tj}^2 \right] + \gamma J \qquad (1)$$

where G_{tj} and H_{tj} are the sums of the first-order and second-order derivatives of all input samples for the i^{th} decision tree

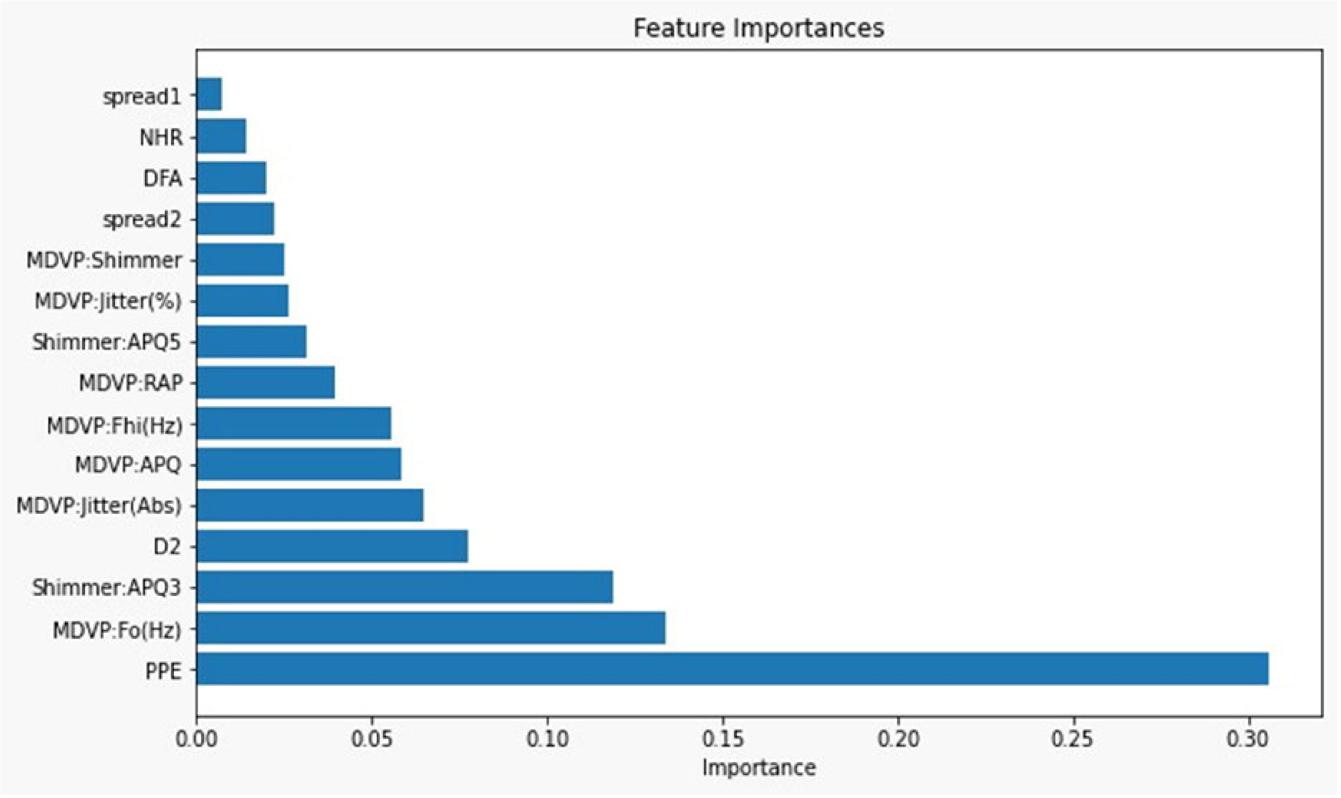

FIGURE 5. Feature ranking based on RFE with XGBoost feature selection.

mapping to the leaf node j, respectively. They are calculated as follows:

$$G_{tj} = \sum_{x_i \in R_{tj}} g_{ti}, \quad H_{tj} = \sum_{x_i \in R_{tj}} h_{ti} \tag{2}$$

where,

- L_t represents the objective function for the tth iteration.
- x_i denotes the input samples.
- G_{tj} is the sum of the first-order derivatives of the loss function with respect to the predictions.
- H_{tj} is the sum of the second-order derivatives of the loss function with respect to the predictions.
- J is the number of leaf nodes.
- γ is a regularization parameter.
- ω_{tj} represents the weight of leaf j.
- λ is the regularization factor.
- g_{ti} and h_{ti} are the first-order and second-order derivatives for the i^{th} sample at the t^{th} weak learner.

The XGBoost method determines the relevance of a feature by counting the instances in which it is used to split the data among all the trees in the model. Next, each feature's relevance is standardized, the total feature importance equals one. As a result, all pertinent features have been found, and the dataset's minor significant features may be eliminated to streamline the model, boost efficiency, and lower its computational complexity.

TABLE 2. Extracted feature importance using RFE with XGBoost classifier.

Index value	Features	Importance
14	PPE	0.305839
0	MDVP:Fo(Hz)	0.133905
6	Shimmer:APQ3	0.118791
13	D2	0.077667
3	MDVP: Jitter (Abs)	0.064809
8	MDVP: APQ	0.058371
1	MDVP: Fhi(Hz)	0.055352
4	MDVP: RAP	0.039256
7	Shimmer: APQ5	0.031283
2	MDVP:Jitter(%)	0.026412
5	MDVP-Shimmer)	0.025119
12	Spread2	0.022268
10	DFA	0.019830
9	NHR	0.013971
11	Spread1	0.007127

Feature Importance measures how much a particular feature contributes to the model's overall performance. In this case, the higher the bar Pitch Period Entropy (PPE), the more influential the feature is to the model's predictions. Feature significance is presented in Figure 5 which shows each feature's importance in an ML model. PPE is the most essential component, followed by MDVP: Fo(Hz) and Shimmer: APQ3. Spread1 is the least significant feature. The patient's voice is one of this model's most crucial components. This implies that the PD can be predicted efficiently by the model using these factors.

Algorithm 1 Mathematical Analysis of Interpretable Feature Ranking XGBoost Model

Input dataset: $D \in \mathbb{R}^{n\times m}$, D is a matrix with dimensions nm, where n represents the number of samples (rows) and m represents the number of features(columns)
Target variable: $y \in \{0, 1\}^n$, y is a vector representing the PD status for each sample.
Visual analysis of feature distributions: $P(x_i)$, for $i = 1, \ldots, m$
Visual analysis of feature correlations: $\mathrm{corr}(x_i, x_j)$, for $i, j = 1, \ldots, m$
Standardization: $z_i = \frac{x_i-\mu_i}{\sigma_i}$, where μ_i and σ_i are mean and standard deviation of feature x_i
Class imbalance addressing: output of class balance is $D_{\text{resampled}} = \text{SVMSMOTE}(D_{\text{std}}, y)$ where SVMSMOTE method applied to the standardized dataset $D_{resampled}$ and the target variable y.
Recursive Feature Elimination (RFE) with XGBoost algorithm:
Define classifier: $h(x) \in \{0, 1\}$, a binary classifier
Initialize: Set $S = \{1, 2, \ldots, m\}$ (all features)
Set k = desired number of features to select
while $|S| > k$ **do**
Train XGBoost model: $h_S = \text{train}(XGBoost, D_S, y)$ on features in S where h_S represents the XGBoost model trained on the selected features S and D_s represents the dataset with only the features in S.
Compute feature importance scores: $\phi_i = \text{feature\_importance}(h_S, i)$ for $i \in S$
Remove feature with lowest importance: $S = S \setminus \{\arg\min_i \phi_i\}$
end while
Output: Selected features S
Model training:
for $k = 1$ to K **do**
Train classifier: $h_k = \text{train}(M_k, D_{\text{resampled}})$, where h_k represents the classifier trained using an ML algorithms and M_k denotes different classifiers such as KNN, SVM, etc.
end for
Train classifiers: KNN, SVM, Random Forest, Logistic Regression, Decision Tree, MLP, Gaussian Naive Bayes
Hyperparameter tuning: Grid search $\theta_k^* = \arg\max_{\theta_k} \text{evaluate}(\text{train}(M_k(\theta_k), D_{\text{resampled}}))$ where θ_k^* Hyperparameter vector for model M_k
Model explainability:
Feature importance(SHAP explainer object): $\Phi = \text{SHAP}(h, D_{\text{resampled}}.X)$, SHAP values calculated for model h on dataset $D_{\text{resampled}}$
Output: Contribution of individual features to model prediction and transparency for decision-making

Algorithm 1 shows the mathematical analysis of the Interpretable Feature Ranking XGBoost (IFRX) model which influences feature importance to rank the features of the voice dataset according to their importance, providing an interpretable way to understand which features are most influential or informative for the model's predictions. Feature importance is concentrating on the most crucial features to enhance the model performance and to reduce its complexity. The RFE feature selection technique relies heavily on feature PPE, the most vital feature in the model, with a ranking of 14 for PPE, as shown in Table 2 also, it represents the first column as the index of the feature, which is a number that uniquely identifies the feature in the model; the second column is the name of the feature; and the third column is the importance of the feature. PPE, a feature that measures the patient's pulse pressure, is the most crucial component to detect PD. MDVP:Fo(Hz), a measurement of the patient's vocal fold vibration frequency, comes next. The patient's vocal fold shimmer, measured by Shimmer:APQ3, is the third most significant feature. Spread1, a measurement of the vocal fold spread of the patient, is the least significant feature. NHR, a measurement of the patient's vocal fold noise-to-harmonic ratio, comes next. The patient's vocal fold jitter, or DFA, is measured as the third least significant parameter.

The importance of MDVP: Fo(Hz) and Shimmer: APQ3 has decreased compared to the PPE. Vocal features and other features remain less critical. Their importance scores are still relatively low, suggesting these features play a minor role in the model's predictions. Figure 6 shows a method for selecting important features from a dataset using RFE with XGBoost model. The RFE with the XGBoost technique specifies the number of features to choose and initializes a counter variable. The data is used to train the XGBoost model, with feature significance calculated after training. The least significant feature is eliminated, and the process is repeated with updated counter variables. The output is a list of chosen features, with the XGBoost model.

D. TIME COMPLEXITY

The time complexity of the IFRX model is computed as follows:

- Visual analysis of distributions and Standardization: $O(n \cdot m)$ where n represents the number of samples and m represents the number of features.
- Visual analysis of correlations: $O(m^2)$ where m represents the number of features.
- Class imbalance addressing (using SVMSMOTE): $O(n)$ where n is the number of samples.
- RFE with XGBoost algorithm: $O(k \cdot f \cdot n \cdot \log(n))$ where k is the number of iterations in RFE, f is the number of features, and n is the number of samples.

Therefore, summarizing the time complexity of the IFRX model can be expressed as:

$$O(n \cdot m + m^2 + n + k \cdot f \cdot n \cdot \log(n))$$

The overall time complexity of the IFRX model is considered moderately complex. It includes linear, quadratic, and potentially logarithmic factors depending on the parameters k, f, n, and m.

It is observed that Pitch Period Entropy (PPE) measures the variability in the time intervals between consecutive glottal pulses in the speech dataset and captures the irregularity of the PD patient's voice pitch. Selecting the features using RFE with XGBoost shows that PPE is a useful feature for PD prediction at an early stage due to the changes in voice features associated with the disease. PPE may even be helpful in early-stage PD prediction, potentially allowing for earlier intervention and treatment.

E. MODEL TRAINING AND EVALUATION

The Training and Evaluation in ML is crucial in building and fine-tuning models to perform tasks effectively. The training set, which includes 70% of the data, is used to train the model. The most crucial features that enhanced the model's overall performance and data quality were selected, and after assessing each feature's significance, the remaining features were eliminated from the dataset. To deliver an accurate prediction and its percentage, we employ a pipeline of ML models to determine which model is most suited for training with high precision. This work tried to solve the oversampling issue, class imbalance, in the suggested model. Recursive Feature Elimination (RFE) selects the most influential features. XGBoost, a robust gradient-boosting algorithm, is chosen as the primary classifier for its effectiveness in handling complex relationships in the data and enhancing model interpretability. Multiple classifiers, including K-Nearest Neighbors, Logistic Regression, Random Forest, Support Vector Machines, Decision Tree, MLP, and Gaussian Naive Bayes, are evaluated using a comprehensive model evaluation function.

The evaluation process involves validating each ML model using the test set that comprises 30% of the data, comparing the outcomes, and ultimately choosing the best model for better prediction. Performance measures like accuracy, recall, precision, and F1 score are applied to compare the models. The proposed technique illustrates the model's ability to distinguish between positive (PD) and negative (Healthy)

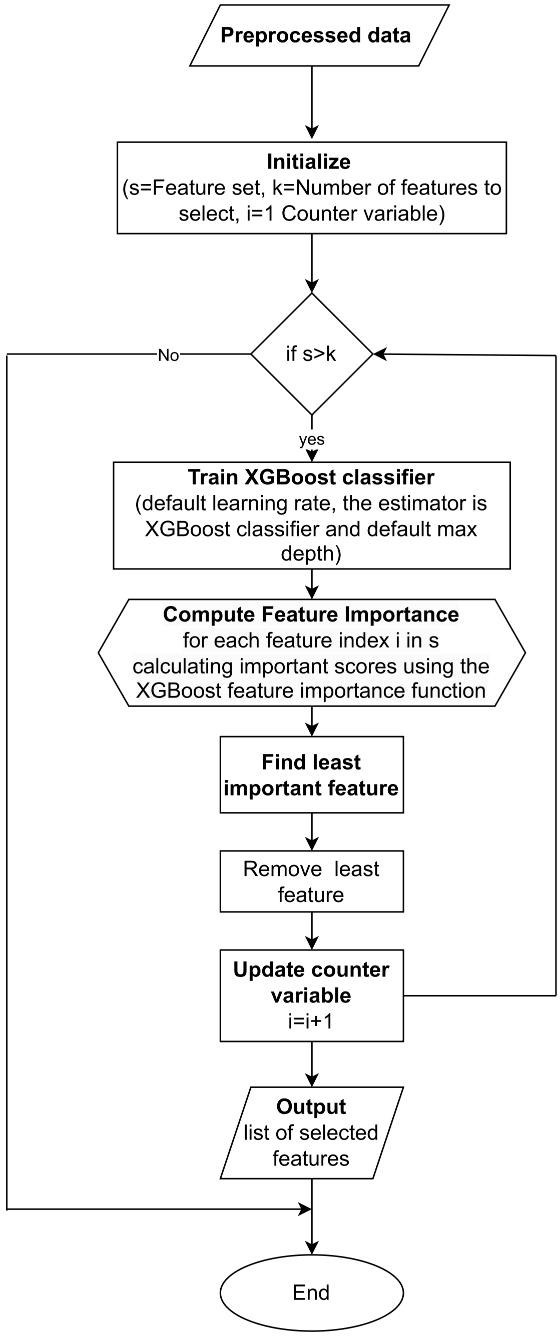

FIGURE 6. Recursive feature elimination with eXtreme gradient boost.

classes, as shown in Algorithm 1. Grid search is used to find the optimal hyperparameters for XGBoost, enhancing the predictive performance. The SHAP (SHapley Additive exPlanations) library creates model explanations, providing insights into the importance of features and their impact on predictions.

F. MACHINE LEARNING CLASSIFIERS

The classification algorithm aims to create a model that can characterize and discriminate between data classes and then utilize that model to predict which class an unclassified item will fall into. The classification process enables data to be categorized into predefined classes according to their features [36]. In this work, we implemented eight ML classifiers such as KNN [37], SVM [38], RF [39], LR [40], DT [41], MLP [42],XGBoost [43], and Gaussian NB [44] for the PD diagnosis based on speech data. To achieve optimal performance, hyperparameter tuning can be applied to the classifiers.

G. EVALUATION METRICS

The values of different available metrics determine a classification model's effectiveness. Metrics for measuring the performance aid in assessing and evaluating the model's effectiveness. These aid in contrasting the other classification models and identifying the optimal model concerning each parameter. Confusion matrix, recall, precision, F1-Score, and accuracy are frequently utilized performance measures [45]. Using the Confusion Matrix is a straightforward method that makes it easy to assess if a model is accurate. Accuracy, precision, recall, and F1 score statistical measures were used to evaluate the performance of classification algorithms applied to the PD speech dataset. The procedure for computing classified instances is outlined in equations (3) to (6), whereby successfully classified examples are represented by TP (True Positive) and TN (True Negative). In contrast, incorrectly categorized instances are represented by FP (False Positive) and FN (False Negative).

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

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

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

H. EXPLAINABLE AI

Explainable Artificial Intelligence (XAI) techniques like SHAP and Local Interpretable Model-Agnostic Explanation (LIME) explain machine learning model predictions. SHAP relies on a concept from game theory called Shapley values. These values help distribute credit (or blame) among players in a collaborative game. In the context of machine learning, the "players" are the different features used by the model, and the "game" is the prediction process. SHAP refers to SHapley Additive exPlanations. It's a technique used to explain the inner workings of any machine learning model, particularly how it arrives at predictions. Cooperative game theory provides a method for calculating the Shapley value of the model, which may be determined as follows:

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S) \right] \tag{7}$$

where,

- ϕ_i: Shapley value for feature i.
- S: Any subset of $F \setminus \{i\}$, where F is the set of all features.
- $|S|$: Cardinality (number of elements) of subset S.
- $|F|$: Total number of features.
- $f_{S\cup\{i\}}(x_{S\cup\{i\}})$: The output of the model when feature i is added to subset S.
- $f_S(x_S)$: The output of the model with subset S of features.

With SHAP, elucidating the importance of a feature in prediction is popular. Based on how each feature affects the performance, SHAP assigns a value to each in the prediction and shows the global feature importance by the class. SHAP makes it possible to comprehend how each feature affects the model's prediction more thoroughly and makes it easier to determine which features are most important for better prediction. SHAP analysis allows health practitioners to identify the most influential features for making predictions and their potential interactions.

The main contribution of the proposed work is the design of the Interpretable Feature Ranking XGBoost (IFRX) model which integrates preprocessing, feature selection, model training and evaluation strategies, and model interpretability methods to create a comprehensive pipeline for Parkinson's disease detection. Advanced techniques such as SVMSMOTE and SHAP enhance the algorithm's robustness by providing valuable insights into the decision-making process of the models and provide a systematic approach for comparing the effectiveness of different classifiers in Parkinson's disease detection.

V. RESULTS AND DISCUSSION

In this section, we outlined and discussed the experimental results, addressing the outcomes of feature extraction, feature selection, oversampling, and model evaluation.

We selected RFE with an XGBoost classifier for feature ranking and eXplainable Artificial Intelligence for model explainability. Overall, the XGBoost classifier model achieved the highest performance across all metrics, with a cross-validation of 94.24, training accuracy of 100%, testing accuracy of 96.61, precision score of 97.73, recall score of 97.73, and F1-score of 97.73. Performance comparison is done with the eight ML techniques and it is observed that the XGBoost classifier performs better for PD prediction based on the speech dataset, as shown in Table 3 and Figure 7.

The other models also performed well, with Random Forest and Decision tree classifiers achieving cross-validation scores of 93.76% and 92.76%, respectively. The RF and DT models might potentially be suitable options for predicting PD. Logistic regression, MLP, and GaussianNB underperformed compared to the other models and may not be appropriate for predicting Parkinson's disease using speech data.

A confusion matrix was applied to calculate the performance measures in terms of accuracy, precision, recall, and F1 score. The confusion matrix shows the True Positive

TABLE 3. Performance comparison of ML techniques in IFRX model.

ML Model	Cross Validation Score	Training Accuracy	Testing Accuracy	Precision Score	Recall Score	F1 Score
K neighbors Classifiers	91.26	94.66	88.14	100	84.09	91.36
Support Vector Classifier	89.4	90.29	88.14	93.02	90.91	91.95
Random Forest Classifier	93.76	100	91.53	93.33	95.45	94.38
Logistic Regression	79.96	80.1	84.75	90.7	88.64	89.66
Decision Tree Classifier	92.76	100	91.53	97.56	90.91	94.12
MLP Classifier	76.83	80.58	74.58	87.18	77.27	81.93
Proposed Classifier(RFE with XGBoost)	94.24	100	96.61	97.73	97.73	97.73
GaussianNB	77.74	80.1	79.66	92.11	79.55	85.37

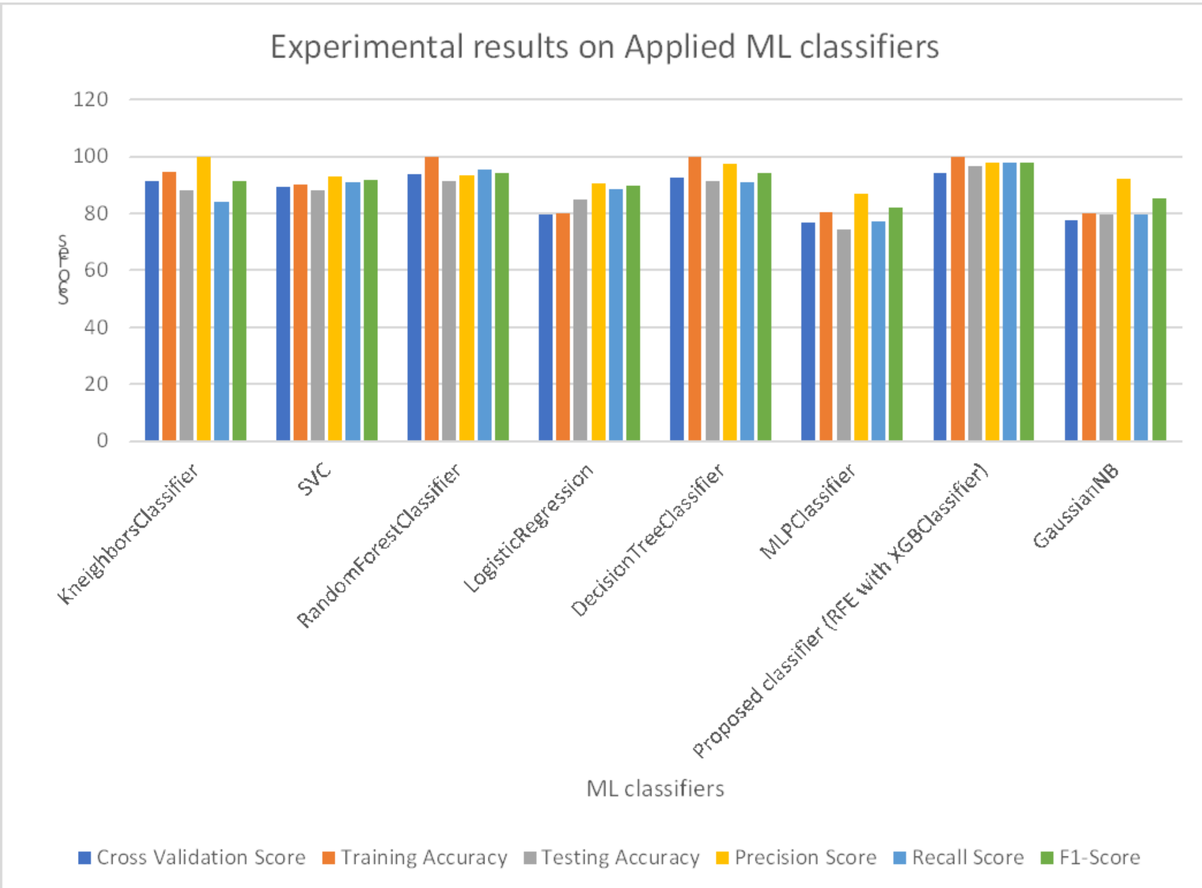

FIGURE 7. Experimental analysis of Machine Learning algorithms in IFRX model.

(TP) value as 43, the True negative (TN) value as 14, the False Positive (FP) value as 1, and the False Negative (FN) values also as 1, as shown in Figure 8. The proposed model has shown an accuracy of 96.61% and Scatter plots are useful for exploring relationships between variables. SHAP scatter plot displays how each attribute affects a single prediction. Figure 9 shows the influence of individual features on the model's output. The horizontal axis represents the SHAP value, which is a measure of how much each feature influences the model's output. The vertical axis represents the feature value. Each dot in the plot represents a feature, and the position of the dot shows how much that feature influences the model's prediction.

The feature at the top of the plot positively affects the model's output, while the feature at the bottom of the plot affects the model's prediction. The features in the middle of the plot have a smaller influence on the model's output. The PPE had an enormous positive influence on its prediction, compared to a lesser positive impact from the MDVP:Fhi(Hz) and a minor negative effect from the D2, as seen in

TABLE 4. Ablation study results.

Experiment Description	Configuration	Classifier	Accuracy (%)
Baseline IFRX Model	SVMSMOTE, (RFE with XGB), SHAP	XGB Classifier	96.61
Remove SVMSMOTE	None, RFE with XGB, SHAP	XGB Classifier	94.10
Replace SVMSMOTE with SMOTE	SMOTE, RFE with XGB, SHAP	XGB Classifier	95.30
Remove (RFE with XGB)	SVMSMOTE, None, SHAP	XGB Classifier	92.45
Replace (RFE with XGB) with PCA	SVMSMOTE, PCA, SHAP	XGB Classifier	93.70
Remove SHAP	SVMSMOTE, (RFE with XGB), None	XGB Classifier	96.61
Replace XGB Classifier with SVM	SVMSMOTE, (RFE with XGB), SHAP	SVM	94.80

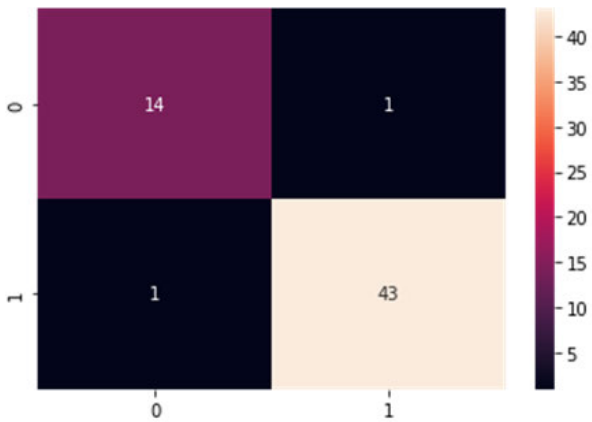

FIGURE 8. RFE with XGBoost confusion matrix.

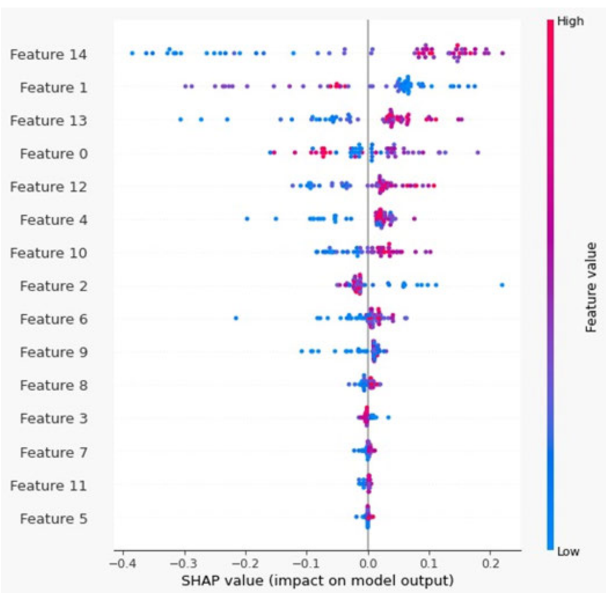

FIGURE 9. Feature impact distribution.

Figure 9. Overall, the results suggest that the XGBoost classifier, particularly when coupled with RFE for feature selection, is identified as the most suitable model for PD prediction based on speech data in the dataset. Other models like Random Forest and Decision Tree classifiers also showed promising results but may not consistently outperform the XGBoost classifier. Conversely, models like Logistic Regression, MLP, and GaussianNB exhibit lower performance and may not be ideal for PD prediction in this context.

A. ABLATION STUDY

To understand the contribution of each component in the proposed IFRX model, we conducted an ablation study by systematically removing or altering individual components. Table 4 summarizes the results of this study.

The ablation study indicates that the combination of SVMSMOTE, RFE with XGB, SHAP, and the XGB classifier collectively leads to the highest performance in the IFRX model. Each component plays a vital role in achieving the optimal accuracy of 96.61%.

B. COMPARATIVE ANALYSIS OF FEATURE SELECTION ALGORITHMS

Table 5 presented a comparative analysis with the state-of-the-art models for PD prediction. It is done based on various preprocessing techniques, feature selection algorithms, and classification methods. The proposed approach shows better performance than other algorithms by introducing the SVMSMOTE technique for preprocessing, RFE with XGBoost for feature selection, and SHAP for model explainability.

It is also observed that Khan et al. [46] employed feature encoding and Evolutionary Wavelet Neural Networks (EWNNs) yielding an accuracy of 90%. Little et al. [47] utilized Support Vector Machines (SVM) without explicit preprocessing or feature selection, achieving an accuracy of 91.40%. Behroozi et al. [48] applied the Pearson Correlation Coefficient for feature selection and ensemble learning, resulting in an accuracy of 87.50%, which is predominantly lesser than the proposed approach. Parisi et al. [49] employed a combination of Multi-Layer Perceptron with Lagrangian Support Vector Machine (MLP-LSVM) and SVM, achieving an accuracy of 78.23%. Mostafa et al. [50] utilized the

TABLE 5. Comparison of the proposed work with the state of the art works for PD prediction.

Study	Speech Dataset	Preprocessing Method	Feature Selection Method	Explainable AI	Classification Method	Accuracy (%)
Khan et al. [46]	UCI ML Repository	Feature Encoding	Evolutionary Wavelet Neural Networks (EWNNs)	No	Not applied	90
Little et al. [47]	UCI ML Repository	Not applied	No	No	SVM	91.40
Behroozi et al. [48]	UCI ML Repository	Not applied	Pearson Correlation Coefficient	No	Ensemble Learning	87.50
Parisi et al. [49]	UCI ML repository	Not applied	Multi-Layer Perceptron with Lagrangian Support Vector Machine (MLP-LSVM)	No	SVM	78.23
Mostafa et al. [50]	UCI ML Repository	Not applied	Multiple Feature Evaluation Approach (MFEA)	No	DT, NB, NN, RF, SVM	95.43
Wroge et al. [51]	UCI ML Repository	Not applied	Mel-frequency cepstral coefficients (MFCCs)	No	DNN	85
Alalayah et al. [30]	UCI ML Repository	SMOTE	RFE	No	Decision Tree	95
Aishwarya et al. [52]	UCI ML Repository	SMOTE	Fisher Score based Recursive Feature Elimination (FRFE)	No	Light GBM	81.35
Proposed Work (IFRX Model)	UCI ML Repository	SVMSMOTE	RFE with XGB	SHAP	XGB Classifier	96.61

Multiple Feature Evaluation Approach (MFEA) for feature selection and achieved better accuracy using various classifiers. Wroge et al. [51] employed Mel-Frequency Cepstral Coefficients (MFCCs) and Deep Neural Networks (DNN), achieving an accuracy of 85%. The proposed model performed better than the other models in terms of accuracy, as seen in Table 5.

Furthermore, our work differs from the existing research works in several ways. Alalayah et al. [30] utilized the Synthetic Minority Over-sampling Technique (SMOTE) and Recursive Feature Elimination (RFE) with a Decision Tree, where we employed an SVMSMOTE to balance the data. Aishwarya et al. [52] applied SMOTE and Fisher Score-based Recursive Feature Elimination (FRFE) with Light GBM, achieving an accuracy of 81.35%. The proposed IFRX model achieved the highest accuracy (96.61%) for Parkinson's Disease (PD) diagnosis using speech data compared to other approaches. This improvement is attributed to several factors, including addressing class imbalance through SVMSMOTE, selecting the most relevant features using RFE with XGBoost, and incorporating explainability through SHAP. These advancements demonstrate the potential of the IFRX model for accurate and comprehensive PD assessment.

IFRX model produces better results compared to existing works on PD diagnosis because the most relevant features are chosen during the feature selection process and address the issue of class imbalance. Overall, the IFRX model demonstrates promising advancements in PD diagnosis using machine learning with speech data, achieving superior accuracy with additional efforts in preprocessing, feature selection, and interpretability.

The comparative analysis highlights the evolution of methodologies over time, ranging from basic techniques like SVM to more sophisticated ones like EWNNs and ensemble learning. The proposed work stands out with a remarkable accuracy of 96.61%, attributed to the utilization of advanced preprocessing and feature selection techniques - SVMSMOTE and RFE with XGBoost, respectively and the improvement indicates the effectiveness of these methods in enhancing classification performance, particularly in handling imbalanced datasets and selecting relevant features. In summary, the experimental results and comparative analysis in this section provide valuable insights into the performance of various machine learning techniques for predicting PD based on speech data. RFE with an XGBoost classifier is selected for feature ranking and XAI, revealing the XGBoost classifier's superior performance across all metrics. With a cross-validation score of 94.24%, training accuracy of 100%, and testing accuracy of 96.61%, the XGBoost model demonstrates robustness and generalizability. Additionally, it achieves high precision, recall, and F1-score, indicating its effectiveness in accurately predicting PD cases.

VI. CONCLUSION AND FUTURE WORK

Based on this work's comparative analysis and findings, we conclude that the RFE with XGBoost classifier is a reliable and effective method for early PD detection based on speech data. Using speech data, the proposed IFRX model can accurately predict PD at an early stage. The combination of RFE and XGBoost is crucial for high accuracy. Compared to other classification algorithms, RFE helps select significant features and remove irrelevant ones. Oversampling using SVMSMOTE helps to address the class imbalance issue, resulting in a more balanced dataset and better model performance. Furthermore, by providing interpretability through SHAP, health practitioners can gain insights into model predictions and identify important features associated with PD, which aids in decision-making and enhances patient outcomes. Our proposed IFRX model showcases the potential of XGBoost for PD detection using speech data, demonstrating its effectiveness as a tool for early diagnosis and personalized treatment planning. The model's accuracy may not be achieved for a wider population since the training data is restricted. Furthermore, there is a disparity in class distribution, where the artificial instances produced by SVMSMOTE may not accurately represent the intricacies of actual data in the real world. Advanced approaches like GANs may achieve more realistic synthetic data creation. Although the study indicates the possibility of using the findings in real-time, more research is needed to confirm its efficacy in those contexts. Future studies could investigate the use of additional modalities, such as medical imaging or clinical data, to further enhance the accuracy and real-time applicability of the prediction model. Additionally, exploring other feature selection techniques and interpretability methods can yield valuable insights into PD diagnosis and contribute to the development of more robust and reliable prediction models. By leveraging these advancements, we can strive to improve the early detection and management of Parkinson's disease, ultimately leading to better patient outcomes and a higher quality of life.

REFERENCES

[1] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 31, pp. 174–180, Jan. 2017.

[2] World Health Org. *Parkinson Disease*. Accessed: Mar. 7, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[3] R. Lamba, T. Gulati, and A. Jain, "Comparative analysis of parkinson's disease diagnosis system," *Adv. Math. Sci. J.*, vol. 9, no. 6, pp. 3399–3406, 2020.

[4] A. Ascherio and M. A. Schwarzschild, "The epidemiology of Parkinson's disease: Risk factors and prevention," *Lancet Neurol.*, vol. 15, no. 12, pp. 1257–1272, Nov. 2016.

[5] S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr, and H. Adeli, "Parkinson's disease: Cause factors, measurable indicators, and early diagnosis," *Comput. Biol. Med.*, vol. 102, pp. 234–241, Nov. 2018.

[6] A. Ma, K. K. Lau, and D. Thyagarajan, "Voice changes in Parkinson's disease: What are they telling us?" *J. Clin. Neurosci.*, vol. 72, pp. 1–7, Feb. 2020.

[7] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, "A systematic approach to diagnose parkinson's disease through kinematic features extracted from handwritten drawings," *J. Reliable Intell. Environments*, vol. 1, pp. 1–10, May 2021.

[8] R. Gupta, M. Khari, D. Gupta, and R. G. Crespo, "Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction," *Inf. Sci.*, vol. 530, pp. 201–218, Aug. 2020.

[9] H. M. R. Afzal, S. Luo, M. K. Afzal, G. Chaudhary, M. Khari, and S. A. P. Kumar, "3D face reconstruction from single 2D image using distinctive features," *IEEE Access*, vol. 8, pp. 180681–180689, 2020.

[10] R. Raj, P. Rajiv, P. Kumar, M. Khari, E. Verdu, R. G. Crespo, and G. Manogaran, "Feature based video stabilization based on boosted Haar cascade and representative point matching algorithm," *Image Vis. Comput.*, vol. 101, Sep. 2020, Art. no. 103957.

[11] R. Gupta, M. Khari, V. Gupta, E. Verde Xing Wu, E. Herrera-Viedma, and R. G. Crespo, "Fast single image haze removal method for inhomogeneous environment using variable scattering coefficient," *Comput. Model. Eng. Sci.*, vol. 123, no. 3, pp. 1175–1192, 2020.

[12] S. Sivaranjini and C. Sujatha, "Deep learning based diagnosis of parkinson's disease using convolutional neural network," *Multimedia Tools Appl.*, vol. 79, pp. 15467–15479, 2020.

[13] R. Lamba, T. Gulati, H. F. Alharbi, and A. Jain, "A hybrid system for parkinson's disease diagnosis using machine learning techniques," *Int. J. Speech Technol.*, vol. 1, pp. 1–11, Aug. 2021.

[14] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou, and S. Mohammed, "Data-driven based approach to aid Parkinson's disease diagnosis," *Sensors*, vol. 19, no. 2, p. 242, Jan. 2019.

[15] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for parkinson's disease diagnosis from EEG signals," *Neural Comput. Appl.*, vol. 32, pp. 10927–10933, May 2020.

[16] C. Loconsole, G. D. Cascarano, A. Brunetti, G. F. Trotta, G. Losavio, V. Bevilacqua, and E. Di Sciascio, "A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis," *Pattern Recognit. Lett.*, vol. 121, pp. 28–36, Apr. 2019.

[17] Ö. F. Ertusrul, Y. Kaya, R. Tekin, and M. N. Almalı, "Detection of parkinson's disease by shifted one dimensional local binary patterns from gait," *Expert Syst. Appl.*, vol. 56, pp. 156–163, Sep. 2016.

[18] J. Goyal, P. Khandnor, and T. C. Aseri, "A comparative analysis of machine learning classifiers for dysphonia-based classification of Parkinson's disease," *Int. J. Data Sci. Analytics*, vol. 11, no. 1, pp. 69–83, Jan. 2021.

[19] G. Solana-Lavalle, J.-C. Galán-Hernández, and R. Rosas-Romero, "Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 505–516, Jan. 2020.

[20] T. Tuncer, S. Dogan, and U. R. Acharya, "Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 211–220, Jan. 2020.

[21] Y. Liu, Y. Li, X. Tan, P. Wang, and Y. Zhang, "Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of Parkinson's disease," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102165.

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

[23] J. Goyal, P. Khandnor, and T. C. Aseri, "Analysis of parkinson's disease diagnosis using a combination of genetic algorithm and recursive feature elimination," in *Proc. 4th World Conf. Smart Trends Syst., Secur. Sustainability*, 2020, pp. 268–272.

[24] O. Bchir, "Parkinson's disease classification using Gaussian mixture models with relevance feature weights on vocal feature sets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 1–20, 2020.

[25] A. S. Ashour, M. K. A. Nour, K. Polat, Y. Guo, W. Alsaggaf, and A. El-Attar, "A novel framework of two successive feature selection levels using weight-based procedure for voice-loss detection in Parkinson's disease," *IEEE Access*, vol. 8, pp. 76193–76203, 2020.

[26] K. Polat and M. Nour, "Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals," *Med. Hypotheses*, vol. 140, Jul. 2020, Art. no. 109678.

[27] Z. Soumaya, B. Drissi Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of Parkinson disease using the genetic algorithm and SVM classifier," *Appl. Acoust.*, vol. 171, Jan. 2021, Art. no. 107528.

[28] M. Hoq, M. N. Uddin, and S.-B. Park, "Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection," *Diagnostics*, vol. 11, no. 6, p. 1076, Jun. 2021.

[29] P. K. Chawla, M. S. Nair, D. G. Malkhede, H. Y. Patil, S. K. Jindal, A. Chandra, and M. A. Gawas, "Parkinson's disease classification using nature inspired feature selection and recursive feature elimination," *Multimedia Tools Appl.*, vol. 1, pp. 1–24, Sep. 2023.

[30] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method," *Diagnostics*, vol. 13, no. 11, p. 1924, May 2023.

[31] M. Little, "UCI machine learning repository: Parkinsons data set," Univ. California, Irvine, Irvine, CA, USA, Tech. Rep., 2008.

[32] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *J. Reliable Intell. Environments*, vol. 7, no. 3, pp. 263–275, Sep. 2021.

[33] E. S. TC and V. D. RS, "Prediction of Parkinson's disease using XGBoost," in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2022, pp. 1769–1772.

[34] T. Maguire, L. Manuel, R. Smedinga, and M. Biehl, "A review of feature selection and ranking methods," in *Proc. 19th SC@ RUG*, 2022, pp. 15–20.

[35] X. Shi, Y. D. Wong, M. Z.-F. Li, C. Palanisamy, and C. Chai, "A feature learning approach based on XGBoost for driving assessment and risk prediction," *Accident Anal. Prevention*, vol. 129, pp. 170–179, Aug. 2019.

[36] E. M. Senan, I. Abunadi, M. E. Jadhav, and S. M. Fati, "Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–16, Dec. 2021.

[37] E. M. Senan and M. E. Jadhav, "Diagnosis of dermoscopy images for the detection of skin lesions using SVM and knn," in *Proc. 3rd Int. Conf. Sustain. Comput.*, 2021, pp. 125–134.

[38] E. M. Senan, M. E. Jadhav, and A. Kadam, "Classification of PH2 images for early detection of skin diseases," in *Proc. 6th Int. Conf. Converg. Technol. (I2CT)*, Apr. 2021, pp. 1–7.

[39] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features," *Diagnostics*, vol. 13, no. 6, p. 1026, Mar. 2023.

[40] A. I. Schein and L. H. Ungar, "Active learning for logistic regression: An evaluation," *Mach. Learn.*, vol. 68, no. 3, pp. 235–265, Aug. 2007.

[41] L. Steigmann, R. Di Gianfilippo, M. Steigmann, and H.-L. Wang, "Classification based on extraction socket buccal bone morphology and related treatment decision tree," *Materials*, vol. 15, no. 3, p. 733, Jan. 2022.
[42] E. Wilson and D. W. Tufts, "Multilayer perceptron design algorithm," in *Proc. IEEE Workshop Neural Netw. Signal Process.*, Aug. 1994, pp. 61–68.
[43] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Cham, Switzerland: Springer, 2003, pp. 63–71.
[44] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," in *Proc. Artif. Intell. Signal Process. Conf. (AISP)*, Oct. 2017, pp. 209–212.
[45] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance analysis and comparison of machine and deep learning algorithms for IoT data classification," 2020, *arXiv:2001.09636*.
[46] M. M. Khan, A. Mendes, and S. K. Chalup, "Evolutionary wavelet neural network ensembles for breast cancer and Parkinson's disease prediction," *PLoS ONE*, vol. 13, no. 2, Feb. 2018, Art. no. e0192192.
[47] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *Nature Precedings*, p. 1, Sep. 2008.
[48] M. Behroozi and A. Sami, "A multiple-classifier framework for Parkinson's disease detection based on various vocal tests," *Int. J. Telemedicine Appl.*, vol. 2016, pp. 1–9, Jun. 2016.
[49] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease," *Expert Syst. Appl.*, vol. 110, pp. 182–190, Nov. 2018.
[50] S. A. Mostafa, A. Mustapha, M. A. Mohammed, R. I. Hamed, N. Arunkumar, M. K. A. Ghani, M. M. Jaber, and S. H. Khaleefah, "Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease," *Cognit. Syst. Res.*, vol. 54, pp. 90–99, May 2019.
[51] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's disease diagnosis using machine learning and voice," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2018, pp. 1–7.
[52] R. Aishwarya, K. Pavitra, P. V. Miranda, K. Keerthana, and L. Kamatchi Priya, "Parkinson's disease prediction using Fisher score based recursive feature elimination," in *Proc. Int. Conf. Advancement Comput. Comput. Technol. (InCACCT)*, May 2023, pp. 1–8.

K. SHYAMALA is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Tamil Nadu, India. Her research interests include machine learning and deep learning.

T. M. NAVAMANI received the M.E. and Ph.D. degrees in CSE from Anna University, Chennai. She is currently a Senior Associate Professor with the School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, India. She has 26 years of teaching experience. She has published more than 40 papers in international and national journals, conferences, and book chapters. Her research interests include data sciences, human–computer interaction, wireless ad hoc networks, network security, and mobile computing. She is a Life Member of the Indian Society for Technical Education (ISTE).