



A Mutual-Information-Guided and ADASYN-Augmented Machine Learning Framework for Early Prediction of Parkinson's Disease

Ghadeer Aqil Ali¹, Leila Sharifi², Parviz Rashidi-Khazaei^{3*}, Hossein Nahid-Titkanlue⁴

¹ Department of Electrical and Computer Engineering, Urmia University, Urmia, Iran

² Assistant Professor, Department of Electrical and Computer Engineering, Urmia University, Urmia, Iran

³ Assistant Professor, Department of Information Technology and Computer Engineering, Urmia University of Technology, Urmia, Iran

⁴ Assistant Professor, Department of Industrial Engineering, Payame Noor University, Tehran, Iran

* Corresponding author email address: p.rashidi@uut.ac.ir

Received: 2025-06-11

Revised: 2025-10-12

Accepted: 2025-10-19

Initial Publish: 2025-10-19

Final Publish: 2026-05-10

Abstract

Early detection of Parkinson's disease (PD) is essential for timely medical intervention and improving patient outcomes. Speech signal analysis offers a non-invasive, cost-effective, and easily deployable diagnostic pathway. However, achieving reliable early prediction remains challenging due to data imbalance, redundant features, and model instability. This study aims to develop an optimized and robust machine learning framework that enhances the predictive accuracy and stability of PD detection from speech data. An optimized machine learning model based on eXtreme Gradient Boosting (XGBoost) was developed for early PD prediction. The model's hyperparameters were tuned using the Tree-structured Parzen Estimator (TPE), while Mutual Information (MI) was employed to select the most informative features from the speech dataset. To address class imbalance, the Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN) was applied to generate synthetic minority samples. Model performance and stability were evaluated using ten independent runs of Stratified 10-Fold Cross-Validation (SCV). The proposed framework achieved superior predictive performance with an average accuracy of 97.27%, precision of 98.79%, F1-score of 97.18%, recall of 95.77%, and ROC-AUC of 98.11% across multiple evaluations. Comparative analysis with similar studies demonstrated improved robustness, reliability, and balance between sensitivity and specificity. The integration of MI-based feature selection and ADASYN-based data augmentation significantly enhanced the performance and stability of the XGBoost model for early PD prediction. The proposed model demonstrates strong potential for clinical use as a decision support system, providing a low-cost, non-invasive, and remotely deployable tool for early PD diagnosis using patient speech signals.

Keywords: XGBoost, Tree-Structured Parzen Estimator, Data Augmentation, ADASYN, Feature Selection, Mutual Information.

How to cite this article:

Aqil Ali, G., Sharifi, L., Rashidi-Khazaei, P., & Nahid-Titkanlue, H. (2026). A Mutual-Information-Guided and ADASYN-Augmented Machine Learning Framework for Early Prediction of Parkinson's Disease. *Management Strategies and Engineering Sciences*, 8(3), 1-12.

1. Introduction

In modern medicine, the early detection of diseases—prior to the onset of severe clinical symptoms—is of critical importance [1-3]. The main objective of early diagnosis is to prevent disease progression, enable timely intervention, and ideally achieve rapid recovery through early treatment. To support this goal, machine learning (ML) technologies have garnered significant attention from researchers and

clinicians alike, proving their effectiveness in various aspects of healthcare, including diagnosis, treatment planning, monitoring, and, notably, early disease prediction [4-7]. Parkinson's disease (PD), characterized by the loss of approximately 80% of dopamine-producing neurons in the brain [8], is a progressive neurodegenerative disorder and ranks as the second most prevalent condition affecting the central nervous system [9]. Due to its increasing incidence



worldwide, PD has even been referred to as a “Parkinson’s pandemic” [10].

The diagnosis and treatment of Parkinson’s disease (PD) involve several global and region-specific challenges. First, neurologists and movement disorder specialists can confirm PD only after an extensive review of the patient’s medical history, supported by multiple clinical examinations and imaging tests [11, 12]. This process is often time-consuming and costly, making it particularly difficult to implement effectively. These difficulties are even more pronounced in developing countries, where there is often a shortage of experienced specialists and limited access to advanced medical facilities [13].

Second, although conventional clinical diagnostic tools can help assess disease severity and distinguish PD from other neurological disorders [14], the absence of specific biomarkers and the overlap of PD symptoms with those of other conditions make early-stage diagnosis particularly difficult [1].

Third, as PD progresses, its symptoms intensify, leading to a significant increase in treatment and care costs [15], and the disease eventually affects multiple organs beyond the brain [16, 17]. Therefore, early prediction and diagnosis of PD are highly beneficial, offering not only substantial cost savings but also improved life expectancy and quality of life for elderly individuals [3, 18-21].

A review of recent scientific literature reveals a notable increase in the use of machine learning (ML) techniques for the early detection of Parkinson’s disease (PD) over the past several years. Researchers have explored a variety of ML-based approaches to develop reliable and efficient diagnostic tools. Central to these studies is the implementation of various classification algorithms, including Artificial Neural Networks (ANN)[22, 23], K-Nearest Neighbor (KNN)[23-26], Support Vector Machine (SVM)[22, 23, 27, 28], Deep Neural Network (DNN)[29-31], Random Forest (RF)[25, 28, 29], eXtreme Gradient Boosting(XGB)[32, 33], and Convolutional Neural Network (CNN)[21, 34, 35].

To improve ML classifier performance, feature selection (FS) algorithms help reduce the feature set, select more important input features, and enhance classifier performance, especially when the number of dimensions is high and the size of the data is small[36]. To reduce the dimension of the PD dataset, which had only 195 records, different techniques like principal component analysis(PCA) [28, 30], recursive feature elimination(RFE) and feature importance(FI) [22], extra tree(ET) [37, 38], mutual information(MI) [26], genetic algorithm(GA) [26], chi2

feature selection[38], correlation matrix [38], SelectKBest(SKB) [39], multi-agent salp swarm(MASS) [35], cuckoo search algorithm (CSA) [19, 40], SHAP [32], collinearity-based feature elimination(CFE) [28], cuttlefish algorithm(CFA) [24] and modified gray wolf optimization(MGWO) [25] had presented and used. These methods have significantly improved the efficiency of ML methods.

Data augmentation methods provide an effective strategy for improving the performance of machine learning (ML) classifiers, particularly when working with limited datasets. These techniques aim to expand the dataset by creating new, synthetic samples that closely resemble the original data. Among the various approaches, the Synthetic Minority Oversampling Technique (SMOTE) [26, 28, 34, 39, 41], is the most commonly applied in the early prediction of Parkinson’s disease (PD). Studies have demonstrated that using SMOTE can significantly enhance the accuracy and robustness of predictive models.

Optimizing the parameters and hyperparameters of machine learning (ML) models is a critical step in developing effective predictive tools. Various optimization algorithms have been successfully applied for early Parkinson’s disease (PD) prediction, including Particle Swarm Optimization (PSO) [42], grid search cross-validation (GSCV) [27, 28, 39], and tree-structured Parzen estimator (TPE) [33]. These techniques help identify the best model configurations, thereby improving predictive performance and reliability.

In this study, after comparing various feature selection methods, Mutual Information (MI) [26] was employed to identify the optimal subset of features. To address the class imbalance in the Parkinson’s disease (PD) dataset, the Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN) [43] was applied for the first time. Subsequently, an optimized XGBoost (XGB) model [44] tuned using the Tree-Structured Parzen Estimator (TPE)[45], was proposed for early PD prediction. The integration of these techniques provides a robust and reliable framework that outperforms existing methods while ensuring stability and accuracy in PD prediction.

The rest of this paper is structured as follows: Section 2 presents the methodology, Section 3 details the results, Section 4 provides the discussion, and Section 5 concludes the study.

2. Methodology

2.1. Proposed PD Early Prediction Framework Structure

In this study, to have a stable and strong PD early prediction tool, an advanced practical framework has been proposed. Figure 1 shows its structure. Initially, the MI[26] was employed to identify the most significant features. Second, ADASYN [43] was used for the first time to address the PD dataset imbalance problem and generate new

synthetic data. Third, stratified 10-fold cross-validation (S10FCV) was employed to split the data into training and testing sets. Fourth, a new optimized XGB-TPE [46] was proposed for the early prediction of PD. Fifth, at each iteration, each tool was evaluated on different unseen test data. Finally, a new rigid evaluation strategy, described in section 2.7, was proposed to consider the average of 10 different runs of S10FCV (A10S10FCV) evaluation metrics as a measure of the tool's overall performance. The proposed framework is named MAXT(MI-ADASYN-XGB-TPE).

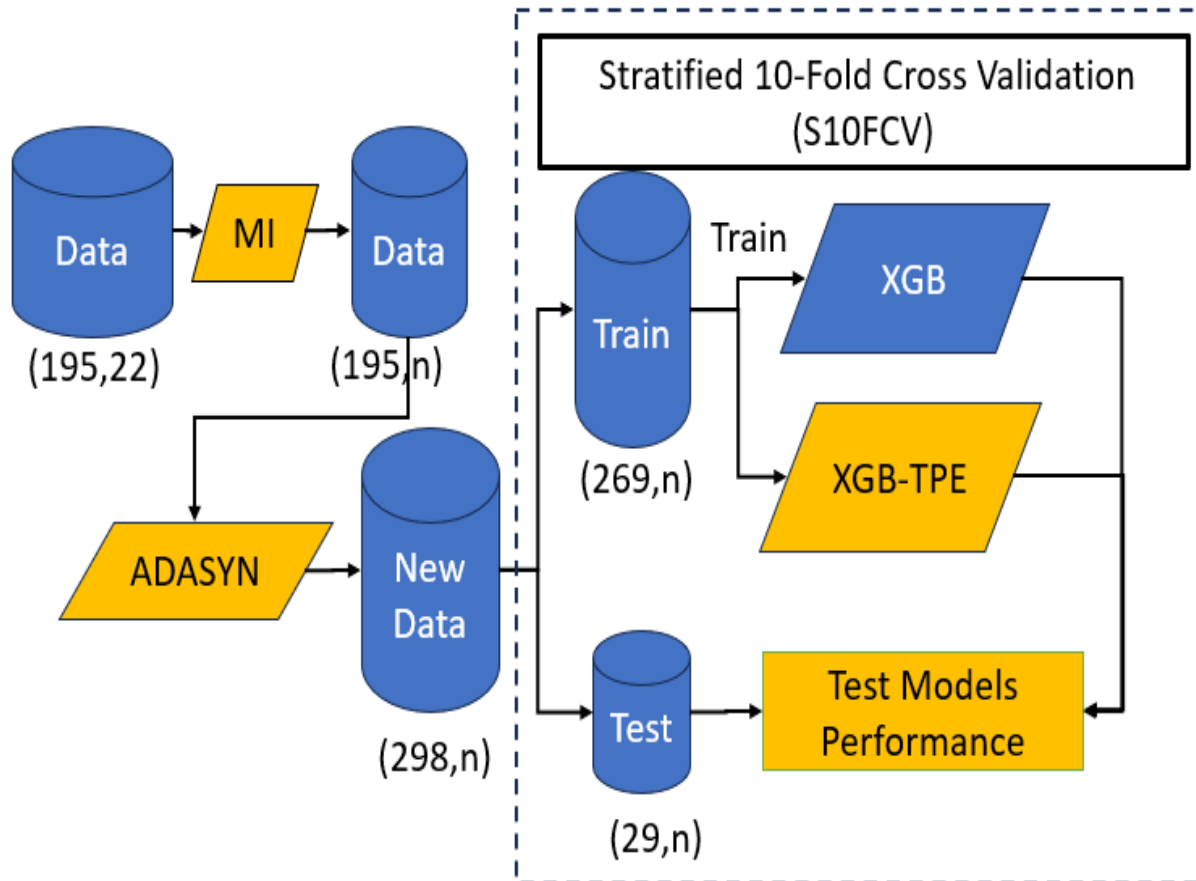


Figure 1. The structure of the proposed MAXT framework for PD prediction

2.2. Dataset Description

The voice speech signal dataset, collected by researchers at the University of Oxford, is publicly accessible through the UCI Machine Learning Repository [47]. This dataset supports the use of non-invasive techniques for the early identification of Parkinson's disease (PD). It includes data from 31 participants—both male and female—ranging in

age from 46 to 85 years. Among these individuals, 23 were diagnosed with PD, while the remaining 8 formed the healthy control group. The dataset comprises 195 distinct biomedical voice features. Table 1 presents a comprehensive description of these attributes and their corresponding meanings. As the features exist on different numerical scales, all input variables were standardized using a standard scaler so that each feature achieved a mean of 0 and a standard deviation of 1.

Table 1. The Voice Speech PD Dataset Characteristics

#	Feature	Selected*	Description
Input Variables			
1	MDVP:F0(Hz)	✓	Average vocal fundamental frequency
2	MDVP:Fhi(Hz)	✓	Maximum vocal fundamental frequency
3	MDVP:Flo(Hz)	✓	Minimum vocal fundamental frequency
4	Jitter (%)	-	Several measures of variation in fundamental frequency
5	Jitter (Abs)	✓	Measure of Jitter in absolute terms
6	RAP	-	Measure of Rapid Jitter
7	PPQ	-	Measure of Jitter using the PPQ method
8	DDP	-	Measure of Jitter using the DDP method
9	Shimmer	-	Several measures of variation in fundamental frequency amplitude
10	Shimmer(dB)	✓	Shimmer in decibels
11	APQ	✓	Average perturbation quotient
12	APQ3	-	Amplitude perturbation quotient measured in the first three instants
13	APQ5	✓	Amplitude perturbation quotient measured in the first five instants
14	DDA	-	Shimmer—difference between the amplitudes of consecutive periods
15	NHR	-	Noise-to-harmonics ratio
16	HNR	✓	Harmonics-to-noise ratio
17	RPDE	-	Recurrence period density entropy
18	DFA	-	Signal fractal scaling exponent
19	Spread1	✓	Nonlinear measure of fundamental frequency variation
20	Spread2	✓	Another nonlinear measure of fundamental frequency variation
21	D2	-	Correlation dimension
22	PPE	✓	Pitch period entropy
23	Status	Target	Health status of the subject

*-The Marked feature was selected by the feature selection algorithm as the final input of classifier.

2.3. Feature Selection Module

After evaluating different feature subset selection algorithms, the MI [26] combination with the proposed optimized framework brings better performance and is considered the optimal feature subset selection module. Mutual information (MI) is a key concept in information theory that measures how much one random variable can reveal about another. When it comes to feature selection, MI is incredibly beneficial—it helps pinpoint the most important features by assessing how much they reduce uncertainty about the target variable. In simple terms, it tells us how valuable a feature is in predicting or explaining the outcome we are interested in.

2.4. Data Augmentation Module

In this research, after comparing SMOTE and several other data augmentation techniques, the ADASYN method [43] was selected and applied for the first time to generate synthetic data for Parkinson's disease (PD) as part of the data

augmentation process. ADASYN is an advanced and adaptive oversampling approach specifically developed to handle class imbalance in machine learning tasks. Rather than merely replicating samples from the minority class, ADASYN synthesizes new instances by concentrating on the more difficult samples—those located near the classification boundary or surrounded by majority class samples. By dynamically determining the number of synthetic data points needed for each minority instance, the method effectively balances the dataset and improves the model's ability to learn from underrepresented categories. This makes ADASYN particularly valuable in medical diagnosis applications, where rare yet significant cases could otherwise be missed.

2.5. Classifier Module

In this study, among various evaluated classifiers, the XGBoost (XGB) classifier was chosen as the primary model. The overall architecture of the XGB classifier is illustrated in Figure 2 [44]. This algorithm is built upon an ensemble of weak regression trees, where each tree is trained sequentially

to minimize the errors made by its predecessor. Through this iterative process, every weak learner (WL) focuses on correcting the mistakes of the previous one, thereby enhancing the model's overall predictive strength. During

the training phase, the optimal XGB tree ensemble—composed of multiple weak learners—is generated, which is then employed to make predictions on new, unseen data samples.

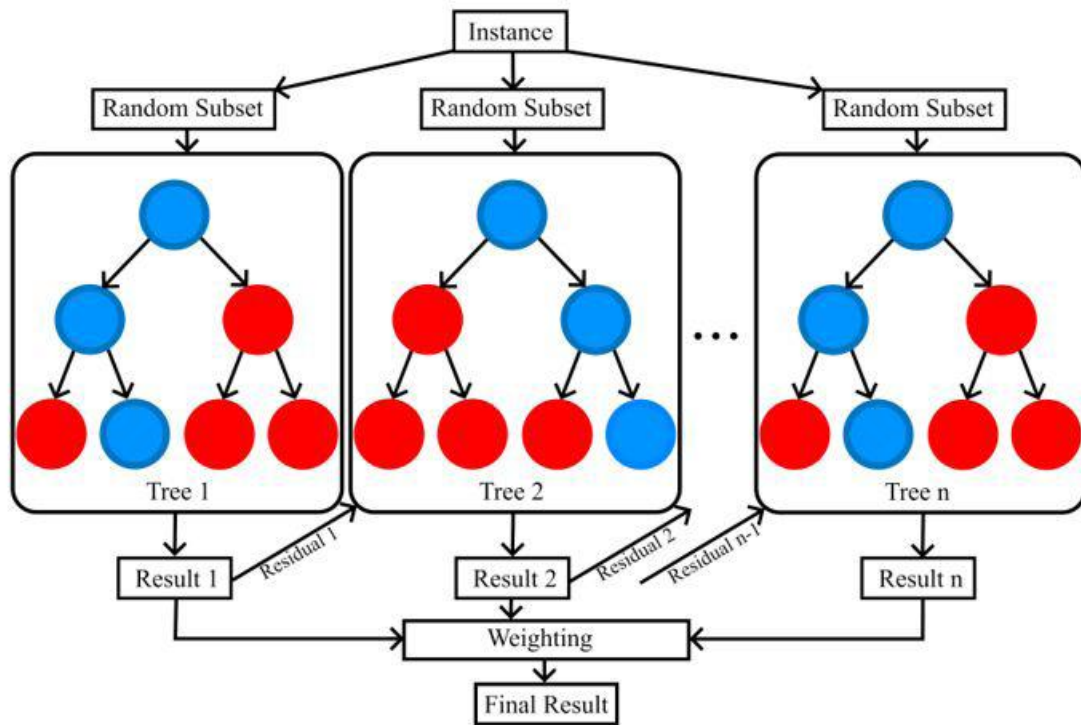


Figure 2. The basic structure of the XGB classifier [46]

XGBoost (XGB) offers several distinctive characteristics that allow it to surpass many other classification algorithms and position itself as a state-of-the-art technique across multiple domains [44]. Its primary strengths include fast computation, high predictive accuracy, scalability to large datasets, flexibility for customization, the capability to manage missing data, and ease of interpretability—all of which contribute to its widespread adoption and effectiveness. Nonetheless, XGB is not without limitations. It can be computationally intensive, may be prone to overfitting in certain scenarios, and involves a large number of hyperparameters that must be meticulously tuned. Therefore, these challenges should be taken into account when utilizing XGB in practical applications.

2.6. Hyperparameter Tuning Module

Among various hyperparameter optimization techniques—such as random search and grid search cross-validation (GSCV)—the Tree-Structured Parzen Estimator (TPE) method offers a more intelligent approach by

adaptively exploring the search space and concentrating on regions with high potential based on prior evaluations. TPE has demonstrated superior performance in several domains, including energy prediction [46]. Bergstra et al. introduced the TPE algorithm by applying Bayesian principles and formulating it around conditional probability distributions [48]. Specifically, $p(x|y)$ represents the conditional probability of a parameter configuration x given an observed loss value y . The objective in this approach is to determine the optimal parameter value x^* that minimizes the loss, denoted by y . A threshold y^* for the loss—often set as a statistic such as the median—is first established based on the available data. The final x^* obtained corresponds to the optimized hyperparameter set for the model. In this work, the original implementation of the TPE algorithm provided within the Hyperopt library was employed [48].

2.7. Proposed Evaluation Strategy

To ensure a fair evaluation, particularly when working with a small dataset, it is crucial to assess the framework's

performance under various conditions and test it against every sample. K-Fold Cross-Validation (KFCV) is a widely used method in such cases. In this research, the stratified version of K-Fold Cross-Validation (SKFCV) was employed for performance assessment. The key difference between KFCV and SKFCV lies in the stratification process, which guarantees that each fold of the test data maintains the same proportion of samples for each class label.

In this study, stratified 10-Fold Cross-Validation (S10FCV) with $K=10$ was applied, and the entire procedure was repeated ten times using distinct random seeds—specifically, ten different prime numbers less than 200—to generate unique training and testing sets for each iteration. This approach helps address dataset imbalance by ensuring that each data subset accurately represents the original class distribution, thereby providing a more reliable evaluation. In contrast, a standard 10-fold split may lead to uneven distributions and biased results.

This repetitive process also serves to verify the stability of the proposed framework. The final performance measure was calculated as the average of ten runs of S10FCV, denoted as A10S10FCV. The evaluation and comparison of models were conducted using common performance metrics,

including Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

3. Results

Initially, by using MI as feature selection module, 11 important features were identified from a total of 22 features. The third column of Table 1 marked the selected features. To address the dataset's imbalance problem, ADASYN generated new synthetic data and incorporated it into the dataset. As a result, the positive and negative class distributions of the samples in the dataset (147, 48) changed to (147, 146).

The performance of the basic XGB and optimized using TPE (XGB-TPE) was investigated. The results of their combination with ADASYN and SMOTE methods for data augmentation, as well as the results of their combination with optimal feature selection by MI, are presented in Table 2 and Figure 3. The stability and robustness of the proposed framework, along with their performance, were evaluated using A10S10FCV, and the average value of 10 different runs was considered the final performance criterion of the tool.

Table 2. The Prediction Performance of Different tools using A10S10FCV

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
XGB	92.15	93.21	97	94.92	96.81
ADASYN-XGB	96.57	98.26	94.81	96.38	99.46
ADASYN-XGB-TPE	97.01	98.74	95.22	96.84	98.83
MI-ADASYN-XGB-TPE (MAXT)	97.26	98.79	95.77	97.18	98.11

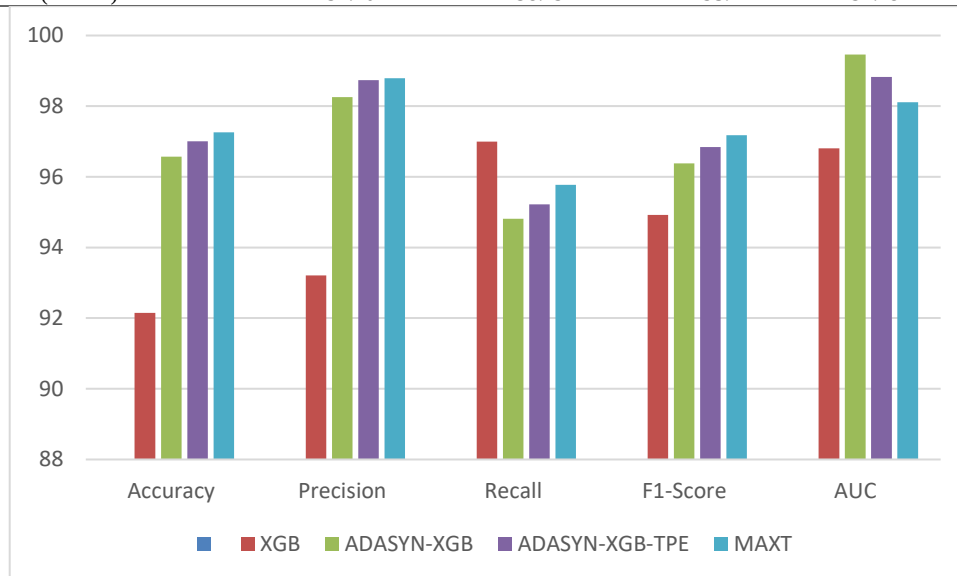


Figure 3. Comparison of different tools evaluation metrics with each other.

The results presented in Table 2 showed that ADASYN and TPE have improved the performance of the XGB classifier, and ADASYN has performed better than SMOTE in early PD prediction. In addition, dimensionality reduction using MI has helped improve the performance of the proposed MAXT framework. Based on various evaluation criteria, the proposed MAXT framework has higher precision, accuracy, and F1 score than other works, which are 97.26%, 98.79%, and 97.18%, respectively, and the

XGB classifier has higher recall, 97%, and the ADASYN-XGB classifier has higher ROC-AUC criterion, 99.46%.

The results of 10 different runs of the proposed MAXT framework are presented in Table 3. As shown, the accuracy of S10FCV with random-state = 179 is 97.94%, which is higher than the average value of 97.26%. However, as shown in Table 2, the average value based on A10S10FCV is considered as the evaluation criterion for the MAXT framework.

Table 3. The Proposed MAXT Framework Stability Checking Using S10FCV

Random State	Accuracy	Precision	Recall	F1-Score	AUC
17	97.25	98.66	95.86	97.19	97.29
29	96.92	97.32	96.67	96.94	99.06
37	96.55	98.56	94.52	96.47	97.43
42	96.90	98.71	95.14	96.75	97.71
53	97.25	99.33	95.19	97.19	98.54
89	97.60	99.33	96.00	97.45	97.98
101	97.62	98.57	96.57	97.51	97.97
139	97.60	99.38	95.90	97.53	98.77
179	97.94	99.33	96.57	97.90	98.78
199	96.93	98.67	95.29	96.89	97.57
Average	97.26	98.79	95.77	97.18	98.11

To compare the prediction accuracy of the proposed framework in the present study with some published works (e.g., tools evaluated using different training and test data splits, such as 80:20), the results of implementing the

proposed MAXT framework with the best performance of S10FCV, with random-state = 179, are shown in Table 4. The presented results showed that the proposed framework achieved 100% accuracy in different test folds.

Table 4. The Proposed MAXT framework Best Run Performance based on S10FCV

Test Fold	Accuracy	Precision	Recall	F1-Score	AUC
1	100.00	100.00	100.00	100.00	100.00
2	96.67	100.00	93.33	96.55	96.89
3	96.55	100.00	93.33	96.55	96.19
4	100.00	100.00	100.00	100.00	100.00
5	93.10	93.33	93.33	93.33	97.14
6	100.00	100.00	100.00	100.00	100.00
7	100.00	100.00	100.00	100.00	100.00
8	96.55	100.00	92.86	96.30	98.57
9	100.00	100.00	100.00	100.00	100.00
10	96.55	100.00	92.86	96.30	99.05
Average	97.94	99.33	96.57	97.90	98.78

4. Discussions

To compare the proposed MAXT framework performance with previous published works, various factors

were considered, including feature selection algorithm, number of features selected, data augmentation method, type of classifier, and the evaluation method used. Table 5 provides complete information about these methods and their comparison with the proposed MAXT framework.

Table 5. Comparison of the Proposed MAXT PD Early Prediction framework with Other Works.

Work	FS	#	DA	Classifier	Tuning Alg.	Testing Method	Best Acc (%)	Recall (%)
AL-Fatlawi et al. [49]	-	16	-	DBN	-	N/A	94	N/A
Pahuja and Nagabhushan[23]	-	22	-	ANN, KNN, SVM	-	10FCV	95.89	93.75
Gupta et al. [24]	CFA	14	-	KNN, DT	-	70:30	92.19	N/A
Kadam and Jadhav [50]	-	22	-	DNN	-	10 # of 10FCV	93.84	95.23
Sharma et al. [25]	MGWO	-	-	KNN, DT, RF	-	70:30	93.87	N/A
Das et al. [42]	-	22	-	Ensemble	PSO	10FCV	93.78	84
Senturk [22]	RFE, FI	7, 13	-	SVM, ANN, RT	-	N/A	93.84	N/A
Lamba et al. [26]	ET, MI, GA	11, 11, 5	SMOTE	Naïve Bayes, KNN, RF	-	10FCV	95.58	93.19
Jain et al. [41]	-	22	SMOTE	DNN	-	80:20	91.47	97.12
Rahman et al. [30]	PCA	N/A	-	Different ML, DNNs	-	80:20	95.41	95
Das et al. [40]	CSA	10	-	VWELM	-	N/A	99.21	100
Yadav et al., [38]	Chi ² , ET, CM	11, 10, 10	-	DT, Gradient Boosting	-	80-20	94.87	N/A
Alsham... et al. [39]	SKB	8	SMOTE	MLP	GSCV	70-30	98.31	96
Akila and Nayahi [35]	MASS	N/A	-	CNN	-	80:20	99.1	94.7
Reddy et al. [32]	SHAP	6	SMOTE	RF, XGBoost	-	80-20	95	100
Arasavali et al.[51]	-	22	-	Hybrid DNN, LSTM	-	80-20	95	N/A
Thirapanish et al.[27]	L1-norm SVM, RFE	3-20	-	SVM	GSCV	5FCV	N/A	88.75
Kadhim et al.[19]	CSA	-	-	Gower distance	-	70:30	98.3	100
Patel[21]	-	22	-	Hybrid CNN-LSTM	-	N/A	95	92
Saha et al [34]	-	22	SMOTE	Ensemble of PD-CNN	-	70:20	99.47 96.18*	98.19 95.63*
Balaha et al. [33]	PSO	-	-	Majority Voting	TPE	A10010FCV	95.67	84.5
Baruah et al. [28]	PCA, CFE	N/A	SMOTE	RF, LR, SVM	GSCV	5-FCV	97.44	100
Proposed MAXT	MI	11	ADASYN	XGB	TPE	A10S10FCV	97.27	98.79
						Best S10FCVRun	97.94	96.57
						Best Fold Testing	100	100

*- Our implementation performance of their work using A10S10FCV., # - Number of selected features

In most published works, including the present study, the performance of tools has usually been compared only in terms of accuracy. In the context of medical problems, accurately detecting positive cases is more important than

the tools overall accuracy. Therefore, it is better to compare the overall performance of tools based on their sensitivity (recall). Upon analyzing and comparing the results of this criterion, it was found that the sensitivity of some tools was

not adequately taken into account. However, comparing the sensitivity of the proposed MAXT framework with other works revealed that, in this criterion, the proposed framework performs significantly better than the other works.

The primary distinction of our work from others is the use of the proposed A10S10FCV strategy for evaluation, as described in Section 2.7. In this way, we tested the stability and reliability of the proposed framework, which are fundamental issues in designing ML tools. As shown in Table 5, the evaluation method was not specified in some studies. Other studies like [19, 24, 25, 34, 39] used the 70:30, and [30, 32, 35, 38, 41, 51] used the 80:20 strategy to train and evaluate tools performance, but it is not clear whether they reported best results or the average of different executions. The results presented in Table 4 show that the proposed MAXT framework, with a 90:10 data splitting strategy, achieved 100% accuracy in 5-fold testing. Therefore, it outperforms other published works tested using an 80(70):20(30) data splitting strategy. By considering A10S10FCV or similar strategies for tool evaluation, only [50] and [33] have employed a similar approach, reporting accuracies of 93.84% and 95.67%, respectively, which is weaker than the proposed MAXT framework accuracy of 97.26%. In comparison with other works evaluated using 10-FCV or 5-FCV, it is worth noting that only [23, 26, 42] utilized 10-FCV, while [27, 28] employed 5-FCV. To ensure a fair comparison, the proposed MAXT best run with random-state =179 was considered. As shown in Table 2, the MAXT accuracy was 97.94%, demonstrating superiority over [23, 26, 28, 42], which achieved accuracies of 95.89%, 93.78%, 95.58%, and 97.44%, respectively. The [27] did not report accuracy, but the reported recall of 88.75% is also weaker than the MAXT recall of 98.79%.

In comparison with works that used XGB as a classifier, [32] employed SMOTE for data augmentation and XGB for PD early prediction, and evaluated their tool once, based on an 80:20 data split. Their work accuracy is reported as 95%, which is weaker than the proposed framework's result. Additionally, [33] employed PSO for feature selection and TPE for hyperparameter tuning, utilizing XGB and a majority voting classifier for prediction. Their XGB classifier achieved an accuracy of 90.20%, and their best ensemble classifier achieved an accuracy of 95.67%, which is weaker than the results presented in this study.

The second main distinction of our work is the use of ADASYN as a data augmentation technique, as opposed to the SMOTE method used by others. This method is used for

the first time on PD data. As shown in Table 2 and Figure 3, ADASYN outperforms SMOTE in improving classifier performance. ADASYN generates synthetic data points by focusing on harder-to-learn examples—those that are close to the decision boundary or surrounded by examples of the majority class. This approach is particularly useful in medical diagnostics, where rare but critical cases may otherwise be missed.

The third main distinction of our work is the use of MI as an optimal feature selector. The results presented in Table 2 show that using MI helped improve framework performance. In a medical context, especially when the data size is low and dimension is high, feature selection not only facilitates dimensionality reduction but also mitigates the curse of dimensionality, thereby improving classifier robustness and improving performance in prediction tasks.

The performance of the proposed MAXT prediction framework demonstrates that combining MI as the optimal feature selector, ADASYN as a data augmentation tool to address the data imbalance issue, and XGB hyperparameter tuning using TPE has resulted in a more stable and predictable tool that can detect PD early and efficiently. MI has been used before, but ADASYN, XGB, and TPE were used for the first time for PD early prediction. The proposed combination, as represented by the MAXT framework, can be utilized as a reliable and practical tool by healthcare providers in their daily operations.

The primary limitation of this study is that the framework was trained using a small dataset. Additionally, external validation is not possible because no similar dataset is available. To have a general-purpose tool that can be used in daily care, it is necessary to train the framework using a large dataset of individuals of different genders and ages from different countries and cultures to gain the trust of physicians for operational use as an intelligent assistant.

5. Conclusions

To create a medical decision support system that is efficient, reliable, stable, and practical for assisting healthcare providers in the early prediction of Parkinson's disease (PD), this study introduces an advanced framework based on XGBoost (XGB). The proposed approach leverages speech signal analysis to predict PD in a manner that is both accurate and efficient. Compared to other techniques, this method is cost-effective, highly efficient, and non-invasive, eliminating the need for intrusive medical procedures.

In this study, for the first time, a combination of MI, ADASYN, XGB, and TPE (MAXT) methods has been used to solve the problem of early prediction of PD. After preprocessing the speech signal data, 11 features out of 22 available features were selected as suitable features for classifier input using MI. Then, the problem of low data size and an imbalanced dataset was investigated and balanced using the ADASYN technique. In order to increase the accuracy of the classifier, the parameters of the XGB were optimized using the TPE method. The proposed combination has created a reliable framework.

To test the stability of the framework and examine its performance in different situations, the framework was evaluated using the proposed A10S10FCV strategy. In this strategy, the S10FCV process was run 10 times with different random states to assess the stability and validity of the framework accurately. The presented results and discussions demonstrate that the proposed framework, while exhibiting appropriate stability, also achieves higher efficiency than all the other works. It is recommended that the A10S10FCV strategy be used by all researchers/practitioners to evaluate the performance of tools, especially when the data volume is small.

The proposed results showed that selecting a proper algorithm for each module at each step of designing ML tools could have a significant impact on the final tools performance.

In future work, we will evaluate the MAXT framework performance on other speech datasets. Additionally, we aim to develop a web-based application that can receive the patient's voice through a Microphone, extract features, and utilize the proposed MAXT framework to determine whether the patient is healthy or sick.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

- [1] J. S. Almeida *et al.*, "Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques," *Pattern Recognition Letters*, vol. 125, pp. 55-62, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865519301163>.
- [2] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, "Automatic detection of Parkinson's disease based on acoustic analysis of speech," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 148-158, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0952197618302045>.
- [3] S. Srinivasan, P. Ramadass, S. K. Mathivanan, K. Panneer Selvam, B. D. Shivahare, and M. A. Shah, "Detection of Parkinson disease using multiclass machine learning approach," *Scientific Reports*, vol. 14, no. 1, p. 13813, 2024. [Online]. Available: <https://www.nature.com/articles/s41598-024-64004-9>.
- [4] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of Parkinson's disease: a review of literature," *Frontiers in aging neuroscience*, vol. 13, p. 633752, 2021. [Online]. Available: <https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2021.633752/full>.
- [5] M. A. Islam, M. Z. H. Majumder, M. A. Hussein, K. M. Hossain, and M. S. Miah, "A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets," *Heliyon*, vol. 10, no. 3, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405844024015007>.
- [6] M. Beriich, A. Ouhmida, Z. Alouani, S. Saleh, B. Cherradi, and A. Raihani, "Advancing Parkinson's Disease Detection: A Review of AI and Deep Learning Innovations," in *2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 2025: IEEE, pp. 1-10. [Online]. Available: https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=202502252256845613.
- [7] A. Reddy *et al.*, "Artificial intelligence in Parkinson's disease: Early detection and diagnostic advancements," *Ageing research reviews*, vol. 99, p. 102410, 2024. [Online]. Available: <https://ejce.org/index.php/ejce/article/view/488>.
- [8] A. H. Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," *Nature reviews neuroscience*, vol. 18, no. 7, pp. 435-450, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28592904/>.
- [9] W. Poewe *et al.*, "Parkinson disease," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1-21, 2017.
- [10] E. R. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloem, "The emerging evidence of the Parkinson pandemic," *Journal of*

- Parkinson's disease*, vol. 8, no. s1, pp. S3-S8, 2018. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/30584159/>.
- [11] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: a review," *Jama*, vol. 323, no. 6, pp. 548-560, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32044947/>.
 - [12] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *The Lancet*, vol. 397, no. 10291, pp. 2284-2303, 2021. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/33848468/>.
 - [13] Y. Ben-Shlomo, S. Darweesh, J. Llibre-Guerra, C. Marras, M. San Luciano, and C. Tanner, "The epidemiology of Parkinson's disease," *The Lancet*, vol. 403, no. 10423, pp. 283-292, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/38245248/>.
 - [14] A. Zolin, H. Ooi, M. Zhou, C. Su, F. Wang, and H. Sarva, "Liver fibrosis associated with more severe motor deficits in early Parkinson's disease," *Clinical Neurology and Neurosurgery*, vol. 252, p. 108861, 2025. [Online]. Available: <https://scholar.google.com/citations?user=P4PgpD4AAAAJ&hl=en>.
 - [15] R. Lamba, T. Gulati, and A. Jain, "A hybrid feature selection approach for parkinson's detection based on mutual information gain and recursive feature elimination," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 10263-10276, 2022. [Online]. Available: <https://www.springerprofessional.de/en/a-hybrid-feature-selection-approach-for-parkinson-s-detection-ba/20046808>.
 - [16] R. Kardan, M. Nazari, J. Hemmati, A. Ahmadi, and M. Ashab, "A Novel Therapeutic Strategy for Parkinson's Disease based on the Gut Microbiota: A Review Article," (in eng), *Scientific Journal of Kurdistan University of Medical Sciences*, Review vol. 29, no. 3, pp. 127-138, 2024, doi: 10.61186/sjku.29.3.11.
 - [17] D. Aarsland *et al.*, "Parkinson disease-associated cognitive impairment," *Nature reviews Disease primers*, vol. 7, no. 1, p. 47, 2021. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/34210995/>.
 - [18] Z. Yang *et al.*, "Optimizing parkinson's disease prediction: A comparative analysis of data aggregation methods using multiple voice recordings via an automated artificial intelligence pipeline," *Data*, vol. 10, no. 1, p. 4, 2025. [Online]. Available: <https://www.mdpi.com/2306-5729/10/1/4>.
 - [19] M. N. Kadhim, D. Al-Shammary, and F. Sufi, "A novel voice classification based on Gower distance for Parkinson disease detection," *International Journal of Medical Informatics*, vol. 191, p. 105583, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386505624002466>.
 - [20] İ. Cantürk and O. Günay, "Investigation of scalograms with a deep feature fusion approach for detection of Parkinson's disease," *Cognitive Computation*, vol. 16, no. 3, pp. 1198-1209, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s12559-024-10254-8>.
 - [21] N. Patel¹, R. Srividhya, P. E. Linda, and S. Rajesh¹, "Parkinson's Insight: Leveraging CNN and LSTM Networks for Enhanced Diagnostic Accuracy," in *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, 2025, vol. 189: Springer Nature, p. 157. [Online]. Available: <https://www.atlantispress.com/proceedings/computatia-25/126010054>.
 - [22] Z. K. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Medical hypotheses*, vol. 138, p. 109603, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32028195/>.
 - [23] G. Pahuja and T. Nagabhushan, "A comparative study of existing machine learning approaches for Parkinson's disease detection," *IETE Journal of Research*, vol. 67, no. 1, pp. 4-14, 2021. [Online]. Available: https://www.shs-conferences.org/articles/shsconf/ref/2022/09/shsconf_etltc2022_03027/shsconf_etltc2022_03027.html.
 - [24] D. Gupta *et al.*, "Optimized cuttlefish algorithm for diagnosis of Parkinson's disease," *Cognitive systems research*, vol. 52, pp. 36-48, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1389041718301876>.
 - [25] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, "Diagnosis of Parkinson's disease using modified grey wolf optimization," *Cognitive Systems Research*, vol. 54, pp. 100-115, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1389041718308726>.
 - [26] R. Lamba, T. Gulati, H. F. Alharbi, and A. Jain, "A hybrid system for Parkinson's disease diagnosis using machine learning techniques," *International Journal of Speech Technology*, pp. 1-11. [Online]. Available: <https://dl.acm.org/doi/10.4018/IJSI.292027>.
 - [27] W. Thirapanish, P. Kantavat, D. Wanvarie, E. Chuangsuwanich, and P. Punyabukkana, "Evaluating Machine Learning-Based Feature Selection Methods for Diagnosing Parkinson's Disease Under the SVM Framework," in *2024 7th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2024: IEEE, pp. 409-415. [Online]. Available: https://www.researchgate.net/publication/382718440_Evaluating_Machine_Learning-Based_Feature_Selection_Methods_for_Diagnosing_Parkinson's_Disease_Under_the_SVM_Framework.
 - [28] D. Baruah, R. Rehman, P. K. Bora, P. Mahanta, K. Dutta, and P. Konwar, "Performance Evaluation of Classification Algorithms for Parkinson's Disease Diagnosis: A Comparative Study," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 3, pp. 692-712, 2025. [Online]. Available: <https://jeemi.org/index.php/jeemi/article/view/713>.
 - [29] V. J. Kadam and S. M. Jadhav, "Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease," in *Computing, Communication and Signal Processing: Proceedings of ICCASP 2018*: Springer, 2018, pp. 567-581.
 - [30] S. Rahman, M. Hasan, A. K. Sarkar, and F. Khan, "Classification of Parkinson's disease using speech signal with machine learning and deep learning approaches," *European Journal of Electrical Engineering and Computer Science*, vol. 7, no. 2, pp. 20-27, 2023. [Online]. Available: <https://ejece.org/index.php/ejece/article/view/488>.
 - [31] D. Jain, A. K. Mishra, and S. K. Das, "Machine learning based automatic prediction of Parkinson's disease using speech features," in *Proceedings of International Conference on Artificial Intelligence and Applications: ICAIA 2020*, 2020: Springer, pp. 351-362. [Online]. Available: https://www.researchgate.net/publication/342640627_Machine_Learning_Based_Automatic_Prediction_of_Parkinson's_Disease_Using_Speech_Features.
 - [32] H. Reddy, D. V. S. Jagadeesh, P. B. Pati, and B. P. Kn, "Parkinson's Disease Diagnosis from Patients Speech Analysis," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 2024: IEEE, pp. 1-5. [Online]. Available:

- <https://www.semanticscholar.org/paper/Parkinson's-Disease-Diagnosis-from-Patients-Speech-HarshithaReddy-Aryagopal/e24528e3b84b9b2f6d65f7c8821d1bc9a9f16639>.
- [33] H. M. Balaha, A. E.-S. Hassan, R. A. Ahmed, and M. H. Balaha, "Comprehensive multimodal approach for Parkinson's disease classification using artificial intelligence: insights and model explainability," *Soft Computing*, pp. 1-33, 2025. [Online]. Available: <https://dl.acm.org/doi/10.1007/s00500-025-10463-9>.
- [34] D. K. Saha and T. D. Nath, "A lightweight CNN-based ensemble approach for early detecting Parkinson's disease with enhanced features," *International Journal of Speech Technology*, pp. 1-15, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28592904/>.
- [35] B. Akila and J. J. V. Nayahi, "Parkinson classification neural network with mass algorithm for processing speech signals," *Neural Computing and Applications*, vol. 36, no. 17, pp. 10165-10181, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s10462-025-11347-y>.
- [36] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, "Efficient diagnosis system for Parkinson's disease using deep belief network," in *2016 IEEE Congress on evolutionary computation (CEC)*, 2016: IEEE, pp. 1324-1330.
- [37] 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," *Journal of Reliable Intelligent Environments*, pp. 1-10, 2021.
- [38] S. Yadav, M. K. Singh, and S. Pal, "Artificial intelligence model for parkinson disease detection using machine learning algorithms," *Biomedical Materials & Devices*, vol. 1, no. 2, pp. 899-911, 2023.
- [39] R. Alshammri, G. Alharbi, E. Alharbi, and I. Almubark, "Machine learning approaches to identify Parkinson's disease using voice signal features," *Frontiers in artificial intelligence*, vol. 6, p. 1084001, 2023. [Online]. Available: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1084001/full>.
- [40] P. Das and S. Nanda, "Bio-inspired voting ensemble weighted extreme learning machine classifier for the detection of Parkinson's disease," *Research on Biomedical Engineering*, vol. 39, no. 3, pp. 493-507, 2023.
- [41] D. Jain, A. K. Mishra, and S. K. Das, "Machine learning based automatic prediction of Parkinson's disease using speech features," in *Proceedings of International Conference on Artificial Intelligence and Applications: ICAIA 2020*, 2021: Springer, pp. 351-362.
- [42] P. Das, S. Nanda, and G. Panda, "Automated improved detection of Parkinson's disease using ensemble modeling," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, 2020: IEEE, pp. 1-5. [Online]. Available: <https://www.proceedings.com/content/057/057976webtoc.pdf>.
- [43] H. H. B. Y. G. EA and S. A. Li, "adaptive synthetic sampling approach for imbalanced learning 2008," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) Hong Kong Piscataway: IEEE*, vol. 13221328.
- [44] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.
- [45] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295-316, 2020.
- [46] N. R. Baqer and P. Rashidi-Khazaei, "Residential Building Energy Usage Prediction Using Bayesian-Based Optimized XGBoost Algorithm," *IEEE Access*, 2025. [Online]. Available: <https://ieeexplore.ieee.org/iel8/6287639/10820123/10900361.pdf>.
- [47] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *Nature Precedings*, pp. 1-1, 2008. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3051371/>.
- [48] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *International conference on machine learning*, 2013: PMLR, pp. 115-123. [Online]. Available: <https://proceedings.mlr.press/v28/bergstra13.html>.
- [49] M. H. J. Ali H. Al-Fatlawi, Sai Ho Ling, "Efficient Diagnosis System for Parkinson's Disease Using Deep Belief Network," *2016 IEEE Congress on evolutionary computation (CEC)*, vol. 2016 Jul 24, pp. 1324-1330, 2016.
- [50] V. J. Kadam and S. M. Jadhav, "Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease," in *Computing, Communication and Signal Processing: Proceedings of ICCASP 2018*, 2019: Springer, pp. 567-581. [Online]. Available: <https://dl.acm.org/doi/abs/10.1007/s00521-021-05741-0>.
- [51] N. Arasavali, R. Challapalli, J. Jayalakshmi, K. Kasireddy, C. Moturu, and P. Poornapriya, "Parkinson's Classification based on Vocal Features using a Hybrid DNN Multi-Layered LSTM Model," in *2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 2024: IEEE, pp. 1-5.