

# An ensemble technique to predict Parkinson's disease using machine learning algorithms

Nutan Singh<sup>\*</sup>, Priyanka Tripathi

Department of Computer Applications, National Institute of Technology Raipur, Chhattisgarh, 492010 India



## ARTICLE INFO

**Keywords:**

Parkinson's disease  
Voice disorder  
Machine learning  
Feature selection  
SMOTE  
Ensemble and hyper tuning

## ABSTRACT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting motor and non-motor symptoms. Its symptoms develop slowly, making early identification difficult. Machine learning has a significant potential to predict Parkinson's disease on features hidden in voice data. This work aimed to identify the most relevant features from a high-dimensional dataset, which helps accurately classify Parkinson's Disease with less computation time. Three individual datasets with various medical features based on voice have been analyzed in this work. An Ensemble Feature Selection Algorithm (EFS) technique based on filter, wrapper, and embedding algorithms that pick highly relevant features for identifying Parkinson's Disease is proposed, and the same has been validated on three different datasets based on voice. These techniques can shorten training time to improve model accuracy and minimize overfitting. We utilized different ML models such as K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Machine (SVM), Bagging Classifier, Multi-Layer Perceptron (MLP) Classifier, and Gradient Boosting. Each of these models was fine-tuned to ensure optimal performance within our specific context. Moreover, in addition to these established classifiers, we proposed an ensemble classifier is found on a high optimal majority of the votes. Dataset-I achieves classification accuracy with 97.6 %, F<sub>1</sub>-score 97.9 %, precision with 98 % and recall with 98 %. Dataset-II achieves classification accuracy 90.2 %, F<sub>1</sub>-score 90.2 %, precision 90.2 %, and recall 90.5 %. Dataset-III achieves 83.3 % accuracy, F<sub>1</sub>-score 83.3 %, precision 83.5 % and recall 83.3 %. These results have been taken using 13 out of 23, 45 out of 754, and 17 out of 46 features from respective datasets. The proposed EFS model has performed with higher accuracy and is more efficient than other models for each dataset.

## 1. Introduction

Parkinson's Disease (PD) is a neurological condition that predominantly affects individuals over the age of 60, significantly disrupting their lives (Parkinson disease, 2023; Poewe et al., 2017). With longer life expectancies, the global population of senior individuals is projected to increase by 34 % by 2050 (Ibarra-Gutiérrez et al., 2023). As neurodegenerative disorders are strongly associated with ageing, the incidence and prevalence of Parkinson's Disease are expected to rise correspondingly (Poewe et al., 2017). Numerous Parkinson's disease datasets have been derived from medical studies, often utilizing force sensors worn on patient's feet(gait) or voice analysis to capture symptoms (Liu et al., 2022) The primary motor symptoms of Parkinson's Disease include tremors, rigidity, bradykinesia (delayed movement), and body instability (Devarajan and Ravi, 2019). Notably, Parkinson's Disease affects males 1.5 times more frequently than females, and the current global

count of individuals living with PD stands at 10 million, projected to increase to 1.2 million by 2023. Each year in the United States alone, approximately 60,000 people are diagnosed with PD, with a higher incidence among individuals over the age of 50. Recent years have seen a 4 % identification rate of PD cases in individuals under 50 (Mamun et al., 2022). Patients can self-evaluate their health from home via remote voice data collecting, decreasing the requirement for frequent hospital visits and the risk of virus infections during an epidemic.

Additionally, early disease identification and treatment are made possible by healthcare professional's remote observation (Yuan et al., 2023). Early detection of Parkinson's Disease holds great scientific interest due to its progressive nature and the potential to enhance patient quality of life. Moreover, early diagnosis can aid in anticipating the onset of other neurodegenerative disorders, given the similarity of their symptoms. Artificial intelligence (AI) approaches, particularly machine learning (ML), have gained considerable attention in medical

\* Corresponding author.

E-mail address: nsingh.phd2021.mca@nitrr.ac.in (N. Singh).

**Table 1**

Literature review of each dataset.

Dataset details	Author(s) name and year (s)	Methodology	Feature selection	Performance result (%)	Future work
<b>Dataset-I (UCI Machine Learning Repository: Parkinsons Data Set 2022)</b>	Little et al. (2009)	Pre-selection filter + exhaustive search + SVM	Bootstrap (50 features)	Accuracy 91.4 %	Using these techniques on voice signals captured in acoustic settings is more characteristic of real-world telemonitoring applications.
	Shahbaba and Neal (2009)	Dirichlet process mixtures + MNL, DT, SVM	CV fold ( $k = 5$ )	87.7 %	Introduced a new nonlinear classification model.
	Das (2010)	Neural Networks, DMneural, Regression and Decision Tree	Comparison with kernel SVM using SAS	Achieved highest score by Neural Networks 92 %	Advances in the treatment of Parkinson's disease.
	Sakar and Kursun (2010)	Mutual information-based features + SVM	Bootstrap selection + SVM (50 replicates)	92.75 %	Improved generalizations to unseen test data.
	Sharma et al. (2019)	k-NN, RF and DT.	Modified Grey Wolf Optimization	93.87 %	To improve the accuracy of early detection, merge the HandPD, SpeechPD, and voice datasets models.
	Devarajan and Ravi (2019)	FKNN—CBR classifier	All features	94.87 %	The dimensionality will be reduced while classification accuracy will be maintained by using the more significant vocal components.
	Sakar et al. (2019)	mRMR-50's Tunable Q-factor Wavelet Transform + SVM	Severity prediction using SVM (RBF kernel)	84 %	Assemble a rule basis based on the expertise of the expert to improve proposals for the care of Parkinson's patients significantly.
	Fang (2022)	k-NN classifier	-k-NN + entropy-based method	93 %	To create a reliable PD telemonitoring system, it is possible to predict the Unified Parkinson's Disease Rating Scale (UPDRS) score of PD patients using a classification issue.
	Lamba et al. (2022)	Naive Bayes, k-NN, RF	-Hybrid MIRFE algorithm (MIRFE-XGBoost)	95.58–95.47 %	An entropy-based method was used to improve accuracy.
		mRMR-50 and the TQWT	SVM (RBF kernels)	86 % (with 50 features)	To create a decision support system for Parkinson's diagnosis, the scientists intend to apply deep learning algorithms to handwritten images.
<b>dataset-II (UCI Machine Learning Repository: Parkinson's Disease Classification Data Set 2022)</b>	El-Hasnony I et al. (2020)	119 features  The fog infrastructure provides real-time data processing and analysis and overcoming the cloud limitations	ANFIS+PSOGWO	87.5 %	The Unified Parkinson's Disease Rating Scale (UPDRS) score of PD patients can be predicted using the TQWT technique to create a reliable system.
	Yücelbaş (2020)	Greedy stepwise algorithm-based Simple Logistic hybrid (SLGS)	Feature-based PD and HC prediction	88.71 % (males), 87.15 % (females)	Used a broader range of data sets.
	Al-Sarem et al. (2021)	RF, XGBoost, CatBoost	Diagnosis using various machine learning classifiers	High accuracy	Using modern metaheuristic techniques, such as an optimization algorithm.
		SHAP-LightGBM model and LightGBM	150 features selected	91.62 % (classification accuracy), 94.5 % ( $F_1$ -score)	Future research must examine the proposed hybrid system with greater information.
	Hawi et al. (2022)	Set A-13 Set B-10 Set C-20	Long-term acoustic properties and MFCC Majority of voting	84.88 %	Alternative ensemble methods could be used to acquire the most crucial features.
	Sabeena et al. (2022)	(MFCC + Wavelet + Concat), FCBi-LSTM classifier (%)	Used OBEFSs (optimization-based ensemble feature selections)	98.77 %	Utilizing the significant features acquired and deep learning techniques.
	Bao et al. (2022)	New SSCL method	Diagnostic performance improvement	83.8 % (accuracy), 82.5 % (recall), 85 % (specificity)	Work with a more efficient feature selection technique
					The developed approach still has to be confirmed for early PD identification.
					The suggested CNN's parallel convolution layers allow for simultaneous input of several data types into the network.
					Use the multimodal data for PD categorization.

(continued on next page)

**Table 1 (continued)**

Dataset details	Author(s) name and year (s)	Methodology	Feature selection	Performance result (%)	Future work
	Liu et al. (2022)	SHAP-LightGBM	50 features	91.62 %	The proposed model's limitations must be solved to make remote patient monitoring possible.
<b>dataset-III (UCI Machine Learning Repository: Parkinson Dataset with replicated acoustic features Data Set 2022)</b>	Naranjo et al. (2016)	Bayesian classification	waveform matching algorithm	75 %(mean)	Based on mobile terminal devices, PD detection
	Nahar et al. (2021)	gradient boosting, extreme gradient boosting, bagging and Extra Tree Classifier	Bagging with recursive feature elimination	82.35 %	Assemble additional information to identify Parkinson's illness and use several techniques, including deep learning. Also assist with Parkinson's disease. And Data from MRI
	Dhar (2022)	LGBM algorithm	Author Proposed MI-AE-GOLGBM	84.17 %	To boost the performance of our suggested strategy, include the previously discussed methodologies and classify PD.

diagnostics owing to their ability to process large volumes of data and generate accurate predictions ([Saravanan et al., Oct. 01, 2022](#)). ML models have shown promise in analyzing diverse medical data types, including text, voice, and images, to predict outcomes such as disease presence, absence, or neutrality. ML models trained on extensive electronic medical datasets from wearable devices have proven effective in accurately and efficiently diagnosing Parkinson's disease ([Raheem, 2022](#)). However, the efficiency of ML models relies heavily on the availability of substantial training data to mitigate overfitting ([Fayed and Atiya, 2019](#)). We aim to bridge the gap between efficient and accurate diagnosis ([Hussain et al., 2023; Chang](#)). Effective intervention explores sampling techniques, such as oversampling or undersampling, to balance the datasets to improve performance and address bias towards the majority class ([Shastry, 2023; Ali et al., 2023](#)). Finding computational efficiency for the complexity of high-dimensional datasets requires efficient feature selection. The motivation lies in addressing computational challenges through an Ensemble Feature Selection Algorithm (EFSA) ([Ali et al., 2023](#)). The primary innovative approach to the proposed EFSA integrates filter, wrapper, and embedding algorithms, aiming to streamline feature identification and enhance classification accuracy while minimizing computation time.

### 1.1. Motivations emphasis

Our study is motivated by the pressing challenges in Parkinson's Disease (PD) diagnosis, particularly the difficulty in early identification and the need for efficient diagnostic tools. With the global increase in the senior population and challenges posed by epidemics, there is a growing demand for remote monitoring and early intervention. Our work seeks to contribute to addressing these challenges by leveraging machine learning on voice data for effective PD diagnosis.

### 1.2. Contribution

- Introduction of a novel ensemble-based technique for feature selection and classification in Parkinson's Disease diagnosis.
- Utilization of filter, wrapper, and embedding methods to pinpoint vital features for Parkinson's Disease diagnosis. Evaluation of classifier performance with and without feature selection.
- Implementation of SMOTE for dataset re-balancing, ensuring fair representation of different classes.
- Application of machine learning classification models with hyper-tuning parameter optimization. Recording computational training and testing times for efficiency assessment. We also calculate 10-fold cross-validation accuracy to ensure robust evaluation.

- A noteworthy achievement of high scores in performance metrics with the same number of parameters as existing techniques demonstrating efficiency.

This study utilizes eight machine learning algorithms employing a soft voting classifier to showcase an improved approach ([Kumari et al., 2021](#)). Specifically, the Parkinson's voice dataset is used for comparative analysis to achieve the research objectives. The development of a diagnostic system for Parkinson's Disease would provide significant benefits to both patients and doctors by enabling timely and efficient analysis. Three voice recording datasets from healthy people and PD patients were used to train the machine learning classifiers used in this work. Pitch, jitter, and shimmer are only a few of the acoustic aspects from the speech recordings retrieved for the diagnosis utilizing the suggested EFSA approach ([Little et al., 2009](#)). This work aims to create an effective model to investigate the potential of ensemble-based feature selection algorithms and classification in telemedicine applications. These algorithms have shown promising results in PD diagnosis. The paper is structured as follows: In Section II, a review of the study-related research is given. Section III contains a discussion of the suggested model and the dataset that was used. The characteristics analysis is covered in Section IV, which also looks at how the target variable and the features are correlated. Section V presents the experimental findings for the base and ensemble models. Section VI provides a comparative analysis of the suggested model with other works, and Section VII wraps up the work with ideas for future work.

### 2. Literature review

Numerous studies have been conducted to forecast Parkinson's Disease in a patient. Parkinson's Disease (PD) is the second most prevalent form of neurodegenerative disease that affects speaking and accurate voice. There is no indication of a hereditary component to illness after age 50. The onset of Parkinson's Disease ranges from 50 to 95 years of age ([Hawi et al., 2022](#)). Studies and investigations into the effectiveness of intensive voice therapy are now being conducted to enhance functional communication. Various researchers have performed a number of notable attempts to diagnose Parkinson's disease. The research that has been done to identify Parkinson's disease using subject voice samples will be briefly reviewed here.

[Lamba et al. \(2022\)](#) presented a study that explored the effectiveness of Naive Bayes, k-NN, and Random Forest (RF) algorithms for diagnosing Parkinson's disease. Their methodology involved a Hybrid MIRFE algorithm, specifically MIRFE-XGBoost, showcasing a fusion approach for feature extraction and selection. The performance results ranged between 95.58 % and 95.47 %. In their future work, the authors outlined plans to incorporate deep learning techniques, specifically on

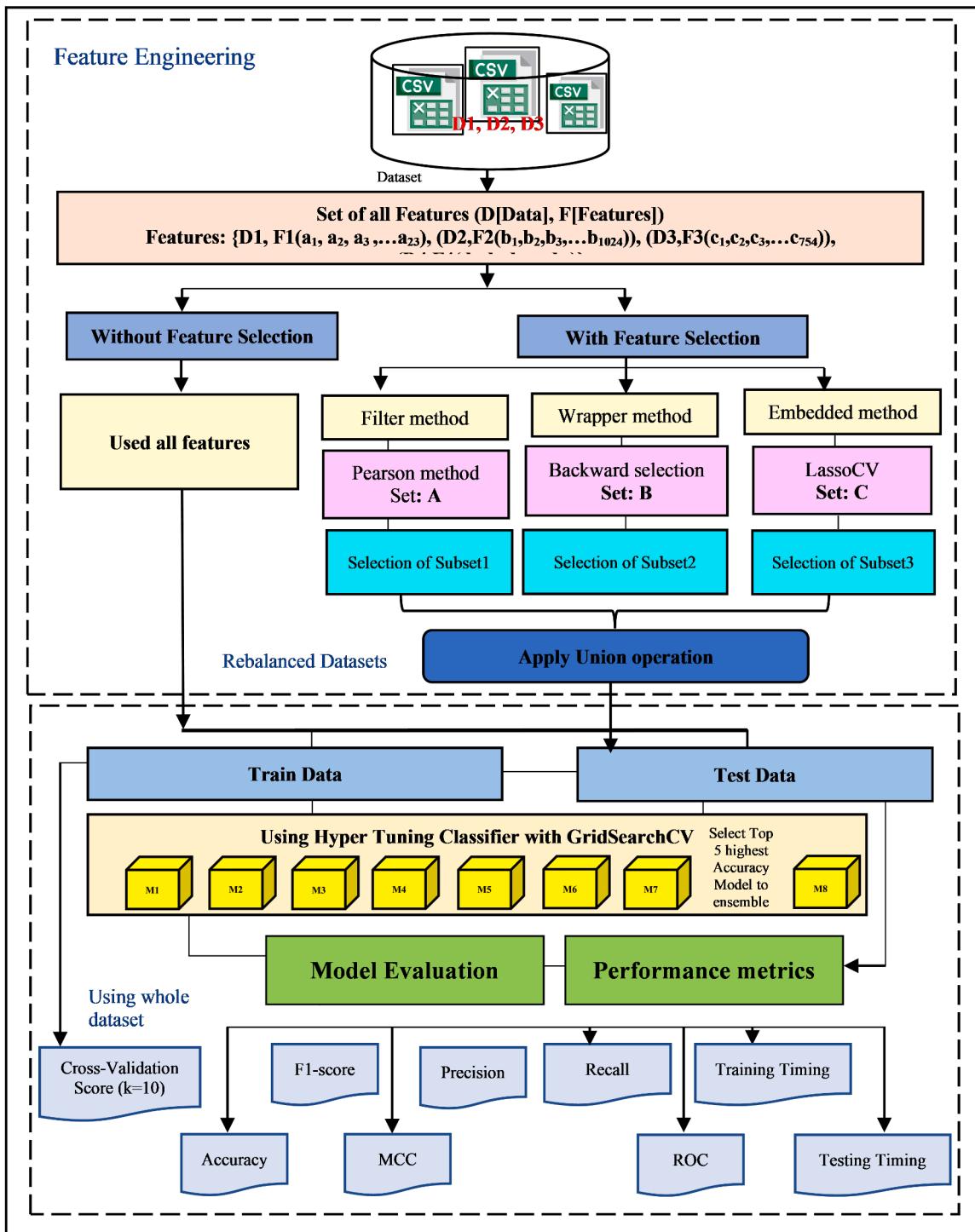


Fig. 1. Proposed framework.

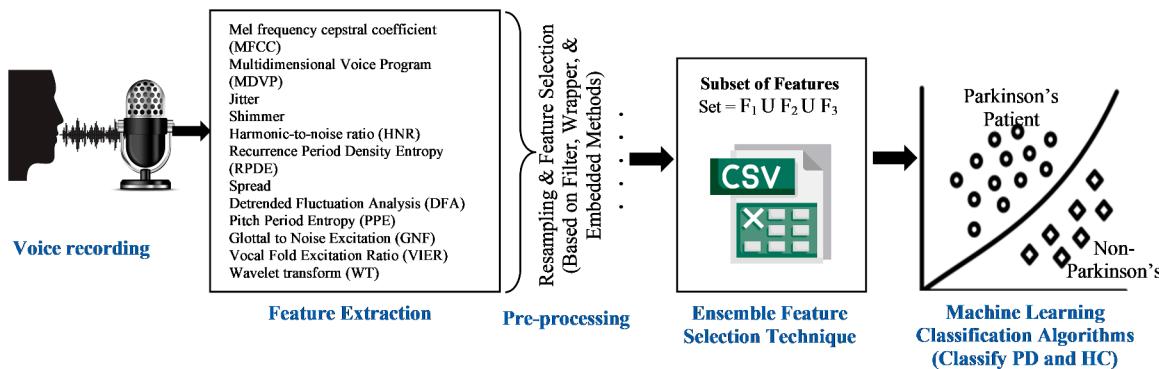
handwritten images, to develop a decision support system for Parkinson's diagnosis. However, they acknowledged a limitation in their proposed system's inability to identify disease progression, given the slow progression of Parkinson's.

Dhar (2022) contributed to Parkinson's disease classification using a distinctive feature set (40PD, 40HC) and the MI-AE-GOLGBM methodology, achieving an 84.17 % performance. Future work emphasizes integrating methodologies for enhanced classification performance.

Liu et al. (2022) introduced a study that focused on remote monitoring of patients using a feature set of 50 features. Their methodology incorporated SHAP-LightGBM, demonstrating a novel feature selection

and modelling approach. The performance result of their proposed model was reported at 91.62 %. The identified limitation pertained to the challenges associated with remote monitoring, indicating an awareness of the contextual constraints within which their model would be implemented.

Shastry (2023), the relevant features were extracted Using feature importance techniques such as F-PER, F-MDI, and correlation method (F-CORR). The normalized training data comprising the selected features were fed into different ML models such as LR, AB, DT, RF, k-NN, GB, XGB, NB, SVM, MLP, ERT, and CB. These trained models were then tested on the test data. The best-performing ML models were then

**Algorithm 2:** The EFSA algorithm**Input:** DI, R, CDADI – a dataset with real-valued features  $F = \{f_1, f_2, \dots, f_k ; i=1,2,3\}$  and a categorical class variable c

R – the feature relevance threshold separating relevant from irrelevant features

CDA<sub>i</sub> – a community detection algorithm {i=1,2,3}**Output:** Final highest accuracy of proposed model.

```
CDA1 //Correlation with output variable and Selecting highly correlated features
    relevant_features = cor_target[cor_target>0.4]
return relevant_features A
```

```
CDA2 //Backward Elimination based wrapper
    cols = list(X.columns)
    pmax = 1
    while (len(cols)>0):
        p = []
        X_I = X[cols]
        X_I = sm.add_constant(X_I)
        model = sm.OLS(y,X_I).fit()
        p = pd.Series(model.pvalues.values[1:],index = cols)
        pmax = max(p)
        feature_with_p_max = p.idxmax()
        if(pmax>0.05):
            cols.remove(feature_with_p_max)
        else:
            break
selected features B
```

```
CDA3 // Feature selection using SelectFromModel and LassoCV
    clf = LassoCV(random_state=42).fit(X, y)
    importance = np.abs(clf.coef_)
    idx_features = (-importance).argsort()[:10]
    C = np.array(X.columns.to_list())[idx_features]
return C
print("A U B U C :", A.union(B, C))
Return relevant features PEF
```

Fig. 2.1 Pseudo code for proposed EFSA Algorithm.

Fig. 2. Process of ensemble feature selection algorithm (EFSA).

combined. The ensemble of k-NN and GB classifiers, the NNB classifier, was found to perform best with respect to multiple performance metrics.

Hussain et al. (2023) using an ensemble classifier-based method to identify PD using ML algorithms and highlight the importance of ML algorithms in assisting with early detection and diagnosis of PD using augmented data. The reported performance result for their approach was 96.6 %. In the future, the study should also discover advanced machine-learning models to make them more efficient and accurate (Table 1).

Further research in this area could shed more light on the comparative effectiveness of these FS methods and their applicability across various technologies. A few publications analyze medical traits only to identify Parkinson's illness. This work focuses on identifying more effective features and efficient disease prediction models. We gathered a voice dataset for Parkinson's Disease from the machine learning repository at UCI. A comparison of existing models has been given in Table 8.

### 3. Proposed methodology

In our proposed framework, we have introduced a predictive model for Parkinson's Disease based on voice data. We divided the analysis into two parts using three datasets from the UCI repository. The first involves classification without feature selection, while the second employs an ensemble feature selection comprising filter, wrapper, and embedded methods. The large dimensionality of datasets is a primary problem for ML prediction models. Therefore, feature selection plays a significant role in efficient model prediction. The details of the dataset used have been discussed in detail in section A. Subsequently, we balanced the dataset using the SMOTE technique. Following dataset re-balancing, we applied optimized machine learning classifiers to distinguish between individuals with PD and healthy subjects.

Additionally, we used a proposed classifier employing a voting classifier. The final step involves a comprehensive assessment of each model's performance. This framework is designed to enhance the accuracy and efficiency of Parkinson's Disease prediction based on voice data (Figs. 1 and 2).

**Table 2**  
Details of benchmark dataset used in the proposed model.

Characteristics	Dataset-I	Dataset-II	Dataset-III
Dataset Repository	UCI Machine Learning Repository: Parkinsons Data Set (2022)	UCI Machine Learning Repository: Parkinson's Disease Classification Data Set (2022)	UCI Machine Learning Repository: Parkinson Dataset with replicated acoustic features Data Set (2022)
Type	Voice signal	- Various speech signal - C. Okan Sakar a et al. - 2018	Voice signal
Author's and year	-Max Little -2008	- Carlos J. - 2019	
Number of Attributes	23	754	46
Number of Instances	197	756	240
No. of PD.	23	188 (107 men and 81 women)	40
No. of HC.	8	64(23 men and 41 women)	40
Balanced Associated Tasks	No Classification	No Classification	Yes Classification

#### 4. Methods

##### A. Benchmark Dataset

In this work, three voice-based datasets from the UCI archive have been used. **Table 2** displays the specifics of each collection.

##### B. Feature selection

EFSA is an approach proposed by the author as part of this research. It involves an ensemble-based feature selection method that combines

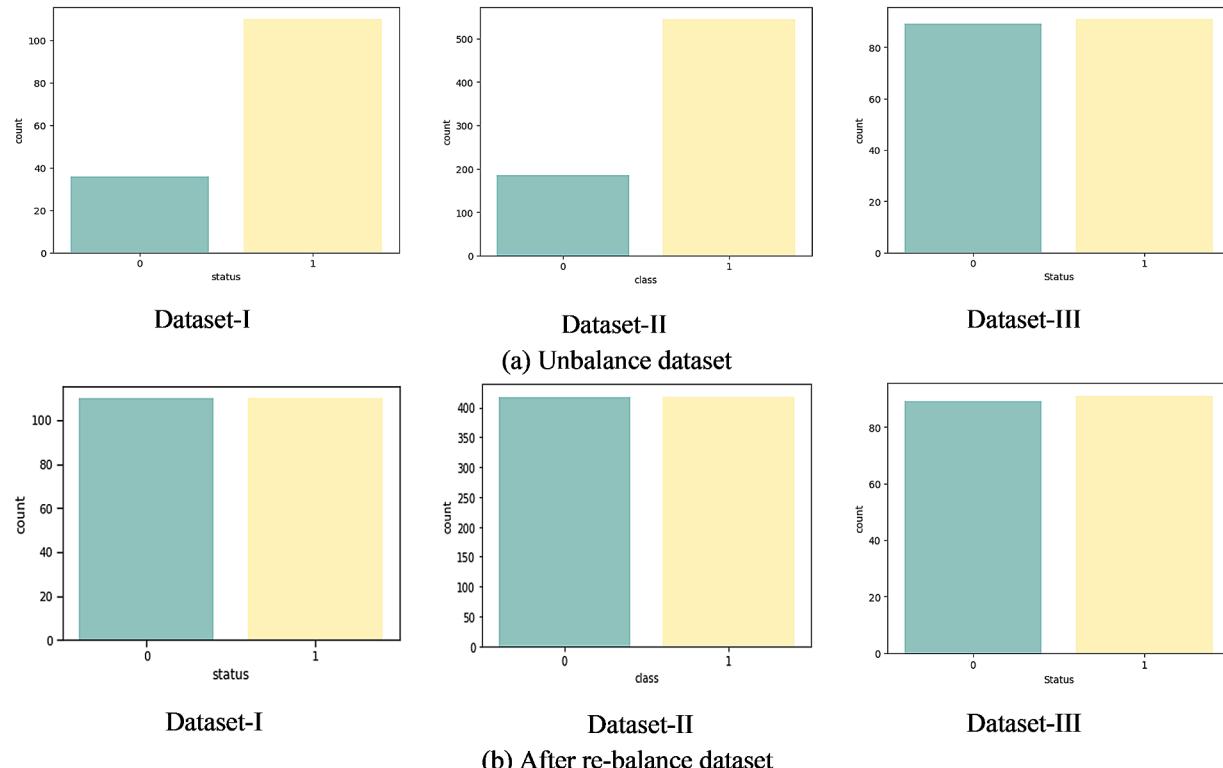
the results of three individual feature selection algorithms: filter (Pearson correlation), wrapper (backward elimination), and embedded (LassoCV). These algorithms are applied to each dataset independently, and their selected features are combined through a union operation. The rationale behind developing EFSA was to leverage the strengths of different feature selection techniques and create a diverse set of selected features. Combining multiple algorithms allows for identifying the most relevant voice features for Parkinson's disease prediction. By incorporating the steps outlined above, EFSA aims to improve the accuracy and efficiency of the predictive models by focusing on the subset of features that have demonstrated high correlation and ranking.

##### C. Data Pre-processing

Data pre-processing is an essential step in the ML life cycle since it facilitates data analysis and improves the accuracy and computation of the ML algorithms. Due to issues with missing null values, a drop of some non-relevant data and class imbalance in the obtained dataset, we used specific pre-processing techniques. The two most common reasons for null or missing values in medical datasets are the absence of data collection or the practitioner's failure to consider the observation. Most relevant features are transformed by MinMaxScaler () by scaling each feature on the training set. To a predetermined range, this estimator scales and translates each feature separately (-1,1). After pre-processing, each dataset was split into 75:25 for training and Testing sets.

##### i. Data resampling

In this step, we apply the Data resampling method through the SMOTE (Synthetic Minority Over-Sampling Technique) technique because our datasets I and II have imbalanced class distribution. SMOTE is used to balance this unbalanced dataset (Polat, 2019). The Synthetic Minority Oversampling Technique is a component of the oversampling phase shown in **Fig. 3(b)**. By creating synthetic samples from the



**Fig. 3.** (a) and (b) show the dataset is unbalanced and balanced after applying SMOTE.

**Table 3**

Description of datasets.

Dataset	Shape of dataset	Before Pre-processing		After Pre-processing		Applying EFSA	
		Training size	Testing size	Training Size	Testing size	Training size	Testing size
Dataset-I	195,23	146, 22	49, 22	202, 22	59, 22	202, 13	59,13
Dataset-II	756,754	529, 754	227, 754	529, 753	227, 753	567, 45	189, 45
Dataset-III	240, 46	180, 46	60, 44	180, 42	60, 44	180, 17	60, 17

minority class, SMOTE oversampling addresses the problem of class imbalance. Every minority instance is considered by SMOTE, which then selects one of its  $k$  closest minority neighbours and establishes a new minority instance between that neighbour and itself. The class imbalance issue is resolved using this process (Jain et al., 2021). Table 3 shows the no. of training and testing sets used in model classification.

#### D. Machine learning classification algorithm

This work used eight classifiers to classify Parkinson's Disease effectively.  $k$ - nearest neighbours, Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Bagging, MLP Classifier, gradient-boosting and proposed classifiers (Lamba et al., 2022; Fang, 2022; Jain et al., 2021). The top Seven ML classifiers apply hyperparameter tuning to select optimal parameters to classify Parkinson's Disease. One must consider the advantages of improved model performance against the time and computational resources required to optimize hyperparameters after the top 5 highest accuracy-based classifiers were used to integrate the ensemble voting classifier, i.e., our proposed classifier. Tables 2, 5 and 8 in Appendix A show each classifier's optimized parameter.

##### i. Hyperparameters of ML models

Each ML classification model needs one or more parameters to control (efficiency) the classifier's prediction results. Choosing the correct parameter values can be challenging because doing so entails balancing the generality and complexity of the models. The grid search value alters before determining the parameter values based on changing

parameter estimates. A GridSearchCV (Fayed and Atiya, 2019) was done through a grid of performance parameters to find the best ones to optimize the performance of the models taken into consideration for the study. In our work, we proposed an ensemble voting classifier model that combines the top five high accuracy-based models to predict disease. We used a soft voting classifier that takes a weighted average of individual performance on the validation set. Optimization (Sabeena et al., 2022) was employed to decrease prediction time, overprocessing, and error to identify the appropriate hyperparameters (Kaur et al., 2020).

## 5. Experimental and results

### A. Experimental Environment Configuration

In this experiment, we use a Windows 11 Ultimate platform with Python 3.5 and an Intel(R)-i5 processor with a CPU clock rate of 4.2 GHz, 256GB SSD and 16 GB of main memory. The investigation aimed to identify the features that improve the Accuracy of PD prediction (see Fig. 5,8 and 11). 75 % of the training data and 25 % of the test data were used to get the best training and test results (Avuclu and Elen, 2020). We evaluated proposed classifiers with and without feature selection methods for the same datasets and compared their classification performance. The data were presented accurately to assess each classifier's performance. In the end, we examined the experimentation's outcomes in Table 8.

### B. Performance evaluation

---

Algorithm 1 The Proposed Classifier

---

Dataset: Dataset DI {I=1,2,3}

**Inputs data:** Training set T {t1..tp}

Set of p features F = {f1, ..., fp}

Ranking method M(T, F)

**Output:** Final highest accuracy of the proposed model.

---

```

procedure (DI)
    return gait.csv("PD", "HC")
//Using ensemble Feature selection
A select subset of relevant features( PEF)
procedure SPLITG( PEF , table)
    Training_DI, Testing_DI
    return Training_DI, Testing_DI
voting="soft"

M1=neighbors.KNeighborsClassifier(Training_DI, Training_label, Testing_DI)
M2=RandomForestClassifier(Training_DI, Training_label, Testing_DI)
M3=DecisionTreeClassifier( Training_DI, Training_label, Testing_DI)
M4=SVC( Training_DI, Training_label, Testing_DI)
M5= BaggingClassifier( Training_DI, Training_label, Testing_DI)
M6=MLPClassifier(Training_DI, Training_label, Testing_DI)
M7=GradientBoosting(Training_DI, Training_label, Testing_DI)

procedure ENSEMBLE VOTING_MODEL( Training_DI, Training_label, Testing_DI)
Soft_voting_classifier= concatenate( M1,M2,M4,M5,M6,M7, voting='soft')
prediction= soft_voting_classifier.predict( Testing_DI)

```

---

Fig. 4. Proposed classifier model based on the majority of the vote.

**Table 4**  
Performance evaluation for each classifier in Dataset I.

Model	(a) without using feature selection								
	Testing Accuracy (%)	F <sub>1</sub> Score (%)	Precision (%)	Recall (%)	MCC (%)	CVS (10-fold)	AUC (10-fold)	Training Time (sec.)	Testing Time (sec.)
KNN	87.8	88.1	88.7	87.8	0.6903	83.3 ± 0.067	0.94±0.05	0.00151	0.10777
Random-Forest	87.8	87.8	87.8	87.8	0.6689	93.3 ± 0.072	0.98±0.03	0.35404	0.02900
Decision-Tree	83.6	83.7	83.7	83.7	0.5585	92.6 ± 0.062	0.94±0.04	77.41696	0.00199
SVM	93.8	91.6	91.7	91.8	0.77140	96.7 ± 0.031	0.99±0.01	0.01575	0.00192
BaggingClassifier	82.0	82.6	85.2	81.6	0.58193	90.7 ± 0.078	0.97±0.04	26.10523	1.73122
MLPClassifier	80.0	80.5	82.4	79.6	0.51427	84.4 ± 0.082	0.94±0.06	3.19833	0.00198
GradientBoosting	85.7	85.9	86.2	85.7	0.62521	91.5 ± 0.05	0.96±0.05	0.03094	0.00150
Proposed Classifier	93.9	93.8	93.8	93.9	0.83100	96.7 ± 0.064	0.99±0.02	3.44650	0.13317
(b) with ensemble features selection									
KNN	93.9	94.0	94.1	93.9	0.84020	94.1 ± 0.01	0.97±0.03	0.00101	0.00346
Random-Forest	93.9	93.6	94.3	93.9	0.83291	96.3 ± 0.036	0.99±0.02	169.1517	0.0242
Decision-Tree	85.7	85.9	86.2	85.7	0.62521	94.1 ± 0.034	0.93±0.06	29.65520	0.00092
SVM	89.8	89.6	89.6	89.8	0.71726	97.0 ± 0.022	0.97±0.04	0.01053	0.00074
BaggingClassifier	85.7	85.5	85.4	85.7	0.60352	94.8 ± 0.066	0.97±0.05	7.40316	0.54193
MLPClassifier	91.8	91.8	91.8	91.8	0.77927	82.2 ± 0.039	0.97±0.02	2.8029	0.00100
GradientBoosting	93.9	93.6	94.3	93.9	0.83291	94.4 ± 0.037	0.99±0.01	0.15124	0.15124
Proposed Classifier	<b>97.6</b>	<b>97.9</b>	<b>98.0</b>	<b>98.0</b>	<b>0.94474</b>	<b>96.9 ± 0.063</b>	<b>0.99±0.01</b>	<b>116.19069</b>	<b>0.01500</b>

The performance of the proposed model is assessed using five formulas, as shown in Eq. (1), (2, 3, 4), and (5). The following performance metrics are considered while assessing the categorization algorithms (**Mittra and Rustagi**) performance: Accuracy, sensitivity, and specificity (Acc, Sen, Spe, F1-score, ROC, CVS and Mathew's Correlation Coefficient (MCC) (**Chandra et al., 2021**).

### C. Confusion Matrix

A classification algorithm's performance is evaluated using a confusion matrix table. A confusion matrix depicts and summarizes a classification algorithm's performance. In our work, the confusion matrix is displayed in Appendix B in Fig 1 to Fig 3 for without feature selection and Fig 4 to Fig 6 for those with EFSA, where 0 is referred to as healthy and 1 as Parkinson's Disease. The total of true positives (TP) and true negatives (TN) represents the accurate forecast, as seen below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{F1 score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (4)$$

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (5)$$

### D. Cross-Validation Score(CVS) Analysis

In this study, we apply 5-fold and 10-fold cross-validation, two different folding algorithms (**Bradshaw et al., 2023**), to assess the performance of machine learning classifiers. The whole dataset is used for this evaluation. Determining the mean and standard deviation of cross-validation scores for both folding procedures provides significant information about the classifier's generalization abilities and the variability of their performance throughout various folds. Because 10-fold cross-validation benefits model evaluation, we deliberately choose it above 5-fold cross-validation in our paper. With this decision, the dataset will be used more effectively, and performance estimations may become more trustworthy. Appendix B Sections I (without feature selection) and II (with ensemble feature selection) demonstrate the application and discussion of this methodical methodology. We hope to thoroughly assess the classifiers' performance through these sections, illuminating their stability and performance in various scenarios.

### E. Experiment-first for Dataset-I:

This section presents a comparative study of an ensemble classifier and feature selection. Three feature selection algorithms were used based on filter, wrapper, and embedded methods, specifically person correlation, backward elimination, and Lassocv. The proposed classifier and other machine learning classifiers were used to classify data for individuals with Parkinson's Disease and healthy individuals. Eight classifiers were trained, including the suggested classifier with GridSearchCV that provides optimal hyper-tuning parameters, Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Bagging Gradient Boost, and K-Nearest Neighbors (k-NN). Additionally, k-fold cross-validation with  $k = 10$  was used to validate the findings, and the average value of the cross-validation score was computed with their standard deviation. Model performance was evaluated using accuracy, recall, f<sub>1</sub>-score, precision, ROC, and CVS (Cross-validation Score mean with

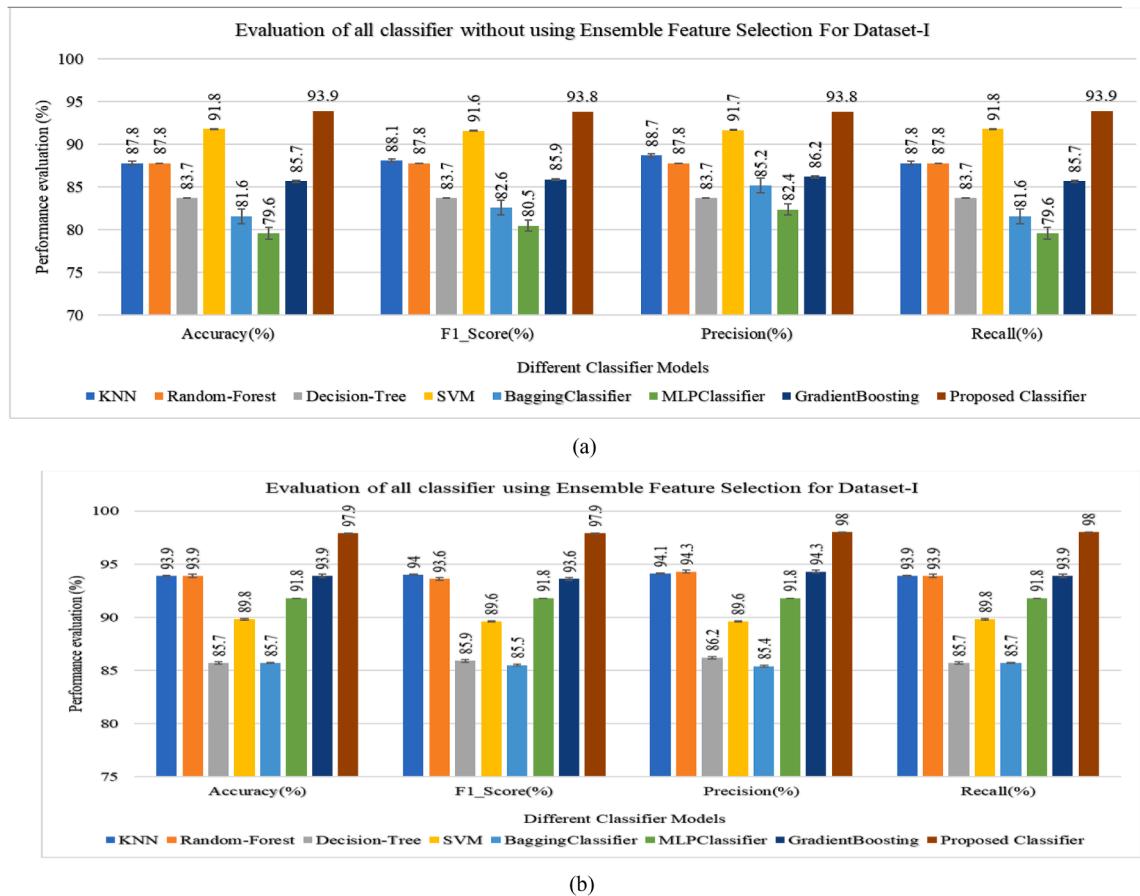


Fig. 5. Comparison of performances (a) without and (b) with Ensemble Feature Selection for dataset-I.

cv=10). Firstly, the Filter-based method was applied for feature selection, using Pearson Correlation, which selected features with a high correlation value of  $\text{cor\_target} > 0.4$ . This resulted in the selection of 11 relevant features. Secondly, the wrapper feature selection method was applied using the Backward Elimination technique to remove features that did not significantly affect the dependent variable, with  $\text{pmax} > 0.05$ . This resulted in the selection of eight significant features. The embedded-based method was used as the third feature selection technique, where LassoCV was used to find the optimal choice of regularization strength, resulting in the selection of the top 10 relevant features. The union operation was then applied to the three subset

features to create an EFSA method, selecting 13 significant and relevant features for a more efficient, fast, and accurate model. The proposed classification approach effectively used ensemble characteristics with high accuracy scores. The ensemble technique that selected just 13 features outperformed the other two feature selection methods. The features selected using EFSA techniques are displayed in Appendix A, Table 4. The performance of classifiers was assessed after selecting features using EFSA techniques. The proposed classifier used hyper-parameter tuning based on the GridSearchCV algorithm to optimize the best parameters for machine learning classifiers. Table 8 displays the performance of the proposed classifier with all features and reduced

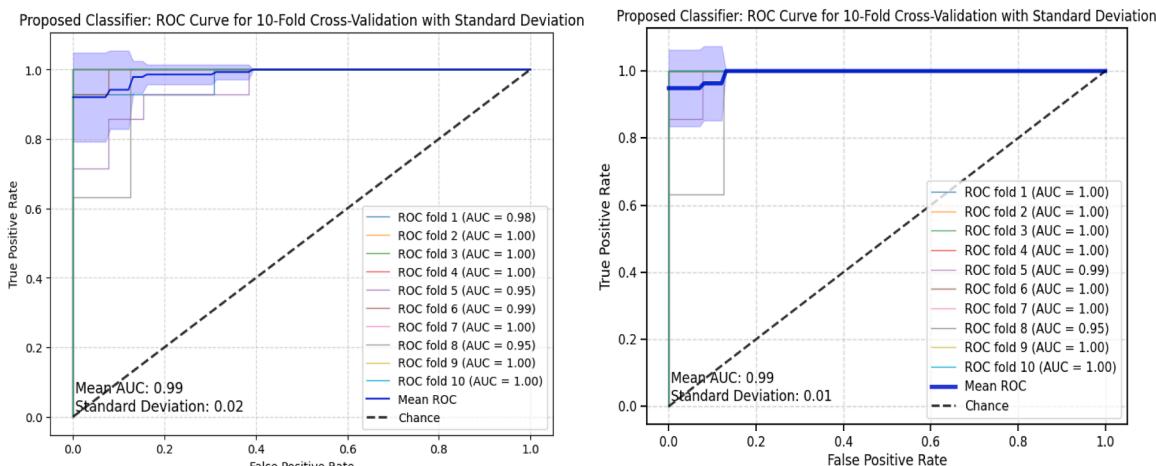
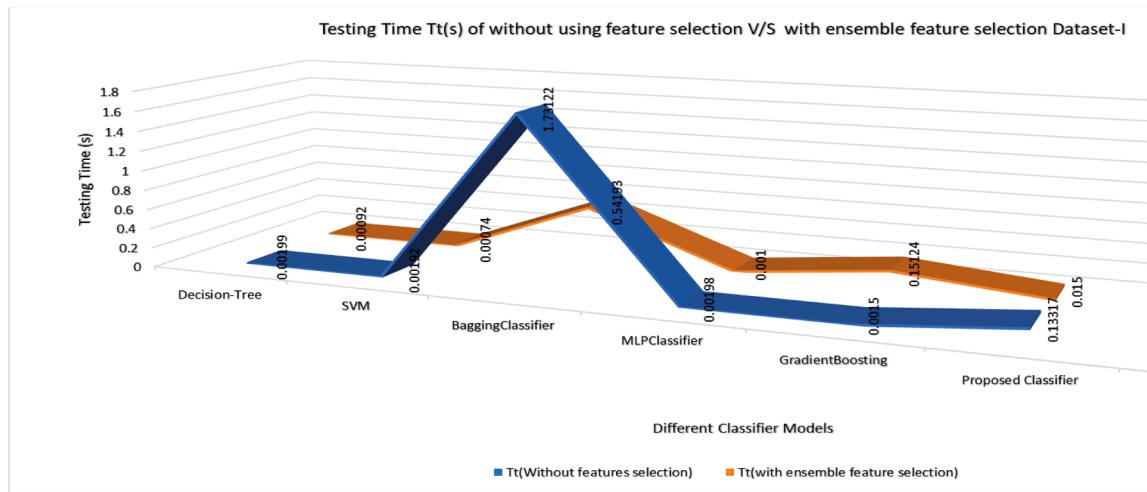


Fig. 6. ROC curve Proposed classifier using 10-fold Cross-Validation in dataset-I (a) without and (b) with ensemble feature selection.



**Fig. 7.** Testing Time without and with ensemble feature selection for Dataset-I.

**Table 5**  
Performance evaluation for Dataset-II.

Model	(a) without ensemble feature selection								
	Testing Accuracy (%)	F <sub>1</sub> -Score (%)	Precision (%)	Recall (%)	MCC (%)	CVS (10-fold)	AUC (10-fold)	Training Time (sec.)	Testing Time (sec.)
KNN	77.8	79.3	84.2	77.8	0.517	86.1 ± 0.034	0.95±0.02	0.00550	0.0540
Random-Forest	86.8	86.2	86.2	86.8	0.600	95.2 ± 0.018	0.99±0.01	0.32385	0.0050
Decision-Tree	78.8	77.8	77.8	77.8	0.367	88.1 ± 0.028	0.89±0.03	285.35831	0.0020
SVM	84.7	83.8	83.8	84.7	0.529	95.2 ± 0.02	0.99±0.00	1.36174	0.0734
BaggingClassifier	76.2	76.6	78.2	75.7	0.374	87±0.036	0.94±0.02	13.56044	1.0339
MLPClassifier	84.0	84.0	83.9	84.1	0.541	93.7 ± 0.01	0.99±0.01	0.9221	0.0009
GradientBoosting	81.5	83.1	84.0	82.5	0.541	87.3 ± 0.058	0.95±0.03	2.06418	0.0029
Proposed Classifier	88.9	88.3	88.6	88.9	0.663	95.8 ± 0.011	0.99±0.00	12.45948	0.0967
Model	(b) with ensemble feature selection								
	Testing Accuracy (%)	F <sub>1</sub> -Score (%)	Precision (%)	Recall (%)	MCC (%)	CVS (10-fold)	AUC (10-fold)	Training Time (sec.)	Testing Time (sec.)
KNN	85.2	85.6	86.4	85.2	0.6083	91.9 ± 0.013	0.93±0.02	0.00121	0.0551
Random-Forest	86.2	86.0	86.2	86.8	0.5963	93.4 ± 0.013	0.99±0.01	139.29991	0.0102
Decision-Tree	81.0	81.5	82.3	81.0	0.4946	86.6 ± 0.006	0.90±0.03	40.88099	0.0009
SVM	84.7	83.6	83.7	84.7	0.5240	90.3 ± 0.026	0.98±0.01	0.24740	0.0119
BaggingClassifier	73.0	74.5	77.4	73.0	0.3466	83.3 ± 0.022	0.91±0.01	3.46079	0.2430
MLPClassifier	86.8	86.5	87.1	86.2	0.6293	94.0 ± 0.014	0.99±0.01	11.59677	0.0024
GradientBoosting	87.8	87.6	87.4	87.8	0.6409	91.6 ± 0.014	0.97±0.02	1.14074	0.0009
Proposed Classifier	90.5	90.2	90.2	90.5	0.7182	94.3 ± 0.007	0.99±0.01	148.14575	0.0675

feature subsets chosen via EFSA approaches.

In the first experiment on dataset D1 without using the feature selection technique, the proposed classifier outperformed seven classifiers with 93.9 % accuracy, f<sub>1</sub>-score of 93.8 %, precision of 93.8 %, and recall of 93.9 %. In the second experiment on the same dataset, D1, using an EFSA algorithm with the proposed classifier, the proposed classifier outperformed seven classifiers with 97.5 % accuracy, f<sub>1</sub>-score of 97.9 %, precision of 98 %, and recall of 98 %, as shown in Table 4. Using EFSA techniques, all eight classifiers performed more accurately. Fig. 5(a) and (b) illustrate the performance of classifiers on features chosen using the EFSA approach and without feature selection, respectively. As shown in

Fig. 6(a), each classifier demonstrated strong ROC performance, and Fig. 7 displays the testing times (in seconds) without using feature selection and with ensemble feature selection. Notably, when benefiting from ensemble feature selection, the Proposed Classifier showcased significant efficiency, reducing its testing time by 0.11817 s. These outcomes underscore the impact of ensemble feature selection on computational efficiency, underscoring the importance of such strategies in optimizing overall model performance on dataset I.

#### F. Experiment-second for Dataset-II

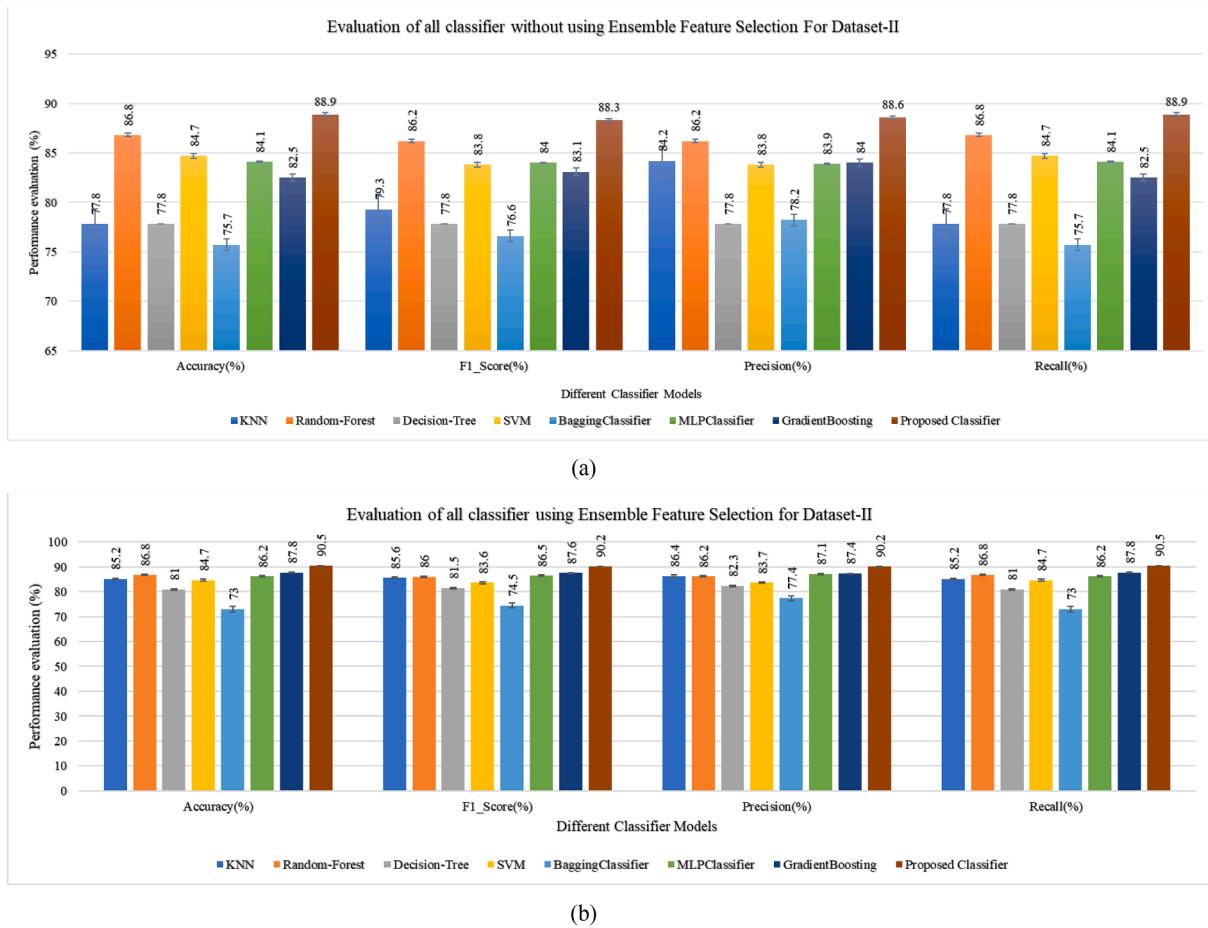


Fig. 8. Comparison of performances (a) without and (b) with Ensemble Feature selection for dataset-II.

In this section, we utilized the D2 dataset, which is the second dataset. Firstly, we applied the Filter-based technique for feature selection, using Pearson Correlation to estimate the correlation for each feature. Relevant features were chosen based on their high correlation value ( $\text{cor}_{\text{target}} > 0.30$ ), resulting in 18 selected features. Secondly, we applied the wrapper feature selection technique using the Backward Elimination technique to remove features that did not significantly affect the dependent variable's strength at a  $p_{\max} > 0.90$ . This approach resulted in the selection of 22 most relevant features. The third feature selection method is embedded-based, using LassoCV to find the optimal

regularisation strength choice and select the top 20 relevant features. We then used the union operation for the three subset features to create an EFSA method. The ensemble feature selection method selected 45 features, the most significant and relevant features, resulting in a more efficient and accurate model. The proposed classification approach showed good accuracy scores with the ensemble-based. The ensemble technique, which selected only 45 features, outperformed the other two feature selection methods. The chosen features using EFSA techniques are displayed in Appendix A, Table 4. After selecting the features using EFSA techniques, we assessed the performance of classifiers. The

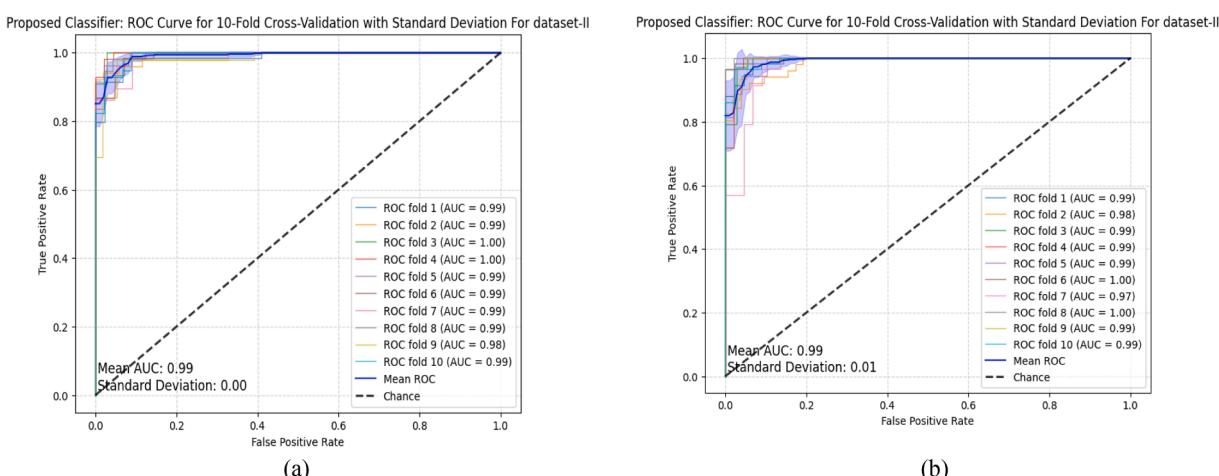


Fig. 9. ROC curve Proposed classifier using 10-fold Cross-Validation in dataset-II (a) without and (b) with ensemble feature selection.

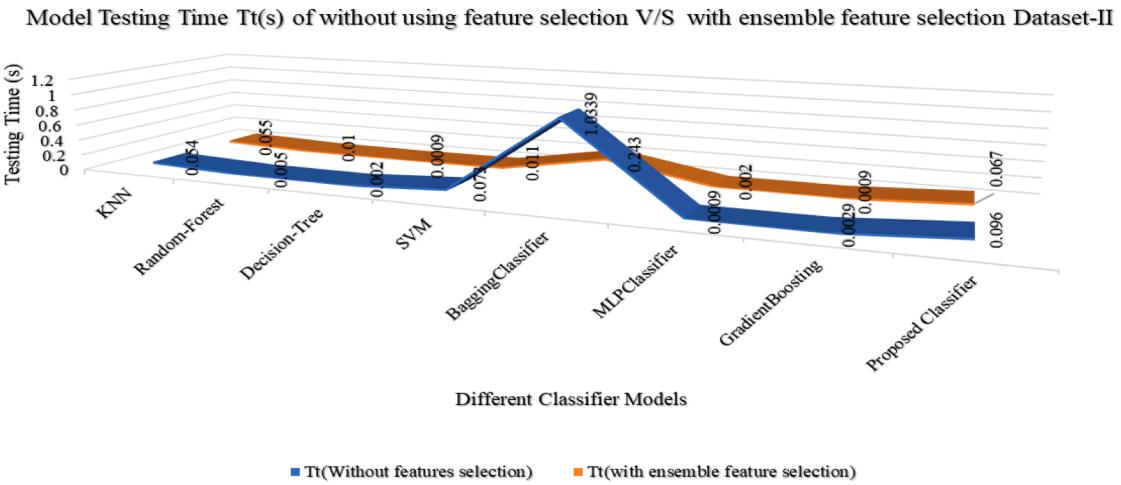


Fig. 10. Testing Time for without and with ensemble features selection for Dataset-II.

Table 6  
Performance evaluation for Dataset-III.

(a) without using feature selection									
Model	Accuracy (%)	F1_Score (%)	Precision (%)	Recall (%)	MCC (%)	CVS (10-fold)	AUC (10-fold)	Training Time (sec.)	Testing Time (sec.)
KNN	76.7	76.7	76.9	76.7	0.5329	82.5 ± 0.108	0.88±0.08	0.00100	0.04712
Random-Forest	78.3	78.4	78.4	78.3	0.5640	80.8 ± 0.077	0.89±0.08	0.28128	0.01700
Decision-Tree	70.0	70.0	70.0	70.0	0.3939	78.8 ± 0.075	0.83±0.11	6.97865	0.00098
SVM	78.3	78.4	78.4	78.3	0.5640	82.1 ± 0.075	0.88±0.10	0.01725	0.00201
BaggingClassifier	70.0	68.4	68.8	68.3	0.3685	76.2 ± 0.095	0.83±0.12	0.19453	0.01406
MLPClassifier	78.3	78.3	78.3	78.3	0.5611	82.2 ± 0.098	0.87±0.11	0.95438	0.00099
GradientBoosting	78.3	78.4	78.4	78.3	0.5640	79.2 ± 0.092	0.87±0.08	0.21783	0.00095
Proposed Classifier	80.0	80.0	80.3	80.0	0.6000	80.8 ± 0.088	0.89±0.09	64.0771	<b>0.06827</b>

(b) with ensemble feature selection									
Model	Accuracy (%)	F1_Score (%)	Precision (%)	Recall (%)	MCC (%)	CVS (10-fold)	AUC (10-fold)	Training Time (sec.)	Testing Time (sec.)
KNN	78.3	78.2	79.1	78.3	0.5747	83.3 ± 0.065	0.90±0.06	0.00295	0.04266
Random-Forest	75.6	73.1	74.3	73.3	0.4762	82.1 ± 0.084	0.90±0.08	85.5759	0.01505
Decision-Tree	65.0	65.0	65.0	65.0	0.3001	76.2 ± 0.071	0.82±0.09	7.52412	0.00099
SVM	81.7	81.5	82.6	81.7	0.6423	85±0.065	0.88±0.09	0.01014	0.00314
BaggingClassifier	78.0	80.0	80.0	80.0	0.6000	81.2 ± 0.095	0.88±0.09	16.8840	1.13042
MLPClassifier	80.0	80.0	80.1	80.0	0.6013	83.8 ± 0.071	0.90±0.06	0.84418	0.00099
GradientBoosting	70.3	69.7	70.8	70.0	0.4082	80.8 ± 0.075	0.89±0.06	0.11934	0.00028
Proposed Classifier	83.3	83.3	83.5	83.3	0.6681	83.08±0.072	0.91±0.07	70.6220	0.06325

classifier employed the GridSearchCV algorithm to tune the hyper-optimized parameter. Table 5 displays the performance of the proposed classifier using all features and reduced feature subsets chosen using EFSA techniques. In the first experiment on dataset D2, without using the feature selection technique, the proposed classifier outperformed seven classifiers with 88.9 % accuracy, an  $f_1$ -score of 88.3 %, a precision of 88.6 %, and a recall of 88.9 %. In the second experiment, using the EFSA algorithm on the same dataset D2 with the proposed classifier, the proposed classifier outperformed seven classifiers with 90.2 % accuracy, an  $f_1$ -score of 90.2 %, precision of 90.2 %, and recall of 90.5 %, as shown in Table 5 and Fig. 8 presents a comparison of the accuracy of classifiers with and without the EFSA approach. The proposed model performance matrix is improved by using EFSA techniques, as shown in Fig. 9, which shows the ROC curve (a) without and (b) with feature selection. Also, calculate performance time in Fig. 10, which displays the testing times (sec) reduced by 0.0292 s.

#### G. Experiment-third for Dataset-III

Similarly, the Third Dataset, or D3, was used in this section. The filter-based strategy was first applied for feature selection using the estimated Pearson Correlation for each feature. The selected relevant

features were those with a high correlation value ( $\text{cor}_{\text{target}} > 0.55$ ), resulting in the selection of 10 features. The next step was to use wrapper feature selection strategies. The Backward Elimination technique was employed to eliminate characteristics that did not significantly affect the dependent variable strength at  $p_{\max} > 0.05$ , resulting in the selection of eight features. The third feature selection method was embedded-based, in which LassoCV found the optimal regularization strength, and the top 10 essential features were selected. The EFSA method was then applied by combining the three subsets of features, resulting in the selection of relevant features that increased the model's effectiveness, speed, and accuracy. The proposed classification approach worked effectively with the ensemble characteristics, producing good accuracy scores. The ensemble technique outperformed the other two feature selection methods by selecting only 17 features. Table 9 in Appendix A lists the features chosen using the EFSA technique. The performance of classifiers was evaluated after selecting features using the EFSA method. The GridSearchCV algorithm was employed to tune the hyper-parameter tuning-based hyper-optimized parameter of the classifier. Table 6 shows the performance of the proposed classifier using EFSA techniques with all features and reduced feature subsets. In the first experiment on dataset D3 without employing the feature selection technique, the proposed classifier outperformed seven classifiers with

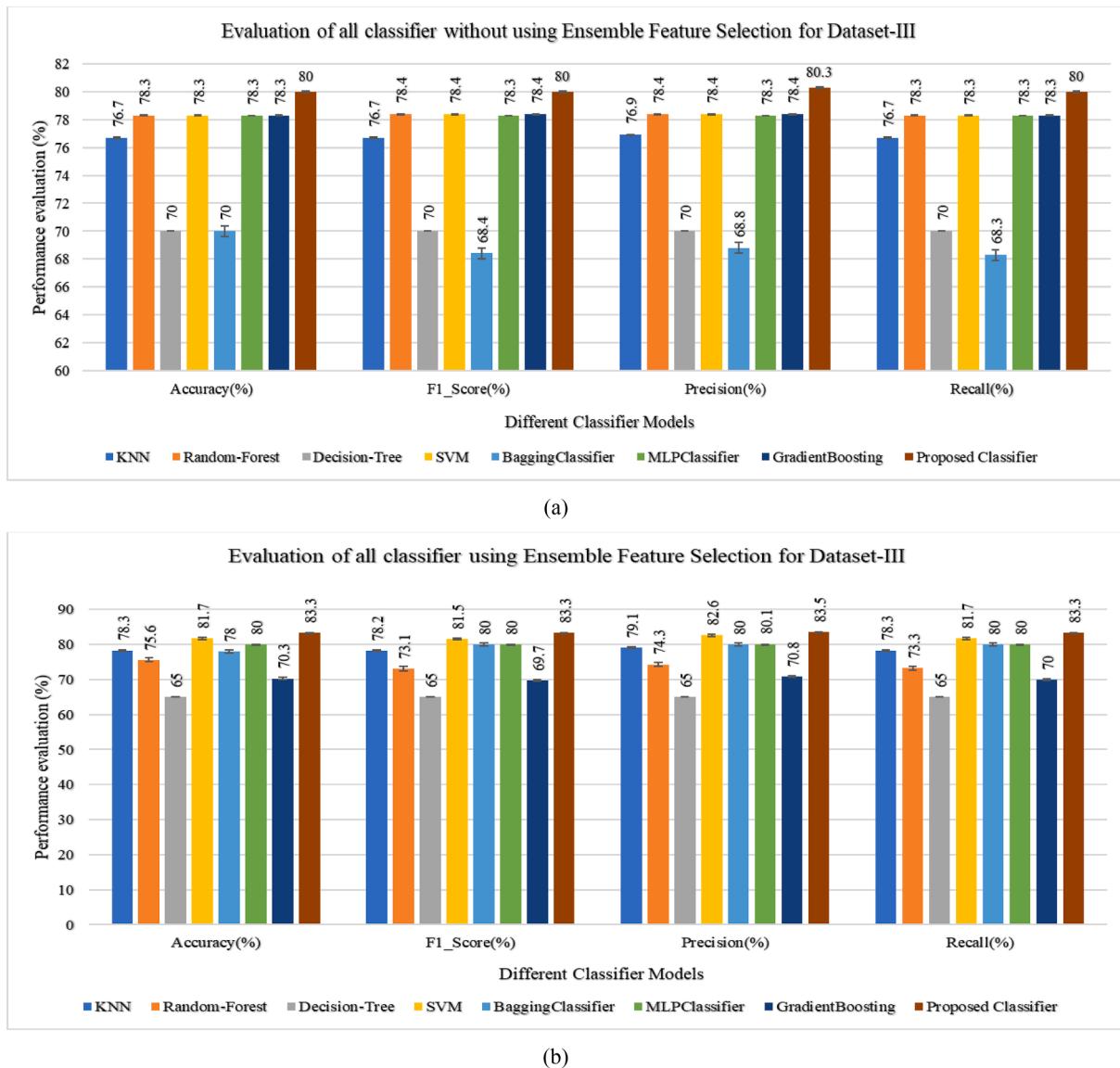


Fig. 11. Comparison of performances (a) without (b) with ensemble features selection for dataset-III.

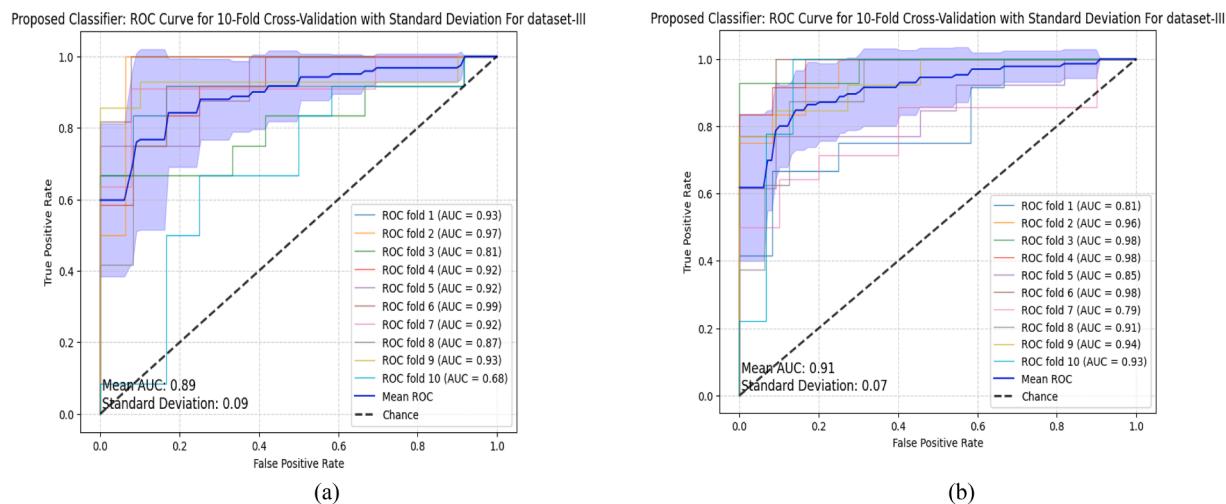


Fig. 12. ROC curve proposed classifier using 10-fold Cross-Validation in dataset-III (a) without and (b) with ensemble feature selection.

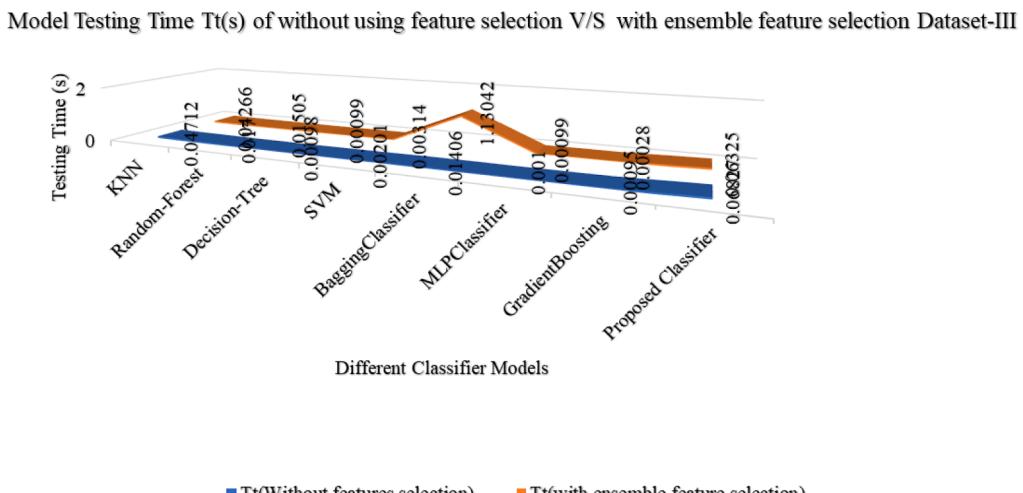


Fig. 13. Testing Time without and with ensemble features selection for Dataset-III.

**Table 7**  
Comparative accuracy (%) analysis of Datasets I, II and III.

Dataset	Model	Without feature selection		With ensemble feature selection	
		Training	Testing	Training	Testing
Dataset - I	KNN	88	87.8	97.2	93.9
	Random-Forest	100	87.8	100	93.9
	Decision-Tree	100	83.7	100	85.7
	SVM	100	91.8	100	89.8
	BaggingClassifier	97.7	81.6	100	85.7
	MLPClassifier	89	79.6	94.5	91.8
	GradientBoosting	97.2	85.7	100	93.9
Dataset - II	Proposed Classifier	99.5	93.9	100	98.0
	KNN	89.1	77.8	100	85.2
	Random-Forest	100	86.8	100	86.8
	Decision-Tree	99.7	77.8	97	81.0
	SVM	100	84.7	100	84.7
	BaggingClassifier	94.1	75.7	87.6	73.0
	MLPClassifier	100	84.1	100	86.2
Dataset - III	GradientBoosting	93.4	82.5	100	87.8
	Proposed Classifier	100	88.9	100	90.5
	KNN	83.8	76.7	86.6	78.3
	Random-Forest	100	78.3	100	73.3
	Decision-Tree	82.7	70.0	87.2	65.0
	SVM	83.8	78.3	95.2	81.7
	BaggingClassifier	85	68.3	82.2	80.0
	MLPClassifier	87.2	78.3	87.2	80.0
	GradientBoosting	100	78.3	100	70.0
	Proposed Classifier	95	80.0	89	83.3

80 % accuracy, 80 % f<sub>1</sub>-score, 80.3 % precision, and 80 % recall. The proposed classifier in the second experiment using the same dataset, D3, outperformed seven classifiers with 83.3 % accuracy, 83.3 % f<sub>1</sub>-score, 83.3 % precision, and 83.3 % recall by employing the EFSA approach. The proposed model performance matrix is improved by using EFSA techniques. Fig. 11(a) and (b) show the performance of the classifiers, and Fig. 12 shows the ROC curve (a) without and (b) with feature selection. Also, calculate performance time in Fig. 13, which displays the testing times (sec) reduced by 0.00502 s.

#### H. Summary of experiment-I, experiment-II and experiment-III

The experiments compare the proposed ensemble classifier with three feature selection techniques (filter, wrapper, and embedded) on three datasets (D1, D2, and D3) to classify patients with Parkinson's disease. The proposed ensemble classifier outperforms seven other

classifiers in accuracy, f<sub>1</sub>-score, precision, recall, and ROC. The ensemble feature selection algorithm selects the most relevant features for the classification task, making the model more efficient and accurate. The experiments show that EFSA (Ensemble Feature Selection Algorithm) techniques improve classifier performance overall more than feature selection alone. In the first experiment on the D1 dataset, the proposed classifier achieved 93.9 % accuracy without using the feature selection technique and 97.6 % accuracy using EFSA. In the second experiment on the D2 dataset, the proposed classifier achieves 88.9 % accuracy without using the feature selection technique and 90.2 % accuracy using EFSA.

Similarly, in the third experiment on the D3 dataset, the proposed classifier achieves 80 % accuracy without using the feature selection technique and 83.3 % accuracy using EFSA. Overall, the experiments demonstrate that the proposed ensemble classifier with the EFSA feature selection technique is effective for Parkinson's Disease classification. Figs. 7, 10, and 13 illustrate testing times for various models with and without ensemble feature selection across different databases. The Proposed Classifier consistently shows efficiency gains with ensemble feature selection, reducing testing times by 0.11817 s (Database-I), 0.0292 s (Database-II), and 0.00502 s (Database-III). These results underscore the positive impact of ensemble feature selection on computational efficiency, highlighting its importance in optimizing model performance across diverse datasets. (Table 7)

#### 6. Comparison of the performance of the proposed classifier

The experimental outcomes of Parkinson's disease diagnosis utilizing the three datasets have been obtained from the literature and contrasted with the proposed methodology in this study. The comparison results are shown in Table 8.

In the above literature (Liu et al., 2022; Devarajan and Ravi, 2019; Mamun et al., 2022; Hawi et al., 2022; Lamba et al., 2022; Das, 2010; Sakar and Kursun, 2010; Sakar et al., 2019; Fang, 2022; El-Hasnony et al., 2020; Yücelbaş, 2020; Sabreena et al., 2022; Naranjo et al., 2016; Nahar et al., 2021; Avuçlu and Elen, 2020; Senturk, 2020; Al-Husban et al., 2022; Tallapareddy and Radha, 2022), To compare the outcomes of the experiments, the authors of the paper also chose a variety of models. This study compares the method in the literature to the examples with the best results and lists them. The Accuracy and F<sub>1</sub>-score are also noticeably better than the outcomes in the comparative literature, demonstrating the efficacy of the strategy used in this work.

**Table 8**

Comparison of results on proposed model accuracy vs literature.

Dataset	References	FS Method	Technology	Accuracy
Dataset-I (Little et al., 2007) ( <a href="#">UCI Machine Learning Repository: Parkinsons Data Set 2022</a> )	Sharma et al. (2019)	Modified Grey Wolf Optimization	K-Nearest Neighbor, RF, DT	93.87 %
	Devarajan and Ravi (2019)	All features	FKNN—CBR	94.87 %
	Das (2010)	All features	Neural Networks, DMneural, Regression and Decision Tree	neural network classifier 92.9 %
	Avuçlu and Elen (2020)	All features	k-NN, RF, Naïve Bayes, SVM	Accuracy from Random Forest 85.81 %
	Senturk (2020)	RFE(selected13 features)	SVM	93.84 %
	Mamun et al. (2022)	All features	LightGBM	95 %,
	Fang (2022)	All features	k-NN	93.88 %
	Lamba et al. (2022)	Extra Tree, Mutual Information Gain, Genetic Algorithm	Naïve Bayes, k-Nearest Neighbor, Random Forest	95.58 %
	<b>Proposed method</b>	<b>EFSA Algorithm (selected 13 features)</b>	<b>Proposed classifier</b>	<b>97.6 %</b>
		All features	Ensemble with voting	81 %
Dataset-II (Sakar et al., 2019) ( <a href="#">UCI Machine Learning Repository: Parkinson's Disease Classification Data Set 2022</a> )			Ensemble with stacking	80 %
	Sakar et al. (2019)	mRMR-50 and the TQWT	SVM (RBF kernels)	86 % (with 50 features)
	El-Hasnony I et al. (2020)	119 features	ANFIS+PSOGWO	87.5 %
	Liu et al. (2022)	50 features	SHAP-LightGBM	91.62 %
	Hawi et al. (2022)		Majority of voting	84.88 %
	Sabeena et al. (2022)		(MFCC + Wavelet + Concat), FCBi-LSTM classifier (%)	98.77 %
	Al-Husban et al. (2022)	All features	DT, SVM, and kNN	SVM
	Tallapureddy and Radha (2022)	All features	Feature Reduction Principal Component Analysis (PCA)	91 %
	<b>Proposed method</b>	<b>EFSA Algorithm (selected 45 features)</b>	<b>Proposed classifier</b>	<b>90.2 %</b>
	Naranjo et al. (2016)	40PD 40HC	Bayesian classification	75 %(mean)
Dataset-III (Naranjo et al., 2016) ( <a href="#">UCI Machine Learning Repository: Parkinson Dataset with replicated acoustic features Data Set 2022</a> )	Nahar et al. (2021)	21, RFE	Bagging with recursive feature elimination	82.35 %
	Dhar (2022)	40PD 40HC	Author Proposed MI-AE-GOLGBM	84.17 %
	<b>Proposed Classifier</b>	<b>EFSA Algorithm (selected 17 features)</b>	<b>Proposed Classifier</b>	<b>83.3 %</b>

## 7. Conclusion and future work

This paper proposed an ensemble technique for predicting Parkinson's Disease (PD) using voice signals. This model's key innovation is using an EFSA and voting classifier to perform PD diagnosis tasks. A comparison is made between the proposed with and without ensemble features selection algorithm with other classifiers and the proposed voting classifier. This work aims to build an efficient model to predict PD patients using voice signals, as there is no standard feature selection technique or classifier for medical datasets. Multiple approaches were tested; the most effective approach was a combination of several. The results show that the feature selection method reduces complexity and enhances accuracy. Three different voice datasets were used in this paper to apply various classifiers, and the results were compared using visualization and statistical analysis. Dataset-I achieved a classification accuracy of 97.6 %, an F<sub>1</sub>-score of 97.9 %, a precision of 98 %, and a recall of 98 %. Dataset-II achieved a classification accuracy of 90.2 %, F<sub>1</sub>-score of 90.2 %, precision of 90.2 %, and recall of 90.5 %. Dataset-III achieved a classification accuracy of 83.3 %, an F<sub>1</sub>-score of 83.3 %, a precision of 83.5 %, and a recall of 83.3 %. These results were obtained using 13 out of 23, 45 out of 754, and 17 out of 46 features from respective datasets. Among all the classifiers, it was found that the proposed classifier with EFSA outperformed other classifiers and some recent existing literature in machine learning algorithms.

We plan to include cutting-edge approaches in our future work to improve our Parkinson's Disease prediction model. Feature extraction

algorithms and enhanced feature selection strategies are used to improve the model's performance. To further help patients and medical professionals, we also intend to implement the model on Android smartphones, allowing for real-time symptom assessment. This method of managing and diagnosing Parkinson's disease improves accessibility and accuracy by combining state-of-the-art machine learning algorithms with mobile technologies.

## Limitation

- The suggested approach is unable to determine the Parkinson's disease severity stage since the illness advances slowly.
- Diagnosing Parkinson's disease at an early stage is challenging due to its initial symptoms, often beginning with tremors in patients.

## Compliance with ethical standards

This study uses three publicly available datasets in the UCI repository related to Parkinson's Disease voice. References are available in the below links ([UCI Machine Learning Repository: Parkinsons Data Set 2022](#); Fang, 2022; Bao et al., 2022). This article does not contain any study performed on animals.

## CRediT authorship contribution statement

**Nutan Singh:** Writing – original draft, Software, Methodology,

**Conceptualization.** Priyanka Tripathi: Writing – review & editing, Supervision.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.spcom.2024.103067](https://doi.org/10.1016/j.spcom.2024.103067).

## References

- Al-Husban, A., Abdulridha, M.M., Mohamad, A.A.H., Ibrahim, A.M., 2022. Biocomposite's multiple uses for a new approach in the diagnosis of Parkinson's disease using a machine learning algorithm. *Adsorpt. Sci. Technol.* 2022, 1–7. <https://doi.org/10.1155/2022/6159392>.
- Ali, A.M., Salim, F., Saeed, F., 2023. Parkinson's disease detection using filter feature selection and a genetic algorithm with ensemble learning. *Diagnostics* 13 (17). <https://doi.org/10.3390/diagnostics13172816>.
- Al-Sarem, M., Saeed, F., Bouila, W., Emara, A.H., Al-Mohaiemeed, M., Errais, M., 2021. Feature selection and classification using CatBoost method for improving the performance of predicting Parkinson's disease. *Advances in Intelligent Systems and Computing*. Springer Science and Business Media Deutschland GmbH, pp. 189–199. [https://doi.org/10.1007/978-981-15-6048-4\\_17](https://doi.org/10.1007/978-981-15-6048-4_17).
- Avuolu, E., Elen, A., 2020. Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements. *Med. Biol. Eng. Comput.* 58 (11), 2775–2788. <https://doi.org/10.1007/s11517-020-02260-3>.
- Bao, G., Lin, M., Sang, X., Hou, Y., Liu, Y., Wu, Y., 2022. Classification of dysphonic voices in Parkinson's disease with semi-supervised competitive learning algorithm. *Biosensors* 12 (7). <https://doi.org/10.3390/bios12070502>.
- Bradshaw, T.J., Huemann, Z., Hu, J., Rahim, A., 2023. A guide to cross-validation for artificial intelligence in medical imaging. In: *Radiology: Artificial Intelligence*, 5. Radiological Society of North America Inc. <https://doi.org/10.1148/ryai.220232>
- Chandra, T.B., Verma, K., Singh, B.K., Jain, D., Netam, S.S., 2021. Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble. *Expert. Syst. Appl.* 165 <https://doi.org/10.1016/j.eswa.2020.113909>.
- L.-C. Chang et al., "Machine learning approaches to identify Parkinson's disease using voice signal features."
- Das, R., 2010. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert. Syst. Appl.* 37 (2), 1568–1572. <https://doi.org/10.1016/j.eswa.2009.06.040>.
- Devarajan, M., Ravi, L., 2019. Intelligent cyber-physical system for an efficient detection of Parkinson disease using fog computing. *Multimed. Tools. Appl.* 78 (23), 32695–32719. <https://doi.org/10.1007/s11042-018-6898-0>.
- Dhar, J., 2022. An adaptive intelligent diagnostic system to predict early stage of Parkinson's disease using two-stage dimension reduction with genetically optimized lightgbm algorithm. *Neural Comput. Appl.* 34 (6), 4567–4593. <https://doi.org/10.1007/s00521-021-06612-4>.
- El-Hasnony, I.M., Barakat, S.I., Mostafa, R.R., 2020. Optimized ANFIS model using hybrid metaheuristic algorithms for Parkinson's disease prediction in iot environment. *IEEE Access*, 8, 119252–119270. <https://doi.org/10.1109/ACCESS.2020.3005614>.
- Fang, Z., 2022. Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease. In: *Proceedings - 2022 International Conference on Machine Learning and Knowledge Engineering, MLKE 2022*, pp. 98–101. <https://doi.org/10.1109/MLKE55170.2022.00024>.
- Fayed, H.A., Atiya, A.F., 2019. Speed up grid-search for parameter selection of support vector machines. *Appl. Soft Comput.* J. 80, 202–210. <https://doi.org/10.1016/j.asoc.2019.03.037>.
- Hawi, S., Alhozami, J., AlQahtani, R., AlSafran, D., Alqarni, M., El Sahmarany, L., 2022. Automatic Parkinson's disease detection based on the combination of long-term acoustic features and Mel frequency cepstral coefficients (MFCC). *Biomed. Signal. Process. Control* 78. <https://doi.org/10.1016/j.bspc.2022.104013>.
- Hussain, M.M., Weslin, D., Kumar, S., Umamaheswari, S., Kamalakkannan, K., 2023. Enhancing Parkinson's disease identification using ensemble classifier and data augmentation techniques in machine learning. *Clin. eHealth* 6, 150–158. <https://doi.org/10.1016/j.ceh.2023.11.002>.
- Ibarra-Gutiérrez, M.T., Serrano-García, N., Orozco-Ibarra, M., 2023. Rotenone-induced model of Parkinson's disease: beyond mitochondrial complex I inhibition. *Molecular Neurobiology*. Springer. <https://doi.org/10.1007/s12035-022-03193-8>.
- Jain, D., Mishra, A.K., Das, S.K., 2021. Machine learning based automatic prediction of Parkinson's disease using speech features. *Advances in Intelligent Systems and Computing*. Springer, pp. 351–362. [https://doi.org/10.1007/978-981-15-4992-2\\_33](https://doi.org/10.1007/978-981-15-4992-2_33).
- Kaur, S., Aggarwal, H., Rani, R., 2020. Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease. *Machine Vision and Applications*. Springer. <https://doi.org/10.1007/s00138-020-01078-1>.
- Kumari, S., Kumar, D., Mittal, M., 2021. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *Int. J. Cognit. Comput. Eng.* 2, 40–46. <https://doi.org/10.1016/j.ijcce.2021.01.001>.
- Lamba, R., Gulati, T., Jain, A., 2022. A hybrid feature selection approach for Parkinson's detection based on mutual information gain and recursive feature elimination. *Arab. J. Sci. Eng.* 47 (8), 10263–10276. <https://doi.org/10.1007/s13369-021-06544-0>.
- Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O., 2009. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 56 (4), 1015–1022. <https://doi.org/10.1109/TBME.2008.2005954>.
- Liu, Y., Liu, Z., Luo, X., Zhao, H., 2022. Diagnosis of Parkinson's disease based on SHAP value feature selection. *Biocybern. Biomed. Eng.* 42 (3), 856–869. <https://doi.org/10.1016/j.bbce.2022.06.007>.
- Mamun, M., Mahmud, M.J., Hossain, M.I., Islam, A.M., Ahammed, M.S., Uddin, M.M., 2022. Vocal feature guided detection of Parkinson's disease using machine learning algorithms. In: *2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022*, pp. 566–572. <https://doi.org/10.1109/UEMCON54665.2022.9965732>.
- Y. Mittra and V. Rustagi, "Classification of subjects with Parkinson's Disease using Gait Data Analysis".
- Nahar, N., Ara, F., Neloy, M.A.I., Biswas, A., Hossain, M.S., Andersson, K., 2021. Feature Selection Based Machine Learning to Improve Prediction of Parkinson Disease. Springer International Publishing. [https://doi.org/10.1007/978-3-030-86993-9\\_44](https://doi.org/10.1007/978-3-030-86993-9_44) vol. 12960 LNAI.
- Naranjo, L., Pérez, C.J., Campos-Roca, Y., Martín, J., 2016. Addressing voice recording replications for Parkinson's disease detection. *Expert. Syst. Appl.* 46, 286–292. <https://doi.org/10.1016/j.eswa.2015.10.034>.
- "Parkinson disease." Accessed 17 January 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>.
- Poewe, W., et al., 2017. Parkinson disease. *Nat. Rev. Dis. Primers.* 3, 1–21. <https://doi.org/10.1038/nrdp.2017.13>.
- Polat, K., 2019. A hybrid approach to Parkinson disease classification using speech signal: the combination of SMOTE and random forests. In: *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/EBBT.2019.8741725>
- M. Raheem, "Machine Learning Based Idiopathic Parkinson's Disease Detection Using Speech Machine Learning Based Idiopathic Parkinson's Disease Detection Using Speech Data," 2022.
- Sabeena, B., Sivakumari, S., Teressa, D.M., 2022. Optimization-based ensemble feature selection algorithm and deep learning classifier for Parkinson's disease. *J. Healthc. Eng.* 2022 <https://doi.org/10.1155/2022/1487212>.
- Sakar, C.O., et al., 2019. 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.* J. 74, 255–263. <https://doi.org/10.1016/j.asoc.2018.10.022>.
- Sakar, C.O., Kursun, O., 2010. Telediagnosis of parkinson's disease using measurements of dysphonia. *J. Med. Syst.* 34 (4), 591–599. <https://doi.org/10.1007/s10916-009-9272-y>.
- Saravanan, S., et al., 2022. A systematic review of Artificial Intelligence (AI) based approaches for the diagnosis of Parkinson's disease. *Archives of Computational Methods in Engineering* 29 (6), 3639–3653. <https://doi.org/10.1007/s11831-022-09710-1>.
- Senturk, Z.K., 2020. Early diagnosis of Parkinson's disease using machine learning algorithms. *Med. Hypotheses*. 138 <https://doi.org/10.1016/j.mehy.2020.109603>.
- Shahbaba, B., Neal, R., 2009. Nonlinear models using Dirichlet process mixtures. *J. Mach. Learn. Res.* 10, 1829–1850.
- Sharma, P., Sundaram, S., Sharma, M., Sharma, A., Gupta, D., 2019. Diagnosis of Parkinson's disease using modified grey wolf optimization. *Cogn. Syst. Res.* 54, 100–115. <https://doi.org/10.1016/j.cogsys.2018.12.002>.
- Shastri, K.A., 2023. An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease. *Healthc. Anal.* 3 <https://doi.org/10.1016/j.healthc.2023.100181>.
- Tallapareddy, G., Radha, D., 2022. Analysis of ensemble of machine learning algorithms for detection of Parkinson's disease. In: *Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAC 2022*. Institute of Electrical and Electronics Engineers Inc., pp. 354–361. <https://doi.org/10.1109/ICAAIC53929.2022.9793048>
- "UCI Machine Learning Repository: Parkinson Dataset with replicated acoustic features Data Set." 6 Accessed 2022. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Parkinson%2BDataset%2Bwith%2Breplicated%2Bacoustic%2Bfeatures%2B>.
- "UCI Machine Learning Repository: Parkinsons Data Set." 6 Accessed 2022. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/parkinsons>.

- "UCI Machine Learning Repository: Parkinson's Disease Classification Data Set." Accessed 6, 2022. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification>.
- Yuan, L., Liu, Y., Feng, H.-M., 2023. Parkinson disease prediction using machine learning-based features from speech signal. *Serv. Orient. Comput. Appl.* <https://doi.org/10.1007/s11761-023-00372-w>.
- Yücelbaş, Ş., 2020. Simple logistic hybrid system based on greedy stepwise algorithm for feature analysis to diagnose Parkinson's disease according to gender. *Arab. J. Sci. Eng.* 45 (3), 2001–2016. <https://doi.org/10.1007/s13369-020-04357-1>.



**Nutan Singh** received her Master Degree in Computer Application and currently pursuing Ph.D. from National Institute of Technology, Raipur, Chhattisgarh, India. Her research interest includes Signal processing, feature selection, image processing and machine learning.



**Dr. Priyanka Tripathi** received her Ph.D. from Maulana Azad National Institute of Technology Bhopal, India. She is Head and Professor in the Department of Computer Applications, NIT Raipur, Chhattisgarh, India. Her research interests include Cloud Computing, Internet of Things, Web Engineering.