

Enhancing Parkinson's Disease Prediction Using Machine Learning and Feature Selection Methods

Faisal Saeed^{1,2,*}, Mohammad Al-Sarem^{1,3}, Muhannad Al-Mohaimeed¹, Abdelhamid Emara^{1,4},
Wadii Boulila^{1,5}, Mohammed Alasli¹ and Fahad Ghabban¹

¹College of Computer Science and Engineering, Taibah University, Medina 41477, Saudi Arabia

²School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, United Kingdom

³Information System Department, Saba'a Region University, Mareeb, Yemen

⁴Computers and Systems Engineering Department, Al-Azhar University, Cairo 11884, Egypt

⁵RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba 2010, Tunisia

Corresponding Author: Faisal Saeed. Email: fsaeed@taibahu.edu.sa

Received: XX Month 202X; Accepted: XX Month 202X

Abstract: Several millions of people suffer from Parkinson's disease globally. Parkinson's affects about 1% of people over 60 and its symptoms increase with age. The voice may be affected and patients experience abnormalities in speech that might not be noticed by listeners, but which could be analyzed using recorded speech signals. With the huge advancements of technology, the medical data has increased dramatically, and therefore, there is a need to apply data mining and machine learning methods to extract new knowledge from this data. Several classification methods were used to analyze medical data sets and diagnostic problems, such as Parkinson's Disease (PD). In addition, to improve the performance of classification, feature selection methods have been extensively used in many fields. This paper aims to propose a comprehensive approach to enhance the prediction of PD using several machine learning methods with different feature selection methods such as filter-based and wrapper-based. The dataset includes 240 recodes with 46 acoustic features extracted from 3 voice recording replications for 80 patients. The experimental results showed improvements when wrapper-based features selection method was used with K-NN classifier with accuracy of 88.33%. The best obtained results were compared with other studies and it was found that this study provides comparable and superior results.

Keywords: Filter-based Feature Selection Methods; Machine Learning; Parkinson's disease; Wrapper-based Feature Selection Methods.

1 Introduction

Parkinson's disease (PD) is a long term degenerative disorder of the central nervous system which causes both motor and non-motor symptoms [1]. The exact causes of PD are unknown and unclear, but it is supposed to include risk factors which are both genetic and environmental. More than 10% of patients with PD have a first-degree relative with PD disease. In addition, PD is more prevalent between people who are disclosed to some pesticides and the people with past history of head injury, while PD risk is lower for patients who smoke [2]. PD mainly affects neurons in a certain region of the mid brain that is known as substantia nigra, dopamine-producing brain cells, which leads to inadequate dopamine secretion in this region [3].

In the early stage of the PD, the main symptoms are shaking, difficulty with walking and slowness of movement. The common symptoms with late phase of PD are anxiety, dementia and depression. Moreover, emotional problems, sleep and sensory symptoms may also occur [4, 5], in addition to Parkinsonian syndrome [6]. These symptoms are mainly used to diagnose typical PD, in addition to examinations such



as neuroimaging. There is no total recovery for PD, however treatment aims to improve the symptoms [7, 8]. The medical decision support systems (MDSS) have increasingly used a significant diagnosis and treatment method that uses artificial intelligence (AI) methods on a clinical dataset to assist clinicians to make better decisions [9,10]. Recent improvements in machine learning, AI and statistical learning have improved decision support system (DSS), which has helped to introduce intelligent decision systems [10,11]. Some studies reported that the artificial intelligence cannot be effective without learning [12]. There are many types of machine learning methods such as Support Vector Machine (SVM), Naïve Bayes (NB), K-nearest Neighbor (KNN), Multilayer Perceptron, Decision Tree (DT) and Random Forests (RF) that have been used to solve medical decision problems.

There is a significant overlapping between ML and data mining which often use the same procedures, but whereas ML concentrates on prediction, based on previously definite properties learned from the training data, data mining concentrates on the detection of unknown properties in the clinical data. The machine learning (ML) techniques have a significant role to play in the medical disease diagnosis field and are widely used in bioinformatics [13, 14].

Recently, the variety of medical data is continuously increasing, therefore, effective classification and prediction algorithms are required. The previous studies on machine learning research reported that the accuracy of a classification algorithm can be influenced by many agents [15]. ML algorithms are used to analyze medical data sets and diagnostic problems [12]. Subsequently, improvement of medical decisions, treatments, and decrease financial costs will occur [16,14].

In addition, feature selection plays an important role in the explanation of medical data. Feature selection technique constitutes a significant issue of global combinatorial optimization in machine learning, which is used to decrease the number of features from the original features, removes irrelevant or redundant features without incurring much loss of information, as well as simplification of models to make them easier to interpret and shortening training times [17]. Therefore, a good feature selection method is required to accelerate processing time and predictive accuracy. There are three types of feature selection algorithms, which are: filter (extract features from the data set without any learning), wrapper (use learning techniques to estimate useful features) and hybrid (gather the feature selection step and the classifier construction) [18, 19].

Recently, the medical field is the most favorable field to use machine learning methods. Therefore, Naïve Bayes (NB), Support Vector Machine (SVM), K-nearest Neighbor, Multilayer Perceptron and Random Forests as well as feature selection methods have been suggested to solve medical decision problems, such as the prediction of Parkinson's disease. In this paper, the main contributions in the domain of prediction of Parkinson's Disease can be summarized as follows.

1. A comprehensive approach was used to investigate the performance of several feature selection methods and machine learning methods in order to enhance the prediction of PD.
2. These feature selection methods include both filter-based methods such as (Information gain IG, Principle Component Analysis PCA) and wrapper methods that include different search methods such as First Best Greedy Stepwise PSO Method.
3. A comparative analysis was conducted to examine the performances of all methods/combinations used and the best prediction results were reported.

This paper is organized as follows: Section 2: the related works. Section 3: discussion of the methods. Section 4: experimental results and discussion. Section 5: conclusions and future works.

2 Related Studies

Several works have investigated the diagnosis of PD, in which many machine learning methods were applied such as Support Vector Machine, neural network, Naïve Bayes, K-nearest neighbor and Random Forests. In this paper, several datasets were used to search for related studies on Parkinson's disease, including Scopus, IEEE Xplore, Science Direct and Google Scholar.

In [20] a supervised ML method was proposed that combined the Principal Components Analysis (PCA) to extract features and SVM as classification method to identify PD patients. The main goal of this method was to determine patients that will be diagnosed with PD or with Progressive Supranuclear Palsy (PSP). The experiments were conducted on data of several patients with clinical and demographic features. The results depicted good accuracy of the proposed method in identifying the PD patients compared to existing related works.

In addition, the authors in [21] proposed an expert system of PD using features extracted from recordings of patients' voice. They developed a Bayesian classification approach to deal with the dependence to match the replication-based experimental design. The experiments were performed on voice recordings involving 80 subjects, 50% of them had PD. The aim was to identify which subjects had no PD and which did have the disease. Naranjo et al. addressed the problem of identifying PD patients using the extracted acoustic features from repeated voice recordings. The proposed method was based on two steps, namely variable selection, and classification. The first step aims to reduce the number of features, while the next step uses a regularization method named LASSO (Least Absolute Shrinkage and Selection Operator) as a classifier. The proposed method was tested on the previously described database and showed a good capacity for PD discrimination.

In addition, the authors in [22] addressed the problem of PD diagnosis by developing an approach that investigated gait and tremor features that were extracted from the voice reordering data. They started by filtering data to remove noises, then, using this data to extract gait features they detected the peak and measured the pulse duration. The average accuracy obtained for the identifying PD patients by the proposed approach was satisfactory.

The authors in [23] proposed a method to automatically detect PD by using the convolutional neural network (CNN). The authors suggested considering electroencephalogram (EEG) signals to build a thirteen-layer CNN model. The proposed approach experimented with EEG signals of 20 Parkinson's disease patients (50% men and 50% women). The CNN method obtained interesting results to identify PD patients; however, its performance should be evaluated using a large population.

Recently, Mostafa et al. [24] tried to enhance the diagnoses of PD by using several methods of feature evaluation and classification. They used a multi-agent system to evaluate multiple features by using five classification methods, namely DT, NB, NN, RF, and SVM. To evaluate the proposed method, they conducted several experiments using original and filtered datasets. The results depicted that this method enhanced the performance of ML methods used by finding the best set of features.

In addition, several methods were applied by [25-27] in order to predict Parkinson's disease. These methods applied several machine learning and feature selection methods to enhance the prediction of Parkinson's disease and other studies utilized machine learning and deep learning to improving prediction of diseases [28-38]. This paper extends these efforts by applying a comprehensive approach to investigate the performance of several machine learning with feature selection methods.

3 Methods

There are many feature selection techniques available, and we have considered the utilization of the following feature selection techniques: Filter-based technique, Correlation-based Feature Subset Selection (CfsSubsetEval), Principle Component Analysis (PCA), and Wrapper technique. The aforementioned techniques use different strategies or search algorithms to generate subsets and progress the search processes including (i) Best First (ii) Greedy Stepwise, (iii) Particle Swarm Optimization (PSO), and (vi) Ranker (see Fig. 1).

The dataset used in this paper is available online at UCI Machine Learning Repository [14]. The dataset contains acoustic features of 80 patients, 50% of them suffering from Parkinson's disease. The data set has 240 recordings with 46 acoustic features extracted from 3 voice recording replications per patient. The data set is well-balanced by gender and class label (whether the patients have Parkinson's disease or not).

The experimental protocol was designed for evaluating the combination of the above techniques and search algorithms when they were used with the following classification models: (i) Naïve Bayes, (ii) Support Vector Machine (SVM)¹, (iii) K-Nearest Neighbor (K-NN), (vi) Multi-Layer Perceptron (MLP) and (v) Random Forest (RF). The experiments were carried out on WEKA tool version 3.8 and MacBook Pro with OS X Yosemite version 10.10.5 as an operating system. To evaluate the performance of each classifier, we first ran feature selection in order to find the representative features and then we applied the classification models. Additionally, 10-fold cross validation was applied and the results have been reported in terms of Accuracy, Recall, Precision and F-score. Finally, we analyzed the results achieved from the experimentations. As stated earlier, the main goal of the research is to enhance the prediction of Parkinson's disease. However, this work also provides a useful guide to selecting the best feature selection technique for different classification models.

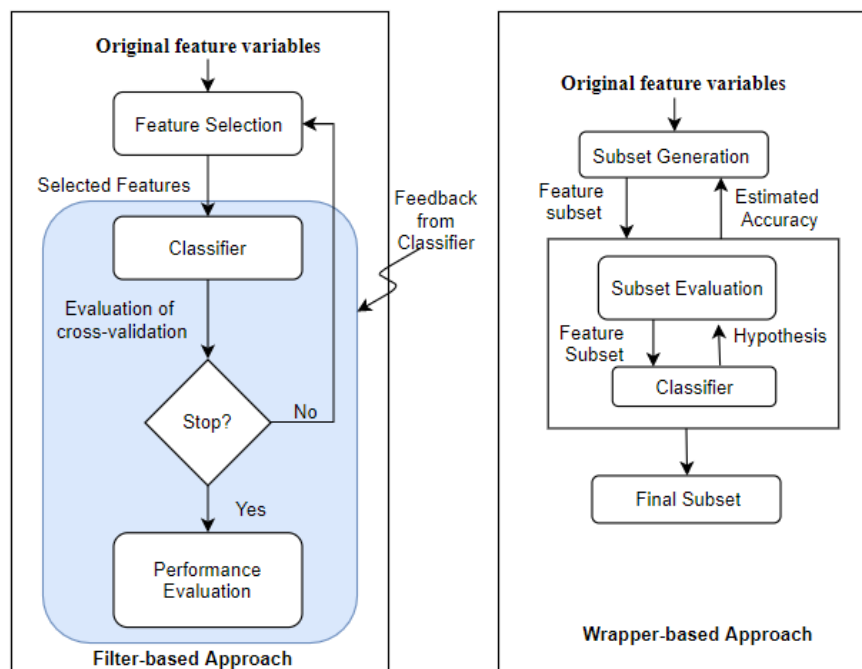


Figure 1: Filter-based Approach vs. Wrapper-based Approach

3.1 Feature Selection Techniques

Several feature selection techniques were applied before feeding the data into the classifier. The filter-based techniques consider the relevance between the features. Thus, they have low complexity, acceptable stability and scalability [39]. A disadvantage of this type of technique is that it might ignore some informative features, especially when the data is coming in stream [40]. The filter-based approaches can be either univariate or multivariate [41]. The univariate methods examine features according to the statistically-based criterion such as Information Gain (IG) [42-44]. Multivariate methods compute feature dependency before ranking the feature. In addition, Principle Component Analysis (PCA) is a common statistical method that is used for data analysis. PCA reduces the size of the data sets by selecting a set of features that represents the whole data set. Since PCA is a conversion technique, the principal components of the first variables is the component with the highest variance value. Then, other principal components are ordered with descending variance values [45]. In addition, the wrapper-based techniques evaluate the quality of the selected features using the performance of the learning classifier.

¹ Both, the c-SVM and nu-SVM are examined.

Regarding the search strategies, the search algorithms follows either sequential forward search (SFS), or sequential backward search (SBS). The SFS starts with a single feature and then iteratively adds or removes features until some terminating criterion is met whereas SBS starts with the whole feature set and then continues with adding and deleting operations. Since the SBS method attempts to find solutions ranged between suboptimal and near optimal regions [41], it is worth fully employing optimization techniques to figure out the subset that leads to maximizing the learner's performance, in particular, with the wrapper approach. At this end, the wrapper-based method can take advantage of various optimization methods such genetic algorithm [46,47] and ant colony optimization algorithm (ACO) [48].

3.2 Machine Learning Classifiers

In machine learning, the data classification is still an attractive domain. Lately, there are many proposed algorithms that have been examined in several domains such as NB, SVM, K-NN, MLP and RF, which are presented briefly in the next subsections.

3.2.1 Support Vector Machine

The basic idea behind SVM algorithm is to construct a hyperplane between groups of data. The quality of the hyperplane is evaluated by measuring to which degree it can maintain the largest distance from the points in either class [39]. Therefore, as it is presented in Fig.2, the higher the separation ability of the hyperplane, the lower the error in the value [49]. The computational complexity of SVM is $O(n^2)$ [50,51].

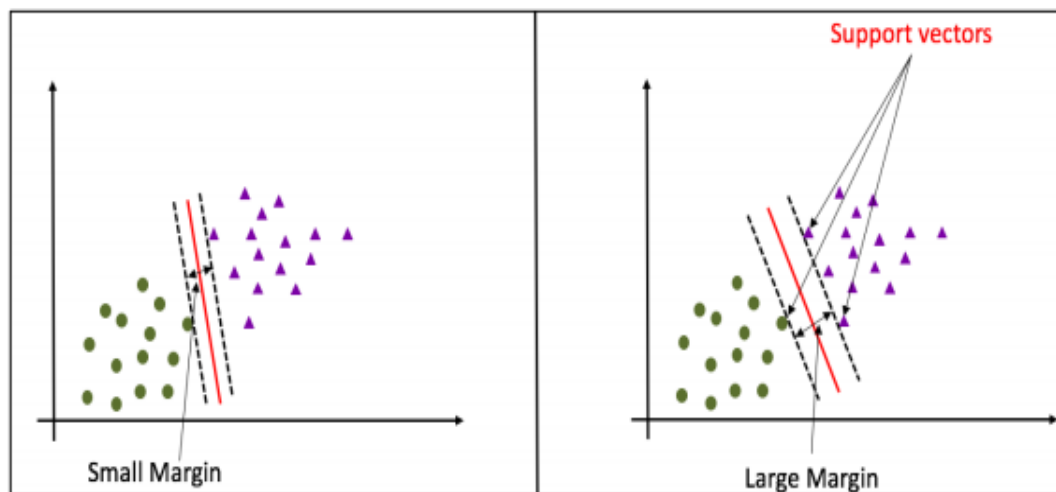


Figure 2: SVM illustration. The larger margin separating the data points, the higher accuracy we obtained.

3.2.2 Naïve Bayes

Naïve Bayes (NB) is a probabilistic classifier that is based on Bayesian theorem. It is called Naïve because the classifier works on a strong features independence assumption. In literature, there are several variants of NB: simple Naïve Bayes, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes and Multi-variant Poisson Naïve Bayes in which the main different among them is the way the probability of the target class is computed. The time complexity of Naïve Bayes is $O(d \times c)$ where d is the query vector's dimension, and c is the total classes.

3.2.3 K-Nearest Neighbor

K-NN is a type of lazy learning, in which there is no explicit training phase and all computations are deferred until classification. It is a method of classifying data based on the nearest training data points in the feature space. The K-NN classifier uses the Euclidean distance measure, or another measure such as Euclidean squared, Manhattan, and Chebyshev, to estimate the target class. The performance of the classifier depends upon the parameter k , while the best value of k depends upon the dataset. In general, the greater the value of k , the lower the noises in the classification, but the boundaries between the classes become less distinct as shown in Fig. 3. The time complexity of K-NN is $O(n \times m)$, where n is the number of training examples and m is the number of dimensions in the training set [52].

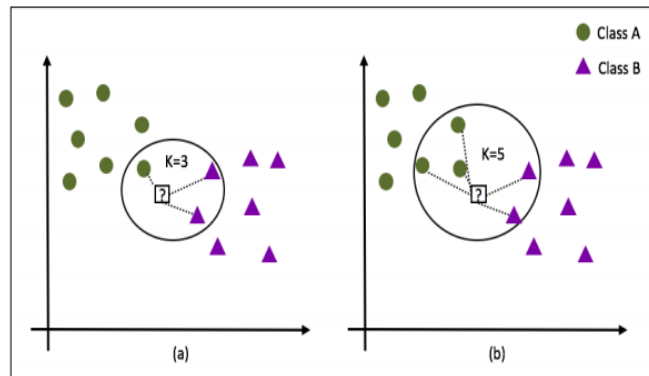


Figure 3: K-NN Model. When $k=3$, the classifier predicts a new point as B class (Fig. a), whilst, when $k=5$, the point is determined as a class A.

3.2.4 Multilayer Perceptron Model

The MLP is a classical feedforward neural network classifier in which the errors of the output are used to train the network [53]. MLP consists of three layers of nodes: (i) input layer, (ii) at least one or more hidden layer(s), and (iii) output layer. The input layer is connected to the hidden layers which are connected to the output layer. All the layers are processed by weighted values. Fig. 4 represents a MLP with a single hidden layer. MLP is one-way error propagation where back-propagation techniques have been utilized to train and test these weight values. The time complexity of MLP is $O(n^2)$.

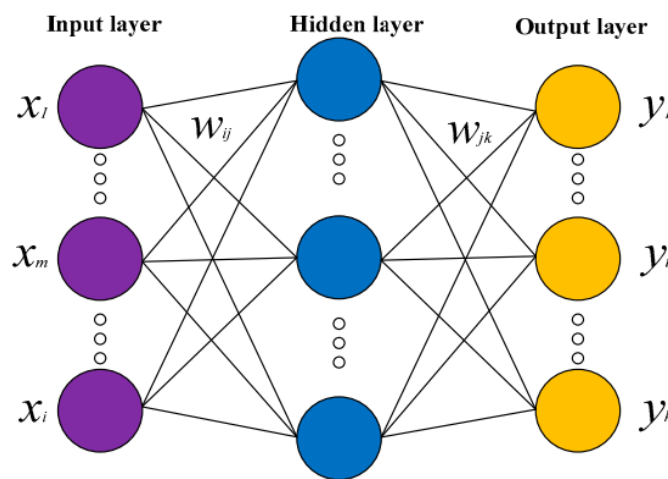


Figure 4: MLP Model with 1 input layer, 1 hidden layer, and 1 output layer.

3.2.5 Random Forests

The Random Forests (RF) classifier is a type of ensemble method that combines multiple decision tree predictions. In RF, the trees are generated randomly by selecting attributes at each node. The output of the ensemble is tree votes with the most popular class. The pseudo-code of the Random Forest ensemble is presented in Tab. 1. The time complexity of Random Forest of size T and maximum depth D (excluding the root) is $O(T \times D)$ [54].

Table 1: Pseudo-code of RF model.

Input:					
D' , a set of d training sets;					
D'' , a set of d test sets;					
k , number of models in the ensemble;					
f , number of attributes that are used to split the D'					
M_i , set of <i>base classifiers</i> .					
Output:					
The ensemble (a composite model M^*)					
Steps:					
(1) for $i = 1$ to k do					
(2) - create bootstrap sample of D' with replacement, D_i ;					
(3) - construct a decision tree classifier by selecting randomly the attributes f ;					
(4) - use CART method to grow the trees.					
(5) end for					
//To use the ensemble for classifying data on test set:					
(1) each of the k models (decision tree) classifies D'' and return the majority vote					

The random forest method is more robust to errors and outliers. Therefore, the problem of over-fitting is not faced. The accuracy of the model depends mainly on the strength of the base classifiers and measure of the dependence between them [55].

4 Experimental Results

The experiments were conducted such that 10-fold cross validation was applied for each classifier. The performance of each classifier was measured by the accuracy, precision, recall and F-score. Tab. 2 to 12 show the experimental results of several machine learning methods both with and without different feature selection methods.

Table 2: The performance of classifiers without features selection

Performance →		Accuracy	Precision	Recall	F-score
Classifier ↓					
Naïve Bayes		0.829	0.830	0.829	0.829
Support	c-SVM	0.779	0.781	0.779	0.779
Vector	nu-SVM	0.779	0.779	0.779	0.779
Machine					
Multilayer Perceptron		0.767	0.767	0.767	0.767
K-nearest Neighbor		0.800	80.00	80.00	80.90
Random Forest		0.800	80.00	80.00	80.00

Table 3: Performance of classifiers with CfsSubsetEval Feature Selection Combinations

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	<u>0.817</u>	<u>0.817</u>	<u>0.817</u>	<u>0.817</u>
	c-SVM	0.767	0.767	0.767	0.767
	Nu-SVM	0.762	0.763	0.763	0.762
	MLP	0.758	0.759	0.758	0.758
	K-NN	0.733	0.733	0.733	0.733
	RF	0.804	0.805	0.804	0.804
Greedy Stepwise	NB	0.825	0.825	0.825	0.825
	c-SVM	0.746	0.746	0.746	0.746
	Nu-SVM	0.754	0.754	0.754	0.754
	MLP	0.746	0.746	0.746	0.746
	K-NN	0.742	0.743	0.742	0.741
	RF	<u>0.829</u>	<u>0.832</u>	<u>0.829</u>	<u>0.829</u>
PSO Method	NB	<u>0.821</u>	0.821	<u>0.821</u>	<u>0.821</u>
	c-SVM	0.754	0.754	0.754	0.754
	Nu-SVM	0.746	0.746	0.746	0.746
	MLP	0.738	0.738	0.738	0.737
	K-NN	0.700	0.701	0.700	0.700
	RF	<u>0.821</u>	<u>0.822</u>	<u>0.821</u>	<u>0.821</u>

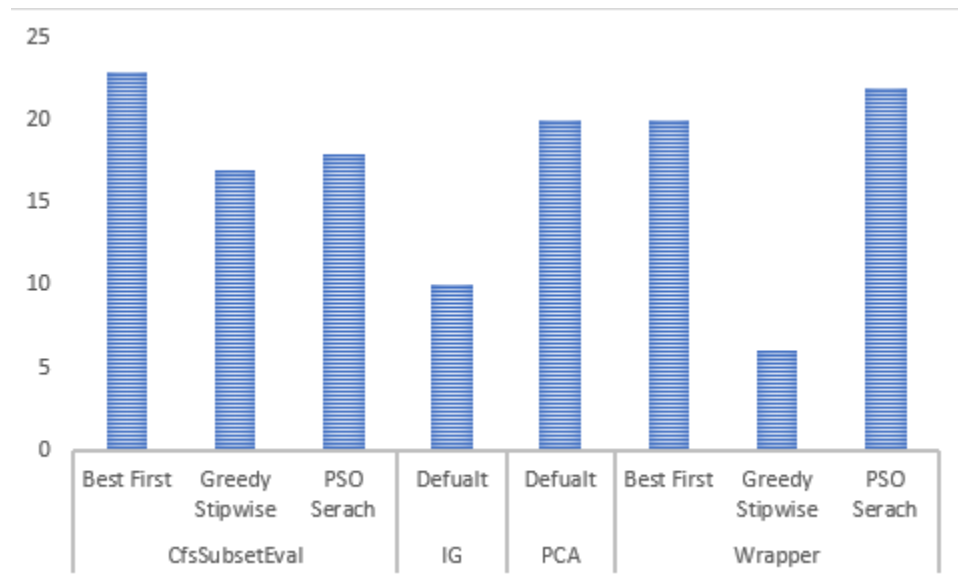


Figure 5: Number of remaining features after applying features selection methods

Table 4: Performance of classifiers with features selection based on information gain

Classifier	Accuracy	Precision	Recall	F-score
NB	0.808	0.808	0.808	0.808
c-SVM	0.696	0.697	0.696	0.695
Nu-SVM	0.700	0.701	0.700	0.700
MLP	0.738	0.738	0.738	0.737
K-NN	0.708	0.712	0.708	0.707
RF	0.800	0.801	0.800	0.800

Table 5: Performance of classifier with features selection based on PCA

Classifier	Accuracy	Precision	Recall	F-score
NB	0.733	0.734	0.733	0.733
c-SVM	0.833	0.834	0.833	0.833
Nu-SVM	<u>0.838</u>	<u>0.838</u>	<u>0.838</u>	<u>0.837</u>
MLP	0.746	0.746	0.746	0.746
K-NN	0.692	0.696	0.692	0.690
RF	0.804	0.807	0.804	0.804

Table 6: Summary of the accuracy of classifiers with filter-based features selection methods

Classifier	Accuracy (before FS)	Best Accuracy with FS	Method
NB	<u>82.92</u>	82.92	IG
c-SVM	77.92	83.33	PCA
nu-SVM	77.92	<u>83.75</u>	PCA
MLP	76.67	76.67	IG
K-nn	80.00	80.00	IG
RF	80.00	80.42	CfsSubsetEval With Greedy Stepwise

Table 7: Performance of classifiers for wrapper-based method with Naïve Bayes as base classifier

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best / Greedy Stepwise	NB	0.838	0.838	0.838	0.837
	c-SVM	0.721	0.721	0.721	0.721
	Nu-SVM	0.746	0.746	0.746	0.746
	MLP	0.800	0.800	0.800	0.800
	K-nn	0.758	0.758	0.758	0.758
	RF	0.792	0.792	0.792	0.792
PSO Method	NB	0.854	0.855	0.854	0.854
	c-SVM	0.717	0.718	0.717	0.716
	Nu-SVM	0.721	0.721	0.721	0.721

MLP	0.761	0.763	0.763	0.762
K-nn	0.792	0.792	0.792	0.792
RF	0.738	0.738	0.738	0.737

Table 8: Performance of classifiers for wrapper-based methods with c-SVM as base classifier

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	0.833	0.834	0.833	0.833
	c-SVM	<u>0.850</u>	<u>0.850</u>	<u>0.850</u>	<u>0.850</u>
	Nu-SVM	0.846	0.846	0.846	0.846
	MLP	0.788	0.788	0.788	0.787
	K-NN	0.775	0.775	0.775	0.775
	RF	0.817	0.820	0.817	0.816
Greedy Stepwise	NB	0.800	0.803	0.800	0.799
	c-SVM	<u>0.850</u>	<u>0.851</u>	<u>0.850</u>	<u>0.850</u>
	Nu-SVM	<u>0.850</u>	<u>0.851</u>	<u>0.850</u>	<u>0.850</u>
	MLP	0.838	0.839	0.838	0.837
	K-NN	0.775	0.783	0.775	0.773
	RF	0.813	0.814	0.813	0.812
PSO Method	NB	0.8212	0.824	0.821	0.82
	c-SVM	0.842	0.842	0.842	0.842
	Nu-SVM	<u>0.846</u>	<u>0.846</u>	<u>0.846</u>	<u>0.846</u>
	MLP	0.808	0.810	0.808	0.808
	K-NN	0.783	0.784	0.783	0.783
	RF	0.821	0.823	0.821	0.821

Table 9: Performance of classifiers for wrapper-based methods with nu-SVM as base classifier

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	0.754	0.778	0.754	0.749
	c-SVM	0.779	0.786	0.779	0.778
	Nu-SVM	0.779	0.782	0.779	0.779
	MLP	0.808	<u>0.813</u>	<u>0.808</u>	<u>0.808</u>
	K-NN	0.775	0.775	0.775	0.775
	RF	0.779	0.780	0.779	0.779
Greedy Stepwise	NB	0.754	0.778	0.754	0.749
	c-SVM	0.779	0.786	0.779	0.778
	Nu-SVM	0.779	0.782	0.779	0.779
	MLP	0.808	<u>0.813</u>	<u>0.808</u>	<u>0.808</u>
	K-NN	0.775	0.775	0.775	0.775

PSO Method	RF	0.779	0.780	0.779	0.779
	NB	0.813	0.813	0.813	0.812
	c-SVM	0.817	<u>0.817</u>	0.817	0.817
	Nu-SVM	0.796	0.796	0.796	0.796
	MLP	0.746	0.746	0.746	0.746
	K-NN	0.817	0.818	0.817	<u>0.816</u>
	RF	0.813	0.813	0.813	0.812

Table 10: Performance of classifiers for wrapper-based methods with MLP as base classifier

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	0.813	0.819	0.813	0.812
	c-SVM	0.733	0.741	0.733	0.731
	Nu-SVM	0.792	0.800	0.792	0.790
	MLP	0.829	0.832	0.829	0.829
	K-NN	0.771	0.771	0.771	0.771
	RF	0.825	0.825	0.825	0.825
Greedy Stepwise	NB	0.813	0.814	0.813	0.812
	c-SVM	0.733	0.739	0.733	0.732
	Nu-SVM	0.800	0.809	0.800	0.799
	MLP	0.825	<u>0.826</u>	<u>0.825</u>	<u>0.825</u>
	K-NN	0.792	0.739	0.792	0.791
	RF	0.821	0.824	0.821	0.820
PSO Method	NB	0.800	0.804	0.800	0.799
	c-SVM	0.754	0.756	0.754	0.754
	Nu-SVM	0.771	0.771	0.771	0.771
	MLP	0.829	<u>0.830</u>	0.829	0.829
	K-NN	0.800	0.800	0.800	0.800
	RF	0.817	0.817	0.817	0.817

Table 11: Performance of classifiers when wrapper-based methods with K-NN are applied

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	0.763	0.781	0.763	0.758
	c-SVM	0.742	0.746	0.742	0.741
	Nu-SVM	0.771	0.773	0.771	0.77
	MLP	0.792	0.792	0.792	0.792
	K-NN	0.883	<u>0.883</u>	0.883	0.883
	RF	0.821	0.825	0.821	0.82

Greedy Stepwise	NB	0.788	0.801	0.788	0.785
	c-SVM	0.704	0.706	0.704	0.704
	Nu-SVM	0.679	0.68	0.679	0.679
	MLP	0.783	0.786	0.783	0.783
	K-NN	0.846	<u>0.848</u>	<u>0.846</u>	<u>0.846</u>
	RF	0.788	0.790	0.788	0.787
PSO Method	NB	0.775	0.793	0.775	0.771
	c-SVM	0.754	0.756	0.754	0.754
	Nu-SVM	0.758	0.759	0.758	0.758
	MLP	0.850	0.850	0.850	0.850
	K-NN	0.883	<u>0.884</u>	<u>0.883</u>	<u>0.883</u>
	RF	0.804	0.806	0.804	0.804

Table 12: Performance of classifiers for wrapper-based methods with RF as base classifier

Search Method	Classifier	Accuracy	Precision	Recall	F-score
First Best	NB	0.775	0.776	0.775	0.775
	c-SVM	0.721	0.723	0.721	0.720
	Nu-SVM	0.767	0.767	0.767	0.767
	MLP	0.796	0.796	0.796	0.796
	K-nn	0.788	0.778	0.788	0.787
	RF	0.838	<u>0.839</u>	<u>0.838</u>	<u>0.837</u>
Greedy Stepwise	NB	0.800	0.800	0.800	0.800
	c-SVM	0.721	0.724	0.721	0.720
	Nu-SVM	0.729	0.731	0.729	0.729
	MLP	0.779	0.780	0.779	0.779
	K-nn	0.779	0.731	0.729	0.729
	RF	0.808	<u>0.809</u>	<u>0.808</u>	<u>0.808</u>
PSO Method	NB	0.817	0.818	0.817	0.816
	c-SVM	0.742	0.743	0.742	0.741
	Nu-SVM	0.763	0.763	0.763	0.762
	MLP	0.838	<u>0.838</u>	<u>0.838</u>	<u>0.837</u>
	K-nn	0.800	0.800	0.800	0.800
	RF	0.838	<u>0.839</u>	<u>0.838</u>	<u>0.837</u>

Discussion

Tab. 2, shows the performance of all classifiers used before applying features selecting methods. The results showed Naïve Bayes obtained the best performance using all evaluation measures compared to the other classifiers. It obtained 82.92%, 83.30%, 82.90% and 82.90 % for accuracy, precision, recall and F-score respectively.

The number of features was reduced using correlation based feature selection (CfsSubsetEval) method to 23, 17, 18 for the search methods of First Best, Greedy Stepwise and POS respectively, as shown in Fig. 5. The performance of with CfsSubsetEval combinations for each classifier is shown in Tab. 3. The results showed that no improvements were obtained by most of the combinations, except for RF with Greedy Stepwise and POS methods.

Tab. 4 showed the performance of classifiers used when features selection method based on information gain was applied. As shown in Fig. 5, the number of features was reduced to 10. The results showed that no improvements were reported on the performance of all classifiers after applying this feature selection method.

In addition, Tab. 5 shows the performance of all classifiers when features selection method based on PCA was applied. The results showed that only SVM methods obtained better performance after applying this features selection method. The number of features was reduced to 20 as shown in Fig. 5.

Tab. 6 summarizes the performance of filter based features selection methods. The results showed that feature selections with PCA obtained the best performance when SVM classifier was applied.

Tab. 7-12 show the performance of wrapper-based features selection methods using different base classifiers. In each table, First Best, Greedy Stepwise and PSO search methods were applied.

Tab. 7 showed that, when Naïve Bayes was used as the base classifier for wrapper-based feature selection method, the performance of NB using PSO search method was enhanced to 0.854, 0.855, 0.854 and 0.854 for accuracy, precision, recall and F-score respectively. The performance of the other classifiers using this method was reduced.

Tab. 8 shows the performance of classifiers when the wrapper-based features selection method with c-SVM as the base classifier was applied. The results showed the enhancements obtained by all classifiers using all search methods. However, the best performance was obtained by SVM using First Best and Greedy Stepwise search methods.

However, Tab. 9 shows the performance of classifiers when wrapper-based features selection method with nu-SVM as the base classifier was applied. The results showed that the enhancements were obtained by applying c-SVM, K-NN and RF, especially when the POS search method was used.

In addition, Tab. 10 shows the performance of classifiers when wrapper-based features selection method with MLP as base classifier was applied. The results showed that the enhancements were obtained by applying MLP and RF for the three search methods. The best results were obtained using MLP classifier.

Moreover, Tab. 11 shows the performance of classifiers when wrapper-based features selection method with K-NN as base classifier was applied. The results showed that the enhancements were obtained by applying K-NN and RF for the First Best and POS search methods. The best results were obtained using K-NN classifier with accuracy, precision, recall and F-scores of 0.883, 0.884, 0.883 and 0.883 respectively.

Tab. 12 shows the performance of classifiers when wrapper-based features selection method with RF as base classifier was applied. The results showed that the enhancements were obtained by applying MLP and RF for the three search methods. The best results were obtained using RF classifier.

Tab. 13 shows a comparison of different wrapper-based features selection methods (using different base classifiers). The results showed that the best performing classifier was K-NN associated with the wrapper-based feature selection with KNN as base classifier, obtaining 88.33% accuracy. The number of features was reduced (with the best performance obtained) to 20, 5 and 22 using First Best, Greedy Stepwise and PSO search methods.

Table 13: Best Results for wrapper-based techniques

Classifier	Accuracy (before FS)	Accuracy with FS	Method
NB	82.92	85.42	PSO

c-SVM	77.92	85.00	c-SVS/nu-SVM with Greedy Stepwise
nu-SVM	77.92	81.67	PSO with RF/MLP
MLP	76.67	82.92	Best First / PSO
K-NN	80.00	88.33	Best First / PSO
RF	80.00	83.75	Best First, PSO with MLP/RF

Finally, Tab. 14 shows a comparison of using different features selection methods (filter and wrapper base methods). It shows that the best performance was obtained by K-NN classifier associated with wrapper-based feature selection method with K-NN as base classifier and using Best First and PSO search method.

Table 14: Comparison between Filter-based and Wrapper-based techniques

Classifier	Baseline	Filter-based	Method	Wrapper-based	Method
NB	82.92	82.92	IG	85.42	PSO
c-SVM	77.92	83.33	PCA	85.00	c-SVS/nu-SVM with Greedy Stepwise
nu-SVM	77.92	83.75	PCA	81.67	PSO with RF/MLP
MLP	76.67	76.67	IG	82.92	Best First / PSO
K-nn	80.00	80.00	IG	88.33	Best First / PSO
RF	80.00	80.42	CfsSubsetEval With Greedy Stepwise	83.75	Best First, PSO with MLP/RF

For this paper a comparison has been conducted between the best performing methods and the previous studies on predicting Parkinson’s disease using the same dataset, and other datasets, as shown in Tab. 15. The comparison results showed that the best performing method (K-NN classifier associated with wrapper-based feature selection method with K-NN as base classifier and using Best First and PSO search method) obtained comparable and superior results.

Table 15: Comparison with previous studies

Methods	Accuracy with FS	Dataset Used
The best performing method	88.33	UCI Dataset [14]
[21]	86.2	UCI Dataset [14]
[56]	57.5%	Other Dataset
[57]	82.5%	Other Dataset
[58]	82.5%	Other Dataset
[59]	87.5%	Other Dataset

5 Conclusions and Future Works

This paper examined the performance of several classifiers with filter-based and wrapper-based features selections methods to enhance the diagnosis of Parkinson's disease. Different evaluation metrics were used including accuracy, precision, recall and F-score. The experiments compared the performance of machine learning on original and filtered datasets. The results showed that wrapper-based features selection method with K-NN enhanced the performance of predicting Parkinson's disease, with the accuracy reached to 88.33%. In future work, more machine learning and deep learning methods could be applied with these combinations of features selection methods. In addition, other features selection methods could be investigated to improve the performance of predicting Parkinson's disease.

Acknowledgement: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work; project number (77 /442). Also, the authors would like to extend their appreciation to Taibah University for its supervision support.

Funding Statement: This research was funded by the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia under the project number (77 /442).

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

- [1] L. V. Kalia and A. E. Lang, "Parkinson's disease," *The Lancet*, vol. 386, no. 9996, pp. 896–912, 2015.
- [2] J. P. Iannotti and R. Parker, "The netter collection of medical illustrations-musculoskeletal system," in *Elsevier Health Sciences*, 2nd edition, vol. 6, 2013.
- [3] M. Fjodorova, E.M. Torres and S.B. Dunnett, "Transplantation site influences the phenotypic differentiation of dopamine neurons in ventral mesencephalic grafts in Parkinsonian rats," *Experimental Neurology*, vol. 291, pp. 8-19, 2017.
- [4] A. Jamak, A. Savatić and M. Can, "Principal component analysis for authorship attribution," *Business Systems Research*, vol. 3, pp. 49-56, 2012.
- [5] M. Can, "Neural networks to diagnose the Parkinson's disease," *Southeast Europe Journal of Soft Computing*, vol.2, no. 1, 2013.
- [6] L.V. Kalia, S. K. Kalia and A. E. Lang, "Disease modifying strategies for Parkinson's disease," *Movement Disorders*, vol. 30, pp. 1442-1450, 2015.
- [7] N. Singh, V. Pillay and Y.E. Choonara, "Advances in the treatment of Parkinson's disease," *Progress in Neurobiology*, vol. 81, pp. 29-44, 2007.
- [8] C. Camara, P. Isasi, K. Warwick, V. Ruiz, T. Aziz *et al.*, "Resting tremor classification and detection in Parkinson's disease patients," *Biomedical Signal Processing and Control*, vol. 16, pp. 88-97, 2015.
- [9] B. Keltch, L. Yuan and B. Coskun, "Comparison of AI techniques for prediction of liver fibrosis in hepatitis patients," *Journal of Medical Systems*, vol. 38, no. 8, pp. 1-8, 2014.
- [10] M. Nasr, K. El-Bahnasy, M. Hamdy and S.M. Kamal, "A novel model based on non-invasive methods for prediction of liver fibrosis," In *13th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt, 27-28 Dec. 2017.
- [11] S. Guerlain, D.E. Brown and C. Mastrangelo, "Intelligent decision support systems", in *Proceeding of SMC 2000 conference proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions'. IEEE International Conference on Systems, Man, and Cybernetics*, Nashville, TN, USA, 2000, pp. 193438.
- [12] F. Meherwar and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *Journal of Intelligent Learning Systems and Applications*, vol. 9, no.1, 2017.
- [13] S. Brunak, "The Bioinformatics: Machine Learning Approach", *MIT press*, 2001.
- [14] L. Naranjo, C.J. Perez, Y. Campos-Roca and J. Martin, "Addressing voice recording replications for Parkinson's disease detection," *Expert Systems with Applications*, vol. 46, pp. 286-292, 2016.
- [15] B. Harish, D. C. Hoyle and S. Singh, "Machine learning in bioinformatics: a brief survey and recommendations

- for practitioners,” *Computers in Biology and Medicine*, vol. 36, no.10, pp. 1104-1125, 2006.
- [16] R. Fernandez-Millan, J. A. Medina-Merodio, R. B. Plata, J. J. Martinez-Herraiz and J. M. Gutierrez-Martinez, “A laboratory test expert system for clinical diagnosis support in primary health care,” *Applied Sciences*, vol.5, no.3, pp. 222-240, 2015.
- [17] Y. Saeys, I. Inza and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics*, vol. 23, no.19, pp. 2507-2517, 2007.
- [18] Z. M. Hira and D. F. Gillies, “A review of feature selection and feature extraction methods applied on microarray data,” *Advances in Bioinformatics*. Pp.198363, 2015.
- [19] P. Drotár and Z. Smékal, “Comparison of stability measures for feature selection,” in *SAMI 2015IEEE 13th International Symposium on Applied Machine Intelligence and Informatics*, Herlany, Slovakia, 2015, pp.71-75.
- [20] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri *et al.*, “Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy,” *Journal of Neuroscience Methods*, vol. 222, pp.230-237, 2014.
- [21] L. Naranjo, C. J. Pérez, J. Martín and Y. Campos-Roca, “A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications,” *Computer Methods and Programs in Biomedicine*, vol. 142, pp.147-156, 2017.
- [22] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan and V. Venkatraman, “Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease,” *Future Generation Computer Systems*, vol. 83, pp.366-373, 2018.
- [23] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar *et al.*, “A deep learning approach for Parkinson's disease diagnosis from EEG signals,” *Neural Computing and Applications*, vol. 32, no. 15, pp. 10927-10933, 2020.
- [24] S.A Mostafa, A. Mustapha, M.A. Mohammed, R.I. Hamed, N. Arunkumar *et al.*, “Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease,” *Cognitive Systems Research*, vol. 54, pp.90-99, 2019.
- [25] M. R. Salmanpour, M. Shamsaei, A. Saberi, S. Setayeshi, I. S. Klyuzhin *et al.*, “Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease,” *Computers in Biology and Medicine*, 111, pp. 103347, 2019.
- [26] A. U. Haq, J. P. Li, M. H. Memon, A. Malik, T. Ahmad *et al.*, “Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings,” *IEEE Access*, vol. 7, pp.37718-37734, 2019.
- [27] C. Gao, H. Sun, T. Wang, M. Tang, N. I. Bohnen *et al.*, “Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease,” *Scientific Reports*, vol. 8, no.1, pp.1-21, 2018.
- [28] H. Khalid, M. Hussain, M. A. A. Ghamdi, T. Khalid, K. Khalid *et al.*, “A comparative systematic literature review on knee bone reports from MRI, X-rays and CT scans using deep learning and machine learning methodologies,” *Diagnostics*, vol. 10, no. 8, pp. 118–139, 2020.
- [29] M. A. Khan, S. Abbas, K. M. Khan, M. A. Ghamdi and A.Rehman, "Intelligent forecasting model of covid-19 novel coronavirus outbreak empowered with deep extreme learning machine," *CMC-Computers, Materials & Continua*, vol. 64, no. 3, pp. 1329–1342, 2020.
- [30] A. H. Khan, M. A. Khan, S. Abbas, S. Y. Siddiqui, M. A. Saeed *et al.*, “Simulation, modeling, and optimization of intelligent kidney disease predication empowered with computational intelligence approaches,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1399–1412, 2021.
- [31] G. Ahmad, S. Alanazi, M. Alruwaili, F. Ahmad, M. A. Khan *et al.*, “Intelligent ammunition detection and classification system using convolutional neural network,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2585–2600, 2021.
- [32] B. Shoaib, Y. Javed, M. A. Khan, F. Ahmad, M. Majeed *et al.*, “Prediction of time series empowered with a novel sreklr algorithm,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1413–1427, 2021.
- [33] S. Aftab, S. Alanazi, M. Ahmad, M. A. Khan, A. Fatima *et al.*, “Cloud-based diabetes decision support system using machine learning fusion,” *Computers, Materials & Continua*, vol. 68, no. 1, pp. 1341–1357, 2021.
- [34] M. W. Nadeem, H. G. Goh, M. A. Khan, M. Hussain, M. F. Mushtaq *et al.*, “Fusion-based machine learning architecture for heart disease prediction,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2481–2496, 2021.
- [35] S. Y. Siddiqui, I. Naseer, M. A. Khan, M. F. Mushtaq, R. A. Naqvi *et al.*, “Intelligent breast cancer prediction empowered with fusion and deep learning,” *Computers, Materials and Continua*, vol. 67, no. 1, pp. 1033–1049, 2021.
- [36] R. A. Naqvi, M. F. Mushtaq, N. A. Mian, M. A. Khan, M. A. Yousaf *et al.*, “Coronavirus: A mild virus turned

- deadly infection,” *Computers, Materials and Continua*, vol. 67, no. 2, pp. 2631–2646, 2021.
- [37] F. Alhaidari, S. H. Almotiri, M. A. A. Ghamdi, M. A. Khan, A. Rehman *et al.*, “Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach,” *Computers, Materials and Continua*, vol. 67, no. 1, pp. 1269–1285, 2021.
- [38] M. W. Nadeem, M. A. A. Ghamdi, M. Hussain, M. A. Khan, K. M. Khan *et al.*, “Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges,” *Brain Sciences*, vol. 10, no. 2, pp. 118–139, 2020.
- [39] Y. Masoudi-Sobhanzadeh, H. Motieghader and A. Masoudi-Nejad, “FeatureSelect: a software for feature selection based on machine learning approaches,” *BMC Bioinformatics*, vol.20, no. 1, pp. 170, 2019.
- [40] M. Rahmaninia and P. Moradi, “OSFSMI: online stream feature selection method based on mutual information,” *Applied Soft Computing*, vol. 68, pp. 733-746, 2018.
- [41] S. Pourbahrami, “Improving PSO Global Method for Feature Selection According to Iterations Global Search and Chaotic Theory,” *arXiv preprint arXiv:1811.08701*.
- [42] L. Yu and H. Liu, “Feature selection for high-dimensional data: a fast correlation-based filter solution,” in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, Washington DC, USA, 2003, pp. 856-863.
- [43] A.L. Blum and P. Langley, “Relevance selection of relevant features and examples in machine learning,” *Artificial Intelligence*, vol. 97, pp.245-271, 1997.
- [44] L.E. Raileanu and K. Stoffel, “Theoretical comparison between the GINI index and information gain criteria,” *Annals of Mathematics and Artificial Intelligence*, vol. 41, pp. 77-93, 2004.
- [45] I. T. Jolliffe, “*Principal Component Analysis*”, 2nd Edition, Springer, New York, 2002
- [46] M.M. Kabir, M. Shahjahan and K. Murase, “A new local search based hybrid genetic algorithm for feature selection”, *Neurocomputing*, vol. 74, no. 17, pp. 2914-2928, 2011.
- [47] L.-F. Chen, C.-T. Su, K.-H. Chen and P.-C Wang, “Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis,” *Neural Computing and Applications*, vol. 21, pp. 2087-2096, 2012.
- [48] B. Chen, L. Chen and Y. Chen, “Efficient ant colony optimization for image feature selection,” *Signal Processing*, vol. 93, pp. 1566-1576, 2013.
- [49] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–97, 1995.
- [50] K. Shaukat, S. Luo, S. Chen and D. Liu, “Cyber threat detection using machine learning techniques: a performance evaluation perspective”. In *International Conference on Cyber Warfare and Security (ICWS)*, Islamabad, Pakistan, 2020, pp. 1-6.
- [51] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.
- [52] J. Yim, “Introducing a decision tree-based indoor positioning technique,” *Expert Systems with Applications*, vol. 34, no. 2, pp.1296-1302, 2008.
- [53] H. Sayoud, “Automatic Speaker Recognition-Connexionnist Approach,” PhD thesis, USTHB University, Algiers, 2003.
- [54] X. Solé, A. Ramisa and C. Torras, “Evaluation of random forests on large-scale classification problems using a bag-of-visual-words representation,” In *Artificial Intelligence Research and Development*. IOS Press, 2014
- [55] J. Han, J. Pei and M. Kamber, “*Data Mining: Concepts and Techniques*,” 3rd edition, Elsevier, 2011.
- [56] I. Cantürk and F. Karabiber, “A machine learning system for the diagnosis of Parkinson’s disease from speech signals and its application to multiple speech signal types,” *Arabian Journal for Science and Engineering*, vol. 41, no. 12, pp. 5049–5059, 2016.
- [57] Y. Li, C. Zhang, Y. Jia, P. Wang, X. Zhang *et al.*, “Simultaneous learning of speech feature and segment for classification of Parkinson disease,” in *Proceeding of IEEE 19th International Conference of e-Health Networking, Application and Services (Healthcom)*, Dalian, China, Oct. 2017, pp. 1–6.
- [58] A. Benba, A. Jilbab and A. Hammouch, “Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson’s disease and healthy people,” *International Journal of Speech Technology*, vol. 19, no. 3, pp. 449–456, 2016.
- [59] A. Benba, A. Jilbab and A. Hammouch, “Using human factor cepstral coefficient on multiple types of voice recordings for detecting patients with Parkinson’s disease,” *IRBM*, vol. 38, no. 6, pp. 346–351, 2017.