Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID APPROACH FOR ENHANCING THE CLASSIFICATION OF THE PARKINSON'S DISEASE USING SWARM OPTIMIZATION

MANAL A. ABDEL-FATTAH, RAHMA HUSSEIN EID, AHMED ELSAYED YAKOUB

Faculty Of Computers and Artificial Intelligence, Department of Information System, Helwan University, Egypt

E-mail: rahma.hussein.eid@gmail.com

ABSTRACT

The second most common neurodegenerative ailment after Alzheimer's disease is Parkinson's disease. It always affects adults from the age of sixty and more. Parkinson's disease symptoms are unnoticed in the early stages, but the symptoms become more apparent as the disease progresses. Recent studies have shown that the disease symptoms can appear in the form of vocal disturbances in the earlier stages and can be used to diagnose Parkinson's disease. This paper proposes an approach for detecting Parkinson's disease (PD) using speech signals. The motivation of this approach is to improve the detection of Parkinson's early diagnosis by determining the most effective speech examinations instead of produce a huge number of examinations to resistance to the disease at an early stage. As feature selection and swarm algorithms play a vital role during classification, this paper has proposed a hybrid approach based on the Emperor Penguin Colony (EPC) swarm algorithm with Correlation-based Feature Selection (CFC), which is called CEPC. A Parkinson's disease classification dataset consisting of 756 voice measures was used in this study. Before using the proposed approach, five classification algorithms were used to compare accuracy results. Also, the Ensemble classifier has been used in this paper. The CEPC proposed approach provides an improvement in the accuracy of results. An accuracy of 89.4% is obtained by the ensemble classifier, which is higher than some recent work.

Keywords: *Swarm Intelligence Algorithms; Feature Selection; Metaheuristic Algorithms; Emperor Penguins Colony; Classification*

1. INTRODUCTION

Neuropsychiatric disorders such as Epilepsy, Parkinson's disease (PD), and Alzheimer's disease are brain disorders that have an impact on human brain functioning in some way[1]. Parkinson's disease follows Alzheimer's disease of popularity among neurodegenerative diseases. Parkinson's symptoms appear when brain cells that produce dopamine begin to die, but the cause of this deterioration is unknown [2]. It is impossible to prevent Parkinson's, but an early diagnosis can reduce the severity of the disease and its effects and control the disease's symptoms. Shaking in the hands or arms, slowness during walking, changes in handwriting, and changes in speech indicate Parkinson's disease, which is minor at the onset, then the disease develops and the symptoms become more severe [3]. Diagnosing the start of Parkinson's is not an easy task because symptoms are unnoticed in the beginning. Therefore, computer-assisted technologies help in diagnosing Parkinson's disease. Data mining techniques and machine learning methods have been used to diagnose Parkinson's and decide whether a patient has PD or not.

There are many techniques for PD diagnosis using machine learning techniques, including feature selection, classification methods, and swarm algorithms. Data mining is a technique for extracting patterns and knowledge from large amounts of data. In many fields of research, finding hidden information from large datasets is critical to analysing and retrieving the results from the datasets in a short time. Classification, regression, and clustering are data mining techniques that are used to make correct or accurate decisions [4].

Classification analysis is used in this study to predict an early automatic diagnosis of patients in a short amount of time. Decision tree, Naïve Bayes, Random Forest, K-Nearest Neighbor, and Support Vector Machine algorithms are all used to build the proposed approach.

A decision tree is a classification algorithm that is simple to understand and interpret. It looks like a tree structure. The decision tree starts from the root node to leaf nodes, where leaf nodes represent the class that is predicted [4]. Naive Bayes is used for classification based on Bayes' theorem. For this purpose, to categorize data sets, Naive Bayes is used to compute the probability of each class based on conditional probability [4][5]. The Random Forest classification method uses a randomly selected subset of many trees. Specifically, it combines the votes from various decision trees to determine the test subset final class [6]. K-Nearest Neighbor (KNN) is a simple and effective classifier that computes the distance between the new sample and all other samples. For calculating the distance, KNN uses many different measures like Euclidean, Manhattan, and Minkowski[7]. Support Vector Machine (SVM) is a method of classification that tries to improve accuracy by increasing the hyperplane's margin that separates the dataset classes. SVM uses kernel functions such as quadratic, polynomial, and radial basis [4].

Typically, a huge number of features are used to represent the data in the real world, so some irrelevant or redundant features may be found in the datasets. Therefore, feature selection is an important and complicated task in machine learning and data mining approaches that are used to avoid the curse of dimensionality and optimize classification accuracy. Furthermore, removing the features that are not relevant or redundant increases the learning process, enhances prediction performance, and improves data interpretation [8].

Swarm intelligence algorithms are nature-inspired algorithms that simulate the behavior of a group of animals [9]. Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Bat algorithm (BS), and Emperor Penguin Colony (EPC) are some examples of those algorithms. The Emperor Penguin Colony (EPC) algorithm is one of the most recent swarm-based optimization and metaheuristic algorithms. The algorithm was developed by Sasan Harif [10] in 2019. EPC mimics the design of emperor penguins' huddling behavior.

This paper proposes a hybrid approach (CEPC) to diagnose Parkinson's based on combining the Correlation-based Feature Selection (CFC) method and the Emperor Penguin Colony (EPC) swarm algorithm. The contribution of this approach is to provide good accuracy to differentiate between PD patients and healthy person.

The rest of the paper is organized as follows: Section 2 provides the literature review about the feature selection algorithms and swarm algorithms. section 3 presents the main structure of the proposed approach and discusses its components. The experimental results are reported in section 4. Finally, the conclusion and some future work make up section 5.

2. LITERATURE REVIEW

Researchers in the literature have proposed a variety methodology for PD diagnosis as they focused on the feature selection and machine learning techniques. The field of study here choose the speech symptoms to detect the Parkinson disease. As the speech problems is one of the first symptoms to appear in Parkinson patients. Sakar, C. Okan et al. [1] proposed an approach for the detection of PD patients. In this study, the UCI Parkinson’s disease classification dataset was employed, which has 252 samples and 757 features. The top 50 features were chosen by applying the mRMR feature selection method. This paper employed 7 classifiers, SVM (Linear), SVM (RBF), Multilayer perceptron, Naïve Bayes, Logistic regression random forest, and KNN. Also, used the ensemble learning to combine the predictions of the 7 classifiers, in 2 ways (voting and stacking with linear kernel SVM). Bchir, Ouiem [11] suggested a new classifier called the Gaussian mixture models (GMM) classifier strategy for Parkinson’s disease diagnosis. GMM classifier incorporates relevance feature weighting for the other classifiers. Naïve Bayes, LR, KNN, multilayer perceptron, RF, and SVM (Linear and RBF) have been applied to 252 samples dataset. The mRMR feature selection technique has been chosen to determine the best relevant features. Anisha, C. D. and Arulanand, N. [12] present a proposed framework for performs an early prediction of PD using ensemble classifiers. The research use a speech dataset from the UCI library which contains voice measurements for 252 cases. They used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to integrate the correlated features. The ensemble classifiers based on bagging classifier, Adaptive Boosting (AdaBoost) classifier, Gradient Boosting Machine (GBM) classifier and Extreme Gradient Boosting (XGBoost) classifier have been fed with the optimal parameters which obtained using PCA and LDA. 5-fold and 10-fold validation technique are used to validate the model. Polat, Kemal and Nour, Majid [13] employed the UCI-Irvine Parkinson's acoustic characteristics dataset with 80 samples. This paper used the one against all (OGA) as a data sampling strategy to divide the dataset into five equal parts. For classification, used k-NN, Logistic Regression, and support vector machine. Goyal, Jinee et al. [14] Provide a two-stage feature selection model, which employed the UCI Parkinson's speech dataset. The dataset has 192 cases and 22 characteristics. The model combines the Genetic Algorithm and Support Vector Machine-Recursive Feature Elimination techniques into one. The outcome was 9 characteristics out of 22. SVM classifier was used to classify the data. Soumaya, Zayrit et al. [15] choose to use genetic algorithm as a feature selection method and support vector machine as a classifier algorithm in this study. The authors used a 34 sample and 21 features dataset. The provided method selected only 14 features. The authors in Goyal, Jinee et al [16] used two datasets. The first dataset has 252 cases and the second dataset has 40 cases. Three different feature selection techniques have been applied to reduce the dimensionality of the dataset. The feature selection techniques are mRMR, GA, and PCA. This paper used a comparison of various classification techniques including SVM, DT, RF, LR, KNN, NB, MLP, AdaBoost, and XGBoost to classify the data. Also the same dataset has been used in Lamba, Rohit et al. [3] the paper suggested a hybrid feature selection methodology called MIRFE-XGBoost. The hybrid approach merges two feature selection methods: mutual information gain method with the recursive feature elimination method. For classification, random forest and XGBoost classifiers were used.

In this study, we aim to build a model to determine the most accurate speech examinations in Parkinson's disease with less time and less cost using a hybrid optimization approach which combined feature selection and swarm algorithm techniques and then improve prediction accuracy using machine learning techniques, While most previous studies have employed the traditional feature selection techniques or depend only on using the swarm algorithms to identify the most practical features used to build their models. In addition, the ensemble method that involved the top three classifiers is proposed to enhance the accuracy of classifying healthy and Parkinson's patients.

3. THE PROPOSED MODEL

A new hybrid approach for enhancing the classification accuracy and determine the most related features of speech examinations in Parkinson's disease for data reduction is proposed in this research. Specifically, the Correlation-based Feature Selection (CFS) method is combined with the EPC swarm algorithm to offer a hybrid approach called CEPC.

The overall architecture of the proposed approach is depicted in Fig 1. The proposed model consists of three phases. The first phase is the data preprocessing phase. In this phase, two preprocessing methods have been applied to the collection of data needed for the study. The second phase includes the hybrid approach in which the Correlation-based Feature Selection (CFS) technique is used and then combined with the EPC swarm algorithm in six different iterations for dimensionality reduction. Finally, the third phase is about building the base classifiers and applying the ensemble classifier. Each framework phase has been discussed in detail in the following subsections.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

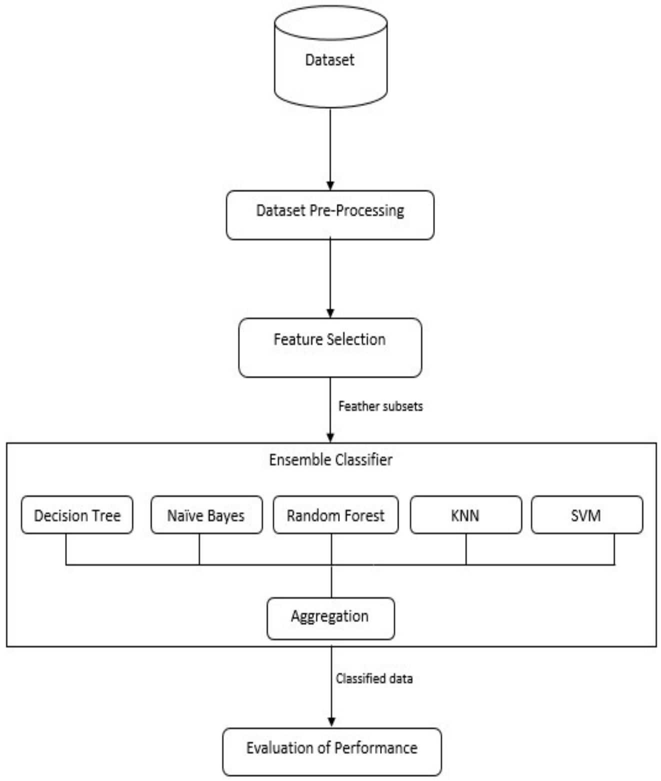

Figure 1: The overall representation Scheme of the Hybrid method

3.1 Data Pre-Processing

In this study, we used the UCI Parkinson's disease dataset for classification [17]. The dataset is imbalanced, as it contains 188 instances belonging to the infected class and only 64 instances belonging to the healthy class. To overcome this issue, the SMOTE technique [18] is employed. In addition, data scaling is done by using the standard scaler normalization, which normalizes the feature set as zero mean and unit variance.

3.2 Feature Selection

Feature selection is the process of selecting the most crucial, relevant, and non-redundant features from an extensive set of features to use in model building. So feature selection techniques are used to remove redundant or irrelevant features and make learning faster [19]. Furthermore, using a small number of features in the classification step can reduce computing time while enhancing the system's efficiency. Correlation-based Feature Selection (CFS) combined with Emperor penguin colony (EPC) have been used as feature selection methods. In the following, we discuss these two methods.

3.2.1 correlation-based feature selection (CFS)

Correlation-based Feature Selection is a method of selecting attributes using a filter technique. It examines the value of a subset of attributes by taking into account each feature's individual predictive ability as well as the degree of redundancy between them. It is a simple, fast, and scalable method that may be completed once and then used as input to several classifiers.

It can be coupled with forward selection, backward elimination, bi-directional search, best-first search, and genetic search to discover the optimal feature subset [20]. This technique is used with Weka's[21] BestFirst searching technique. As a result, a total of 119 out of 754 features were selected using the CFS technique. After that, this reduced feature subset is given as an input to the EPC swarm algorithm.

3.2.2 Emperor penguin colony (EPC)

Metaheuristic algorithms are an interesting and widely used field of research. As swarm intelligence algorithms are one of the population-based algorithms, which consist of a set of solutions

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that are randomly generated in a given search space, and solution values are updated during iterations until the best solution is generated [22].

Swarm algorithms use the behavior of real swarms such as birds, fish, and ants. Examples of swarm intelligence methods are Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), Bacterial Foraging, Cat Swarm Optimization, and Emperor Penguins Colony (EPC).

This proposed approach employs Emperor Penguins Colony (EPC) [7]. That algorithm is based on the behavior of emperor penguins in their colonies. It is a nature-inspired algorithm that is controlled by the body heat radiation of the penguins and the spiral-like movement of penguins from a cold domain to a warmer one.

The algorithm has a high potential for solving various and high-dimensional problems. The EPC swarm algorithm to find the best solution out of all possible solutions to an optimization problem [23]. This technique was used with Matlab. Depending on the proposed study, I used the 119 features that have been selected by the CFS feature selection method as input to the EPC swarm algorithm. There are six iterations of the swarm algorithm 20, 50, 100, 150, 200, and 500 to choose the best set of features. The output from this process was 60, 8, 16, 9, 9, and 12 features.

3.3 Ensemble Classifier

This part of the experiment is divided into two parts: prediction of classifier accuracy and ensemble classification.

In the first part, we used the following classifiers: Decision Tree, Naive Bayes, Random Forest, K-Nearest Neighbor, and Support Vector Machines in four sections. 1) With the dataset features directly, 2) with the dataset features after applying the CFS feature selection method, 3) with the selected features that have been chosen by the EPC swarm algorithm, and 4) with the hybrid approach selected features and then store the results in tables 1, 2, 3, 4, 5, and 6. In the second part, we have done ensemble classifiers and built them on top of these classifiers in an attempt to achieve an improvement in accuracy. After exploratory data analysis, notice that the ensemble classifier used only the top classifiers, which are: Random Forest, K-Nearest Neighbor, and Support Vector Machines. Those base classifiers as well as the ensemble method are briefly discussed below.

3.3.1 classification methods

(i) Decision tree

A Decision Tree is a tree-structured decision-making method. Which is made up of four parts: a root, leaf nodes, branches, and internal nodes. Any path begins from the root to the leaf nodes representing the different classes, the internal leaves reflect the processes, while the branches indicate the outcomes. A decision tree is a graphical representation for obtaining all suitable solutions to a decision depending on certain parameters [24]. It works with numerical and categorical data sets [4].

(ii) Naïve bayes

The Naive Bayes Classifier is a simplified probabilistic classification algorithm based on Bayes' theorem and the assumption of predictor independence. It is assumed that the presence of one feature in a class does not affect the presence of any other feature.

In the Naive Bayes classifier, create a conditional probability by analyzing the relationship between dependent and independent properties.

The probabilities are calculated by using the following equation:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

As in Eq 1: P(X) is the input probability, P(Y) is the probability of a possible exit status and P(Y|X) represents the probability of Y output versus input X [5].

(iii) Random forest

The Random Forest Algorithm is based on the concept of ensemble learning. This method uses many trees to predict the dataset's class, some decision trees may correctly predict the output while others may not. However, when all of the trees are combined, the correct outcome is predicted. As a result, a complex problem was solved and the model's performance was improved [5] [25]. Also on huge datasets, the Random Forest classifier performs well [6].

(iv) KNN

The K-Nearest Neighbor algorithm is based on the Supervised Learning technique and is one of the most basic Machine Learning algorithms. Where the distance between the new sample and all other classified samples must be calculated and analyzed using one of the different methods (Euclid, Manhattan, Minkowski, etc), and the samples must then be sorted according to their distances. An object's classification is determined by a majority vote of its neighbors, with the object being allocated to the most common class among its k closest neighbors [7].

(v) SVM

Support Vector Machine (SVM) is a machine learning algorithm that is supervised. It is usually used with linear and nonlinear data sets to solve binary or multi-class classification problems. After that, find the best hyperplane to separate the different classes by plotting each data item as a point in n-dimensional space, where n is the number of features. Consequently, to improve classification accuracy, SVM needs to increase the margin between classes [4, 5].

3.3.2 ensemble of classifiers

Ensemble methodology aims to combine multiple sets of classifiers to create a final prediction model as shown in Fig 2. Ensemble classifier is well-known for its ability to improve prediction performance and provide better accuracy [26].

Weighting methods and meta-learning are the two main methods. When the base-classifiers execute the same task and have similar success, weighting approaches are effective. Meta-learning approaches are best suited for situations where particular classifiers consistently classify or misclassify certain instances. Majority voting which has been used in this model is a weighing method in which an unlabeled instance is classified based on the class that gets the most number of votes [27].

The previous part of the experimental work has been done using Python programming language for the analysis and for building prediction models. For dataset representation and processing, used NumPy and pandas libraries and scikit-learn for building the machine learning models for predictions.

4. RESULTS AND DISCUSSION

4.1 Dataset and Experimental Setup

This study used the UCI Parkinson's disease dataset for classification[17] consisting of 756 voice measures of 252 individuals. Of the 252 individuals, 188 were Parkinson's patients, and the rest were healthy. Repeat the vowel /a/ three times to collect data, so 754 features were extracted for each voice measure. The details of 754 features are gender, baseline (21), time-frequency (11), Vocal Fold Features (22), Mel frequency cepstral coefficients (MFCCs) (84), wavelet transform-based features (182), tunable Q-factor wavelet transform (TQWT) (432), and class.

First, the CFS feature selection method chooses 119 features. EPC and CEPC methods apply in six iterations, starting from 20 iterations, 50, 100, 150, 200, and 500 iterations.

In the pre-processing phase, the dataset has been divided into 80% train and 20% testing to validate results. Also, SMOTE was used to deal with the class imbalance problem and used the standard scaler to transform the data in such a manner that it has mean as 0 and standard deviation as 1. The dataset was constructed from 252 individuals, including only 64 healthy individuals, and the rest were Parkinson's patients.

4.2 Experimental Evaluation

The performance of five classifiers used in this model, Decision tree, Naive Bayes, Random Forest, KNN, and SVM, were analyzed in accuracy as shown in Eq. 2.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \qquad (2)$$

Where TP is the number of true positives, TN true

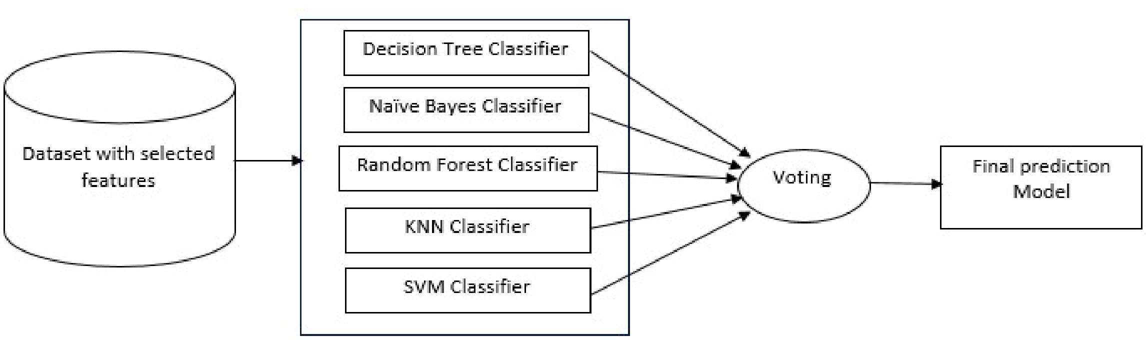

Figure. 2: Ensemble Classifier Model

Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

negatives, FP false positives, and FN false negatives [28].

4.2.1 classifier methods performance

The Parkinson dataset, when used without preprocessing as an input to the classifiers Decision Tree, Naive Bayes, Random Forest, K-Nearest Neighbor (KNN), and SVM (Support vector machine) exhibit the accuracies 79%, 65%, 82%, 66%, and 73%.

As shown in Fig 3, it can be observed that the Random Forest classifier gives the best response without applying any preprocessing methods while the Decision Tree is the second top performer. But Naive Bayes, KNN, and SVM are not performing well. But when we applied the CFS featureselection methodology on the Parkinson dataset in Fig 4, all the classifiers exhibit 79%, 80%, 80%, 83%, and 81%. So here some

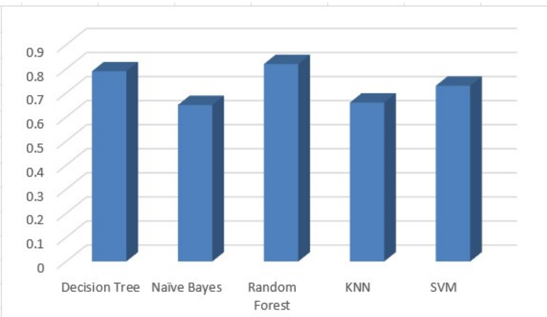

Figure. 3: Accuracy results for using the five classifiers with the 754 dataset features

classifiers have been improved, but others decreased. Also, for the six iterations in EPC swarm algorithm, the results arrived to 84% when applying the random forest with 100 iterations. But after applying the preprocessing methods on the proposed approach (CEPC), the results got improved.

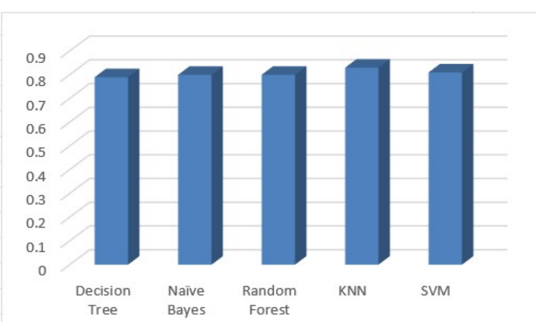

Figure. 4: Accuracy results for using the CFS feature selection with the 754 dataset features

Tables 1, 2, 3, 4, 5, and 6 illustrate the results of the experiments. The performance of the proposed CEPC approach is compared with other standard methods. All the following tables show the accuracy results for using all features without feature selection, after applying the CFS feature selection method, after applying the EPC swarm algorithm, and the CEPC proposed approach.

Table 1: Performance evaluation of using 20 iterations in EPC and CEPC

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.74	0.71
Naïve Bayes	0.65	0.8	0.71	0.66
Random Forest	0.82	0.8	0.81	**0.86**
KNN	0.66	0.83	0.77	**0.87**
SVM	0.73	0.81	0.75	**0.84**

The above table (Table 1) shows that Random Forest, KNN, and SVM in CEPC proposed approach improved the accuracy with 86%, 87%, and 84%. The same case is done with Table 2 and Table 3. Applying the proposed approach with 50 and 100 iterations provide an improvement with the same 3 classifiers.

Table 2: Performance evaluation of using 50 iterations in EPC and CEPC

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.74	0.71
Naïve Bayes	0.65	0.8	0.77	0.72
Random Forest	0.82	0.8	0.75	**0.85**
KNN	0.66	0.83	0.75	**0.87**
SVM	0.73	0.81	0.75	**0.81**

As shown in Table 4, the only enhancement was when apply the random forest classifier which has 86% accuracy. Tables 5, when using 200 EPC iterations, the random forest and SVM getting good accuracy results (85% and 82%). While Table 6 is showing the improvement in random forest, KNN, and SVM with 85%, 87%, and 81% accuracy results.

Table 3: Performance evaluation of using 100 iterations in EPC and CEPC

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.8	0.68
Naïve Bayes	0.65	0.8	0.78	0.73
Random Forest	0.82	0.8	0.83	**0.86**
KNN	0.66	0.83	0.77	0.81
SVM	0.73	0.81	0.75	0.8

Table 4: Performance evaluation of using 150 iterations in EPC and CEPC

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.78	0.66
Naïve Bayes	0.65	0.8	0.79	0.77
Random Forest	0.82	0.8	0.82	**0.85**
KNN	0.66	0.83	0.74	0.79
SVM	0.73	0.81	0.75	**0.82**

Table 5: Performance evaluation of using 200

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.73	0.73
Naïve Bayes	0.65	0.8	0.73	0.68
Random Forest	0.82	0.8	0.84	**0.87**
KNN	0.66	0.83	0.75	**0.88**
SVM	0.73	0.81	0.75	**0.83**

Table 6: Performance evaluation of using 500 iterations in EPC and CEPC

Classifier	All Features	CFS	EPC	CEPC
Decision Tree	0.79	0.79	0.72	0.7
Naïve Bayes	0.65	0.8	0.66	0.66
Random Forest	0.82	0.8	0.81	**0.85**
KNN	0.66	0.83	0.79	**0.87**
SVM	0.73	0.81	0.75	**0.81**

The comparative analysis of the accuracies achieved by various classifiers on six different EPC swarm iterations can be seen in the six previous tables, which always provide an improvement with the Random Forest classifier followed by SVM and KNN classifiers. In contrast, Decision tree and Naive Bayes classifiers show a decreased accuracy compared to the other standard methods.

4.2.2 ensemble method performance

Table 7, shows the ensemble classifier for our work. In this step only used the best three classifiers from the previous performance tables, Random Forest, KNN, and SVM. Noted that the best accuracy result (89.4%) has been obtained from using the EPC swarm algorithm with 50 iterations, followed by (89%) when use the EPC swarm algorithm with 20 iterations as referred to in figure 5.

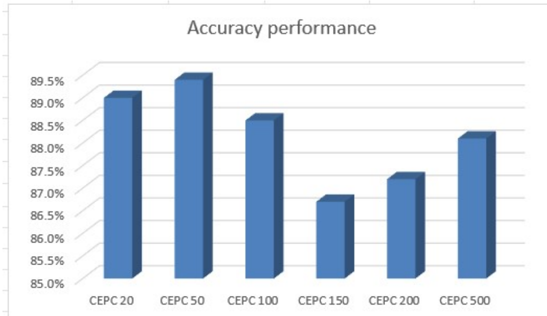

Figure. 5: Accuracy performance for the six different iterations using CEPC

5. CONCLUSION AND FUTURE DIRECTION

This study focuses on the Parkinson's disease diagnosis based on speech signals. In diagnosing Parkinson's disease, it is crucial to determine which of the extracted features from the speech signals may increase the classification performance. Therefore, applying the combination between the feature selection methods and the EPC swarm algorithm technique to find an optimal solution to help finding the best features which predict the Parkinson disease with less time and less cost.

The ensemble classifier methodology is also applied in this study. Five classification methods have been used to evaluate the performance. The proposed approach has outperformed in accuracy compared with the swarm algorithm in six iterations and the CFS feature selection algorithm, each on its own.

The experimental results prove that the proposed approach is effective compared to other features selection methods were tested on the same problem instances. Also prove that Random Forest, KNN, and SVM are the best classifiers with the

Table 7: Ensemble classifier results for the CEPC

	optimized method (CEPC)					
	20 iterations	50 iterations	100 iterations	150 iterations	200 iterations	500 iterations
Precision	0.96	0.94	0.96	0.9	0.94	0.95
Recall	0.82	0.84	0.81	0.82	0.8	0.8
F-Measure	0.88	0.89	0.88	0.86	0.86	0.87
Accuracy	**89.0%**	**89.4%**	**88.5%**	**86.7%**	**87.2%**	**88.1%**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

best performance. Meanwhile, the ensemble classifiers demonstrated that using the ensemble classifier with only three classifiers (Random Forest, K-Nearest Neighbor, and Support Vector Machines) with the hybrid approach (EPC with 50 iterations) provides an improvement in the accuracy results.

In the future, we can further use other disorders than speech disorders of Parkinson's disease as handwritten images to develop a decision-making approach for Parkinson's diagnosis. Also, intend to combine other swarm algorithms with other feature selection methods to enhance the performance and get best results.

REFERENCES:

[1] C. O. Sakar *et al.*, "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.*, vol. 74, 2019, pp. 255–263.

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

[3] R. Lamba, T. Gulati, and A. Jain, "A Hybrid Feature Selection Approach for Parkinson's Detection Based on Mutual Information Gain and Recursive Feature Elimination," *Arab. J. Sci. Eng.*, 2022.

[4] M. Chala Beyene, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques," no. January 2018, 2020.

[5] E. Avuçlu and A. Elen, "Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements," *Med. Biol. Eng. Comput.*, vol. 58, no. 11, 2020, pp. 2775–2788.

[6] P. Kaur, R. Kumar, and M. Kumar, "A healthcare monitoring system using random forest and internet of things (IoT)," *Multimed. Tools Appl.*, vol. 78, no. 14, 2019, pp. 19905–19916.

[7] A. Bourouhou, A. Jilbab, C. Nacir, and A. Hammouch, "Comparison of classification methods to detect the Parkinson disease," *Proc. 2016 Int. Conf. Electr. Inf. Technol. ICEIT 2016*, 2016, pp. 421–424.

[8] T. Saw and P. Hnin, "Swarm Intelligence Based Feature Selection for High Dimensional Classification : A Literature Survey," *Int. J. Comput.*, vol. 33, no. 1, 2019, pp. 69–83.

[9] H. G. Wahdan, H. E. Abdelslam, T. H. M. Abou-El-Enien, and S. S. Kassem, "Two-modified emperor penguins colony optimization algorithms," *Rev. d'Intelligence Artif.*, vol. 34, no. 2, 2020, pp. 151–160.

[10] S. Harifi, M. Khalilian, J. Mohammadzadeh, and S. Ebrahimnejad, "Emperor Penguins Colony: a new metaheuristic algorithm for optimization," in *Evolutionary Intelligence*, vol. 12, no. 2, 2019, pp. 211–226.

[11] O. Bchir, "Parkinson's disease classification using Gaussian mixture models with relevance feature weights on vocal feature sets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, 2020, pp. 413–419.

[12] C. D. Anisha and N. Arulanand, "Early Prediction of Parkinson's Disease (PD) Using Ensemble Classifiers," *2020 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 2020*, 2020, pp. 0–5.

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

[14] J. Goyal, P. Khandnor, and T. C. Aseri, "Analysis of parkinson's disease diagnosis using a combination of genetic algorithm and recursive feature elimination," *Proc. World Conf. Smart Trends Syst. Secur. Sustain. WS4 2020*, 2020, pp. 268–272.

[15] Z. Soumaya, B. Drissi Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of Parkinson disease using the genetic algorithm and SVM classifier," *Appl. Acoust.*, vol. 171, 2021, p. 107528.

[16] J. Goyal, P. Khandnor, and T. C. Aseri, "A Comparative Analysis of Machine Learning classifiers for Dysphonia-based classification of Parkinson's Disease," *Int. J. Data Sci. Anal.*, vol. 11, no. 1, 2021, pp. 69–83.

[17] "UCI Machine Learning Repository- Center for Machine Learning and Intelligent System.," 2018. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Artif. Intell. Res.*, vol. 30, no. 2, 2002, pp. 321–357.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[19] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Feature Subset Selection Problem using Wrapper Approach in Supervised Learning," *Int. J. Comput. Appl.*, vol. 1, no. 7, 2010, pp. 13–17.

[20] A. Gowda Karegowda, M. A. Jayaram, and A. . .Manjunath, "Feature Subset Selection using Cascaded GA and CFS: A Filter Approach in Supervised Learning," *Int. J. Comput. Appl.*, vol. 23, no. 2, 2011, pp. 1–10.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, 2009, pp. 10–18.

[22] S. M. Almufti, "Historical survey on metaheuristics algorithms," *Int. J. Sci. World*, vol. 7, no. 1, 2019, p. 1.

[23] S. Harifi, J. Mohammadzadeh, M. Khalilian, and S. Ebrahimnejad, "Hybrid-EPC: an Emperor Penguins Colony algorithm with crossover and mutation operators and its application in community detection," *Prog. Artif. Intell.*, vol. 10, no. 2, 2021, pp. 181–193.

[24] S. A. Mostafa, A. Mustapha, S. H. Khaleefah, M. S. Ahmad, and M. A. Mohammed, "Evaluating the performance of three classification methods in diagnosis of parkinson's disease," in *Advances in Intelligent Systems and Computing*, vol. 700, 2018, pp. 43–52.

[25] R. G. Ramani and G. Sivagami, "Parkinson Disease classification using data mining algorithms," *Int. J. Comput. Appl.*, vol. 32, no. 9, 2011, pp. 17–22.

[26] I. Gandhi and M. Pandey, "Hybrid Ensemble of classifiers using voting," *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015*, 2016, pp. 399–404.

[27] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, no. 1–2, 2010, pp. 1–39.

[28] F. M. Javed Mehedi Shamrat, M. Asaduzzaman, A. K. M. S. Rahman, R. T. H. Tusher, and Z. Tasnim, "A comparative analysis of parkinson disease prediction using machine learning approaches," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, 2019, pp. 2576–2580.