

A Novel Metaheuristic Data Mining Algorithm for the Detection and Classification of Parkinson Disease

P. Suganya^{1*} and C. P. Sumathi²

¹Bharathiar University, Coimbatore-641046, Tamil Nadu, India;
suganyadgvc@gmail.com

²Department of Computer Science, SDNB Vaishnav College for Women,
Chennai-600044, Tamil Nadu, India; drcpsumathi@gmail.com

Abstract

Objectives: Over the rapidly changing environment, the data mining techniques and its application are applied in the healthcare sector for medical diagnosis. The present study intends to provide a survey to gain the knowledge over current techniques of database discovery for the classification of Parkinson Disease. **Methods:** The study adopted a novel metaheuristic data mining algorithm for the detection and classification of Parkinson Disease were about 195 instances are selected for the investigation. In the initial phase the data underwent five phases, which includes training dataset, data pre-process, feature selection, classification and evaluation. However the research evaluated through performance measure tool, which consist of various techniques. This includes the confusion matrix, precision, recall and error rate. The confusion matrix is evaluated with various attributes like Specificity, Sensitivity, Accuracy and Positive and Negative predictive values. **Findings:** The study also performs a comparative study on five classification algorithms i.e. ABO, SCFW with KELM, FCM, ACO and PSO algorithms. This comparison results from confusion matrix of the selected algorithms which supports in identifying the specificity, sensitivity and accuracy of performance measures index showed that ABO algorithm is found to have best specificity, sensitivity and accuracy compared to all other algorithms, i.e. SCFW with KELM, FCM, PSO and ACO. In addition, the classifiers comparison results of the selected algorithms indicated that 'ABO' has the highest accuracy. **Conclusion:** In the present paper intended to estimate the efficiency and efficacy of the selected algorithm to best detect the Parkinson Dataset using various classifiers, as early detection of any kind of disease is an essential factor. The study reported that ABO algorithm has about have 97 percent accuracy in classifying and features filtering.

Keywords: Artificial Bear Optimization, Metaheuristic Algorithms, Parkinson Disease, Random Tree, Statistical Classifier

1. Introduction

It is well acknowledged that the causes of Parkinson's Diseases (PD) are still unknown due to the lack of blood or laboratory tests. Although various efforts have been adopted to classify and diagnose still achieving precise accuracy classification remains challenging¹⁻⁴. Therefore, given this context, the goal of this research is to design and develop the new algorithm⁵ to detect and classify PD. A new algorithm will be proposed considering the limitations of existing ANN and genetic models. In addition, the research also evaluates existing algorithms in terms of its benefits and drawbacks of each and how best to find the most suitable algorithm. This research

work adopts a novel metaheuristic data mining algorithm for the detection and classification of Parkinson Diseases. The present paper is organized in the below structure where Section 1 discusses the previous studies conducted elsewhere on PD and different algorithms used to detect the same. Section 2 discussed the adopted methodology and procedure which supports the research to obtain the objective. In section 3, the obtained data are analyzed and the results are presented. Finally, in section 4 the analyzed results are concluded. The below section illustrates the empirical evidence on PD.

This section illustrates the previous studies conducted on detection and classification of Parkinson Diseases based on various algorithms, Ene⁶ applied the

* Author for correspondence

Probabilistic Neural Networks (PNN) types for the classification purpose. Incremental Search, Monte Carlo Search and Hybrid search are the three PNN types that have been used for classification purpose, related to smoothening factor. As an outcome, there is no major difference between the three techniques of searching the smoothening factor although the Hybrid search plays better of 81%. The classification work on Parkinson Disease is carried out by several authors. Gil and Manuel⁷ examined the performance of classifier, which was designed for PD detection; this classifier includes SVM and ANN. With the use of three procedures, i.e. SVM and Multilayer Perceptrons (MLP) with of the two kernel types, they deduced a high level of exactness in the confusion matrix with regards to various measurement parameters like accuracy, specificity, sensitivity, negative predictive value and positive predictive value. They concluded by demonstrating a high level of certainty that was above 90%. Some of the parameter elements showed high preciseness of "Sensitivity" and "Negative predictive value" with 99.32% and 97.06%.

Another paper illustrated by Little et al.⁸ on compatibility of Dysphonic measurements predicted that nonstandard estimations along with traditional harmonics to noise ratios will be best suitable to separate active people from PWP. The four features Recurrence Period Density Entropy, Harmonics to Noise Ratio, Pitch Period Entropy and Detrended Fluctuation Analysis are inputted into the kernel support vector machine that produces a comprehensive classification performance of 91.4%. The task conducted by Das⁹ stressed on the comparison of four classification methods. By using SAS software, numerous classifiers have been implemented to determine Parkinson Disease. The four individual classification models used were Neural Network, DM Neural, Regression and Decision Tree are the four independent classification models used. The resulting Neural Network produced best classification rate of 92.9%.

Caglar et al.¹⁰ suggested two types of Multilayer Perceptrons (MLP), Radial Basis Function (RBF) Networks and Artificial Neural Networks. Apart from that, Adaptive Neuro-Fuzzy Classifier (ANFC) along with linguistic hedges is utilized. ANFC along with linguistic hedges provided the best acceptance results with 95.38% for training and test dataset at 94.72%. However, in the study by Tsanas et al.¹¹, the prediction level of PD was determined using the rapid and remote replication of UPDRS evaluation having clinically feasible accuracy

(5% prediction error). Thus the study adopted the speech test, which was administered by the researcher and self-administered, ensured that these speech features were analysed using the signal processing algorithms, and also curved map with the statistical edge over these algorithms to UPDRS. Their findings were assessed through the largest database of PD speech in prevalent, where number touching near to 6,000 recordings collected from 42 PD patients, who were selected for a six-month, multi-centre trial.

Rusz et al.¹² explored the feasible aspect of automated acoustic measures with an aim to recognize voice and speech disorders in PD. The researchers selected 46 Czech native speakers to collect data, out of which; 24 were with early PD, prior to getting administered pharmacotherapy treatment. They have implemented so many conventional and non-standard measurements along with the strategy of statistical decision-making to examine the extent to which selected speakers were having vocal impairment. Then, they resorted to application of support vector machine to identify the most suitable combination of measurements to distinguish PD from healthy subjects. Their proposed method ensured through classification performance of 85%. Again, they have identified the way measures of phonation and articulation are related with bradykinesia and rigidity in PD. As per their conclusion, the acoustic analysis might simplify the clinical assessment of voice and speech disorders, and assists in assessing clinical progression and for monitoring treatment effects.

Ramani and Sivagami¹³ focused to find the best accurate classifier algorithm, in order to diagnose PD. The research utilized Tanagra Data mining tool in order to propose a classification model to determine the patient as PD or non-PD. The research had 22 elements of dataset with an extensive range of values. The results confirmed that Random Tree classifier provides 100% accuracy. Yet, the *telemonitoring* dataset was not utilized, which is set to produce precise classification, which helps the drug discovery procedure.

Yadav et al.¹⁴ said that the PD symptoms showed up only during middle and late middle ages, which make it difficult for the researchers to predict it. There are many indications that prompted PD. This report concentrates on speech articulation and tried to determine a model on account of three data mining procedures. The three data mining methods are obtained from three different domains of data mining, i.e., from statistical classifier, tree classifier and support vector machine classifier. The output performances of the three classifiers are measured

with three performance metrics i.e., accuracy, specificity and sensitivity. Hence, the main aim of this report is to determine the best accurate model for people with PD. Still, only the voice process is treated as the symptom, and other symptoms such as environmental and age factor, difficulties with speech articulation and production and trembling arms, legs, hands were not considered.

In order to trespass the above constraint, another researcher Ramani et al.¹⁵ adopted *telemonitoring* in order to estimate the six feature relevance algorithms and thirteen classification algorithm's output. In addition, Sriram et al.¹⁶ used the patient voice to determine the existence of PD. The research embraced the machine learning method that was determined to give a better understanding from PD dataset in the current decade. The research results prompted that the parallel coordinates showed higher variability in Parkinson disease dataset. SVM has provided good accuracy (88.9%) when compared to majority of and k-NN algorithms. Classification algorithm like Random Forest has provided better accuracy (90.26) and Naïve Bayes has showed a lesser accuracy (69.23). Huge number of clusters in healthy dataset in Fo and low number in diseased data has been determined by Hierarchical clustering and SOM. In the same manner, the study conducted by Ahlrichs and Lawo¹⁷ explored the prediction for PD on the basis of four individual motor symptoms involving freezing of gait, bradykinesia, dyskinesia and tremor. The study implemented the machine learning algorithms and made a comparison with the conventional one to ascertain the efficiency. The study implemented the secondary data collected from past studies, and those were subjected to evaluation. Narmadha et al.¹⁸ in their study emphasized on determining the prevalence of the PD with the intention of ascertaining the features of sensitivity, specificity and accuracy in a given data set. The study implemented classifiers and rules, like Naive bayes, naïve bayes simple, NNge rule and decision table. It was identified from the analysis that the decision table was facilitating highest accuracy, i.e. 96 percent trailed with NNge rule (90%) naïve bayes simple (68%), and Naive bayes (67%). Still, there are so many useful tools such as Naïve Bayes Multinomial, Complement Naïve Bayes, Naïve Bayes Updatable, and Bayes Net, which ensures the estimation of the performance metrics, where features like Specificity, Sensitivity, and Accuracy were not implemented.

Sharma and Giri¹⁹ made an attempt to identify the PD through applying the artificial neural network and support vector machine in automatic means. Three

varieties of classifiers were implemented in the study like Multi-Layer Perceptron (MLP), k-Nearest Neighbors (KNN) algorithm, and Support Vector Machines (SVM) with an aim to assist the experts for making a diagnosis over PD. The research dataset comprises of so many voice signals obtained from 31 people (23 with people having PD and 8 healthier ones). Thus, this study relied on PD data to set obtained from University of California (UCI) machine learning database. The results reflect about the high percentage of accuracy, i.e. approximately the accuracy was observed to be 85 percent. There are various data mining Classification techniques are used to identify various diseases like diabetic²⁰. Thus it is clear evident that there are numerous studies, which have used disordered voice symptom to identify PD, but only few compared the algorithm. Hence, the present study compares various algorithms and identifies the best algorithm for the classification of PD.

2. Method

The data is collected from the healthcare sector through secondary methodology and the each and every healthcare sector huge dataset is available. Thus from the available data, about 195 data set is selected for the investigation. Thus metaheuristic algorithms technique is used to select population to attain the accurate classification methods. To overcome the limitations in the existing algorithms, the present study proposed a new algorithm, i.e. Artificial Bear Optimization (ABO). In simple terms, the ABO algorithm was farmed from the rigorous analytical bear smelling technique as the bear has a keen smelling sense, where a bear detects a dead animal within 20 miles of distance. This incredible sense of smell receptors allowed to formulation of ABO algorithm²¹. The ABO algorithm is designed and implemented to evaluate the performance through Area Under the Curve (AUC) in the ROC curve.

The methodology section is categorized into two sections, which include the data mining phase where the methods for the data using data mining is discussed, and in the second phase performance measure tools used to determine, the best algorithm for PD detection is discussed. Both data mining phase and performance measure section is considered as the key methodology which determined the structure of the present study. The data mining phase consists of various training stages, which allows classifying the PD severity. These stages consist various phases, which include the selection of training data set followed by pre-processing of the data

set, assessing the feature relevance and classification of algorithms, which allows to generate rules for severity prediction. The final phase consists of the evaluation, in which feature filtering algorithms create the classifiers and ranks based on their error rate. In order to ensure the chosen classifier's accuracy the available test phases are verified. Moreover, these attributes like Specificity, sensitivity, and accuracy are common for disease detection. For instance²² used these performance metrics to predict Heart Disease.

2.1 Training Dataset

The training dataset is used to collect the data and in the present study, these data are categorized into two target classes, which include the Healthy individuals and people with PD. However, among the target classes, the features of predictors are found to be common as only two possible values of target classes that classify the training dataset as 0 and 1. In the present study, the training dataset consist of 195 instances with 23 characteristic features, and a wide range of biomedical voice measurement is constituted as features.

2.2 Data Pre-Processing

During the data mining process the data pre-processing is considered as the important stage¹¹, as the redundant or irrelevant data is removed and more importantly, the occurrence of misleading is experienced only if the data is not analysed properly. During this stage the data is selected, normalized, cleaned, transformed and extracted.

2.3 Feature Selection

The feature selection is a process where the input variable's subset are identified and selected through feature's elimination and analysis with little or no predictive information. Moreover, during the feature selection phase, the researcher is more likely to be supported in pattern data mining communities, recognition and statistics^{11,23}. Thus, after the feature selection phase, the process of classification is carried out, which is illustrated in the below section.

2.4 Classification

According to Jacob and Ramani²⁴ the processes of classification consist of finding a set of models, which is distinguished and describe the available data classes. In order to identify the unknown label classification, this

classification process is carried out, which also allows implementing and predicting the model.

2.5 Evaluation

The classifier results obtained with the filtered attributes of the various feature reduction algorithms are analysed and evaluated based on their accuracy. The evaluation is performed using various performance measures.

2.6 Performance Measures

The present study adopted the Han and Kamber²⁵ method to measure the performance of the adopted model, which consist of various parameters' confusion matrix, Precision and Recall, Accuracy and Error-Rate.

2.6.1 Confusion Matrix

According to Nancy et al.²³ confusion matrix is the matrix consist of S*S matrix, in which various variables of 'S' column represent attributes of dataset index and 'S' rows. This tool allows analysing efficiency of the classifier from different classes. Table 1 represents a general structure of the confusion matrix of the predicted and actual class for a given dataset.

Table 1. Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)	False Negative(FN)
Negative	False Positive (FP)	True Negative (TN)

$$\text{sensitivity (\%)} = \frac{TP}{TP + FN} \times 100 \quad (1)$$

$$\text{specificity (\%)} = \frac{TN}{FP + TN} \times 100 \quad (2)$$

2.6.2 Precision and Recall

The fundamental element, i.e. Precision and Recall, allows to measure text performance's²⁶. The algorithm which classifies the data accordingly referred as precision and if the precision value is 1.000, it indicates 100 percent accuracy.

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

However, recall value is considered as the percentage of class relevant information which is classified correctly²⁶.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

2.6.3 Accuracy

According to Skodda et al.²⁷ based on the test cases percentage, i.e. classifying the test set correctly by the classifier is known as the accuracy.

$$\text{sensitivity (\%)} = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \quad (5)$$

Error-Rate

The misclassification rate of the test set is known as the error rate^{13,23}.

$$Re = 1 - \text{Acc (M)} \quad (6)$$

Where R is the Error rate, Acc (M) is the Accuracy of the selected variables

2.7 Test Phase

To verify the accuracy of the designed classifier, the provided test data sets are tested using the predictor attributes of each test data. Later, each of the attributes is being compared to the rules of generated classifier, which results in labelling the test data class. This is labelling affirms the accuracy.

2.8 Parkinson's Data Set

Parkinson Disease is detected based on various symptoms like trembling leg, arms, hands, arms and postural instability, bradykinesia, tremor, etc.,²⁸ out of which Voice measurement technique is the frequent and reliable symptoms which have shown 90 percent great progress in the advancement of PD detection¹³. Moreover, studies^{2,4,16,19} have shown that majority of the PD patients have some kind of vocal deterioration. Thus the present study mainly focused on discriminating the healthy individuals from those with PD, using the speech signals of the selected participants. The Table 2 represent the voice measure, and each row represents voice recording from the study participants, i.e. 195 instances in which the index status has two items, were 1 is for individuals with PD and 0 set is for healthy individuals.

3. Experimental Results

The present study aims to identify and understand the different classifiers which would encounter the performance for the data that was selected. The present

Table 2. Description of the Attributes

Index	Attribute	Description
MDVP	F0 (Hz)	Average vocal fundamental frequency
MDVP	Fhi (Hz)	Maximum vocal fundamental frequency
MDVP	Flo(Hz) -	Minimum vocal fundamental frequency
MDVP	Jitter (%)	Key Pentax MDVP jitter as percentage
MDVP	(Abs)	Key Pentax MDVP absolute jitter in microseconds
MDVP	RAP	Key Pentax MDVP Relative Amplitude Perturbation
MDVP	PPQ	Key Pentax MDVP five-point Period Perturbation Quotient
Jitter:	DDP	Average absolute difference of differences between cycles, divided by the average period
MDVP	Shimmer	Key Pentax MDVP local shimmer
MDVP	Shimmer (dB)	Key Pentax MDVP local shimmer in decibels
MDVP	APQ	Key Pentax MDVP eleven-point Amplitude Perturbation Quotient
Shimmer	APQ3	3 Point Amplitude Perturbation Quotient
Shimmer	APQ5	5 Point Amplitude Perturbation Quotient
Shimmer	DDA	Average absolute difference between consecutive differences between the amplitude of consecutive periods
NHR		Noise to Harmonic Ratio
HNR		Harmonics to Noise Ratio
RPDE		RPDE Recurrence Period Density Entropy
DFA		Detrended Fluctuation Analysis
Spread1		Non Linear measure of fundamental frequency
Spread2		Non Linear measure of fundamental frequency
D2		Correlation Dimension
PPE		Pitch Period Entropy
Status	Health Status	0- Healthy; 1- Parkinson

Note: MDVP stands for (Kay Pentax) Multi-Dimensional Voice Program.

study compared various algorithms like ABO, SCFW with KELM, FCM, PSO and ACO, which supports in detecting Parkinson's disease. The study adopted the metaheuristic algorithms technique which is used to select population to attain the accurate classification methods. The study conducted a comparative analysis of various feature selection algorithms, and it was found that ABO was appropriate and provided best attributes for classification. Every meta-heuristic algorithm is inspired by nature. In the implementation of the concerned new algorithm along with inspiration for nature, the proposed study will use the smelling sense of BEAR to the dynamic and optimal systems for diagnosing Parkinson diseases. The concerned question remains with designing the meta-heuristic problems for dynamic optimization, which remains as the information required during the concerned search that must get memorized and how the information gets used towards the guide in searching and maintaining the adaptability towards changes. Among several optimization problems, concerned environmental variables or the decision variables, in this case, diagnoses and classification gets subjected or perturbed that changes after the final solution gets implemented and obtained for the concerned problem. Thus, ABO is identified to diagnose and classify rightly, there is a necessity to consider the appropriate and acceptable solution with slight changes for the values of decision variable. Thus the present study used the following feature selection process:

```

Step 1. Select_subset = {}
Step 2. Init.all feature weightage = 0
Step 3. for I = 1 to no_of_sample
Step 3.1 Get one instance X from the training data set D.
Step 3.2 Get near hit H = instance in D where dist(X, H)
is closest and X.class=H.class
Step 3.3 Get nearmiss M = instance in D where dist(X, M)
is closest and X.class<>M.class
Step 3.4 Update weightage for all features
Step 4. for j = 1 to no_of_feature
Step 4.1 If weightagej>= Threshold, add featurej to
selected subset T

```

The study used the ABO algorithm for both regression and classification problems were the following steps were adopted:

```

Start{FF = {collection of all predictor features (forest)}
RF = {input data – feature vector}
Repeat {Compare the Attribute Values (av) of RF with FF.
If (RF.av == FF.av) then take the positive branch
Else take the negative branch
} for all RF until leaf node is reached.
End

```

The research used the MATLAB software to classify the dataset by using five classifiers and to determine the performance metrics, i.e. specificity, sensitivity and accuracy. Table 3 illustrates the performance measures

Table 3. Confusion matrix

Algorithm		Confusion Matrix		
ABO	0	0	1	Sum
	1	156	4	160
	Sum	4	156	160
	Sum	160	160	320
SCFW with KELM'	0	0	1	Sum
	1	83.5	5.5	89
	Sum	5.5	83.5	89
	Sum	89	89	178
FCM	0	0	1	Sum
	1	68.5	6.5	75
	Sum	6.5	68.5	75
	Sum	75	75	150
PSO	0	0	1	Sum
	1	204.5	30.5	235
	Sum	30.5	204.5	235
	Sum	235	235	470
ACO	0	0	1	Sum
	1	89	18	107
	Sum	18	89	107
	Sum	107	107	214

using the confusion matrix which consists of actual and predicted classifications information using the classification system. It indicates the confusion matrix of the selected algorithms which supports in identifying the specificity, sensitivity and accuracy and performance measures index. Moreover, they also identified to measure the test performance statistically. From the analysis, it is evident that ABO algorithm is found to have best specificity, sensitivity and accuracy compared to all other algorithms, i.e. SCFW with KELM, FCM, PSO and ACO.

Table 4 presents the Comparison results of classifiers of the selected algorithm. From the analysis, it is evident that algorithm 'ABO' has the highest accuracy of 97 percent, which is followed by SCFW with KELM (94%) and FCM (91%). However, the accuracy of other algorithms PSO and ACO were found be below 90 percent, i.e. 87 percent and 83 percent, and Figure 1 illustrate the same. Moreover, from Table 4 presents the recall and precision values of the comparison results of classifiers where the highest value is observed in ABO (97%), which is followed by while SCFW with KELM (94%) and FCM (91%). This also depicted in Figure 2, 3 and 4. In addition, while considering the F-measure value, the highest value is observed in ABO of 1.05. This followed by SCFW with KELM (0.9381), FCM (0.9134) and PSO (0.8713). However, the least value is observed in ACO (0.8313) which is presented graphically in Figure 4.

Table 4. Comparison results of classifiers of the selective attributes

Performance Measures	ABO	SCFW with KELM	FCM	PSO	ACO
Accuracy (%)	97.5000	93.8202	91.3333	87.0213	83.1776
Recall (%)	0.9762	0.9383	0.9134	0.8719	0.8310
Precision (%)	0.9734	0.9379	0.9133	0.8707	0.8317
F-Measure	1.0533	0.9381	0.9134	0.8713	0.8313

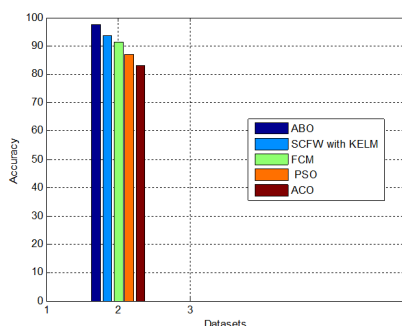


Figure 1. Accuracy values of the algorithms.

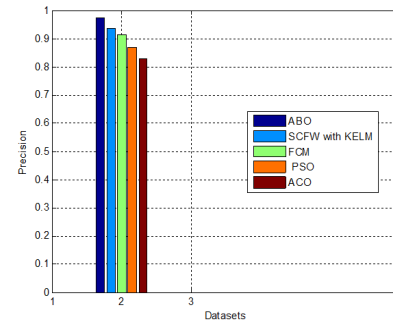


Figure 2. Precision values of the algorithms.

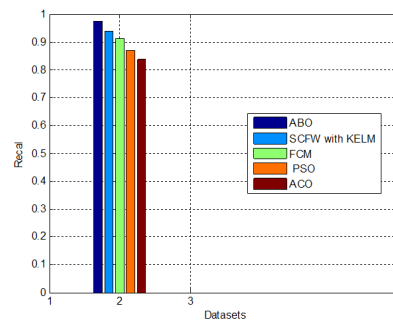


Figure 3. Recall values of the algorithms.

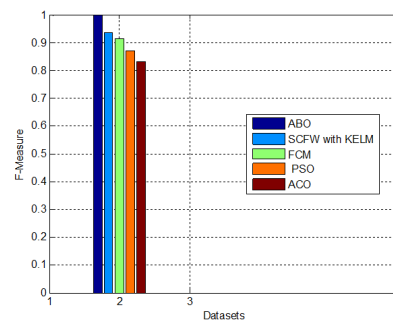


Figure 4. F-Measure values of the algorithms.

Thus, it is clearly evident from the study analysis, various algorithms were analysed to perform the experiment for both training and test dataset with the furnished feature relevance. The ABO Algorithm classifies the Parkinson Disease dataset accurately and provides the 97%. From the Linear Discriminant Analysis, it was evident that ABO, SCFW with KELM and FCM yields the accuracy results above 90 percent. Moreover, Figure 5 illustrates the comparative analysis of true positive rate and false positive rate. From the comparative analysis it is clearly evident that ABO algorithm was identified to have maximum mean value which was followed by SCFW with KELM, FCM, ACO and PSO algorithms.

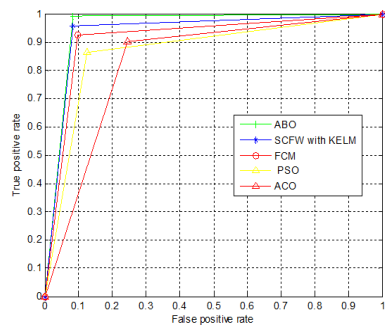


Figure 5. Comparative study of ABO with SCFW with KELM, FCM, PSO and ACO algorithms.

Table 5. Error rate for selective algorithm

Performance Measures	ABO	SCFW with KELM'	FCM	PSO	ACO
Error Rate	0.025	0.062	0.087	0.130	0.168

Table 5 presented the calculated error values of selective algorithm. From the evaluation it is evident that algorithm ABO has the minimum error value i.e. 0.025, which was followed by SCFW with KELM (0.062) and FCM (0.087). Among all, the PSO and ACO algorithm have produced an error rate of 0.13 and 0.168. Thus from the analysis it is evident that the ABO algorithm builds the classification rule based on this characteristic features to obtain the zero error rate and it is graphically illustrated in Figure 6.

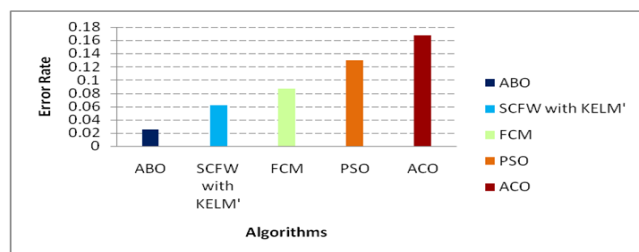


Figure 6. Comparative Error rate for ABO, SCFW with KELM, FCM, PSO and ACO algorithms.

4. Conclusion

In this research work, it is intended to estimate the efficiency and efficacy of the selected algorithm to best detect the Parkinson Dataset using various classifiers, as early detection of any kind of disease is an essential factor.

The present study used 195 instances and 23 Parkinson Dataset features were the study performed a comparative study of several algorithms i.e. ABO, SCFW with KELM, FCM, PSO and ACO on the dataset. In order to overcome the existing algorithm's limitation the present study proposes that the ABO algorithm as the best based on the specificity, sensitivity and accuracy, positive and negative predictive values. Thus, from the study it is concluded that ABO algorithm is identified as the best algorithm which supports in treating the patient well ahead. In this research paper, ABO classifier yields the 97% accuracy.

5. References

1. Babu GS, Suresh S. Parkinson's disease prediction using gene expression – A projection based learning meta-cognitive neural classifier approach. *Expert Syst Appl.* 2013; 40(5):1519–29. doi:10.1016/j.eswa.2012.08.070.
2. Rustempasic I, Can M. Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition. *SOUTHEAST Eur J SOFT Comput.* 2013; 2(1):42–9. Available from: <http://scjournal.com.ba/index.php/scjournal/article/viewFile/43/40>
3. Abinaya S, Sivakumar R, Karnan M, Shankar DM, Karthikeyan M. Detection of Breast Cancer In Mammograms - A Survey. *Int J Comput Appl Eng Technol.* 2014; 3(2):172–8.
4. Shahbakhti M, Far DT, Tahami E. Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. *J Biomed Sci Eng.* 2014; 7(4):147–56. doi:10.4236/jbise.2014.74019.
5. Mozaffari A, Fathi A. A natural-inspired optimization machine based on the annual migration of salmons in nature. 2013; Available from: <http://arxiv.org/ftp/arxiv/papers/1312/1312.4078.pdf>. Accessed 2015 Apr 2
6. Ene M. Neural network-based approach to discriminate healthy people from those with Parkinson's disease. *Ann Univ Craiova, Math Comp Sci Ser.* 2008; 35:112–6.
7. Gil D, Manuel D. Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines. *Global Journal of Computer Science and Technology.* 2009. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.7856&rep=rep1&type=pdf>
8. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomed Eng Online.* 2007; 6:23. doi:10.1186/1475-925X-6-23.
9. Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst Appl.* 2010; 37(2):1568–72. doi:10.1016/j.eswa.2009.06.040.
10. Caglar MF, Cetisli B, Toprak IB. Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using

- ANN and Adaptive Neuro-Fuzzy Classifier. *Mühendislik Bilim ve Tasarım Derg.* 2010; 1(2):59–64. Available from: <http://sdu.edu.tr/edergi/index.php/mbtd/article/view-File/1709/1887>.
11. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng.* 2010; 57(4):884–93. doi:10.1109/TBME.2009.2036000.
 12. Ruzs J, Cmejla R, Ruzickova H, Ruzicka E. Objectification of dysarthria in Parkinson's disease using Bayes theorem. *Recent Res Commun Autom Signal Process Nanotechnology.* Astron Nucl Phys. 2011.
 13. Ramani GR, Sivagami G. Parkinson Disease Classification using Data Mining Algorithms. *Int J Computer Appl.* 2011; 32(9):17–22.
 14. Yadav G, Kumar Y, Sahoo G. Predication of Parkinson's disease using data mining methods: a comparative analysis of tree, statistical, and support vector machine classifiers. *Indian J Med Sci.* 2011; 65(6):231–42. doi:10.4103/0019-5359.107023.
 15. Ramani RG, Sivagami G, Jacob SG. Feature Relevance Analysis and Classification of Parkinson Disease Tele - Monitoring Data Through Data Mining Techniques. *International J Adv Res Comput Sci Softw Eng.* 2012; 2(3). doi:10.5120/3920-5521.
 16. Sriram TV, Rao MV, Narayana GS, Kaladhar DSVGK, Vital TPR. Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms. *Int J Eng Innov Technol.* 2013; 3(3):212–5.
 17. Ahlrichs C, Lawo M. Parkinson's Disease Motor Symptoms in Machine Learning: A Review. 2013. Available from: <http://arxiv.org/abs/1312.3825>
 18. Narmadha D, Marudhadevi D, Santhi B. Parkinson's Disease Detection Using Soft Computing Techniques. *World Appl Sci J.* 2014; 29:89–92. doi:10.5829/idosi.wasj.2014.29.dmsct.16.
 19. Sharma A, Giri RN. Automatic Recognition of Parkinson's disease via Artificial Neural Network and Support Vector Machine. *Int J Innov Technol Explor Eng.* 2014; 4(3):35–41.
 20. Ananthapadmanabhan KR, Parthiban G. Diabetic Retinopathy using Data Mining Classification Techniques. *Indian J Sci Technol.* 2014; 7(10):1498–503.
 21. Prem MJ, Karnan M. Business Intelligence: Artificial Bear Optimization Approach. *Int J Sci Eng Res.* 2013; 4(8):2136–40.
 22. Santhanam T, Ephzibah EP. Heart Disease Prediction Using Hybrid Genetic Fuzzy Model. *Indian J Sci Technol.* 2015; 8(9):797–803.
 23. Nancy P, Ramani RG, Jacob SG. Discovery of Gender Classification Rules for Social Network Data using Data Mining Algorithms. *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCCIC'2011); Kanyakumari, India.* 2011.
 24. Jacob SG, Ramani G. Discovery of Knowledge Patterns in Clinical Data through Data Mining Algorithms : Multi-class Categorization of Breast Tissue Data. *Int J Comput Appl.* 2011; 32(7):46–53.
 25. Han J, Kamber M. *Data Mining: Concepts and Techniques.* 2nd ed. San Francisco, CA: Elsevier; 2000.
 26. Jacob SG, Ramani RG, PN. Efficient Classifier for Classification of Hepatitis C Virus Clinical Data through Data Mining Algorithms and Techniques. *Proceedings of the International Conference on Computer Applications; Pondicherry.* 2012. doi:10.73445/ISBN_0768.
 27. Skodda S, Rinsche H, Schlegel U. Progression of dysprosody in Parkinson's disease over time - a longitudinal study. *Mov Disord.* 2009; 24(5):716–22. doi:10.1002/mds.22430.
 28. Parkinson's Disease Foundation. What is Parkinson's Disease? 2011. Available from: http://www.pdf.org/en/about_pd. Accessed 2015 May 6.