Prediction of Parkinson's Disease using Hybrid Feature Selection based Techniques

Conference Paper · December 2021

DOI: 10.1109/ICCCT53315.2021.9711843

1 author:

Avijit K. Dash
Indian Institute of Technology Bombay

Prediction of Parkinson's Disease using Hybrid Feature Selection based Techniques

Avijit Kumar Dash
Dept. of Computer Science and Engineering
Indian Institute Of Technology Bombay
Powai, Mumbai, Maharashtra, India
avijitdash067@gmail.com

***Abstract*—Parkinson's disease (PD) is one of the significant severe problems globally in recent times. It is a neurological disorder that progresses over time and the most severe problems after Alzheimer's disease. Our article proposes a Hybrid Feature Selection system for the initial detection of PD from speech recordings. This method picks the best set of instances that can lessen instance vector dimensions from 22 to 5. We have proposed a machine learning-based model using five different classifiers named Random Forest, Logistic Regression, XGBoost, AdaBoost, and Gradient Tree Boosting. Gradient Tree Boosting presents the best appearance with a spectacular accuracy of 98.31% and the area under the ROC curve 98.66%, among all classifiers used to predict PD. We showed that the stated design has greater accuracy than the current methods available in the literature, and the number of instances is less than others.**

***Index Terms*—Parkinson's disease, Hybrid Feature Selection, Prediction, Gradient Tree Boosting.**

I. Introduction

Bioinformatics [1] is an interdisciplinary field of science. Here develops various software tools and techniques to gain knowledge about biological data. Bioinformatics is not a single discipline; it combines multiple fields like computer science, biology, information technology, statistics, and mathematics to explore and illustrate biological data. Nowadays, bioinformatics is an emerging research field of science. It has various applications such as Biomedicine, Microbiology, Agriculture, Disease diagnosis. Disease diagnosis is one of the most vital bioinformatics practices like Parkinson's disease prediction, Alzheimer's disease detection, Cancer Prediction. Our research focuses on predicting Parkinson's disease, so that we will discuss Parkinson's disease here. Parkinson's disease (PD) is a well-known disease after Alzheimer's, and it is a neuropsychiatric dysfunction [2]. As PD is a degenerative disease and progresses over time, it mainly affects the central nervous system, which is the motor system. It has numerous introductory signs involve shaking, rigidity, bradykinesia, and postural weakness [3]. PD hits people above the age of 60, and it progresses over time. More than 10 million live with Parkinson's disease, and the number is increasing day by day. Study says that Many people in Bangladesh are suffering from PD. According to WHO, Parkinson's disease Deaths in Bangladesh prolonged to 539 or 0.07% of total deaths in 2017 [4]. The age-adjusted death rate is increasing day by day. Bangladesh ranks 152nd with PD-affected people in the world.

The main reason for PD is still unknown, but many believe that climate and ancestral circumstances perform an essential role [5]. The symptoms of PD are of two types: non-motor and motor symptoms. It is a persistent degenerative disease that influences the motor system. Currently, there exist no cures for PD. Nevertheless, there are a diversity of remedies that can provide exciting relief from the traits. As there are no antidotes for PD, it is inevitable to detect the disorder in the beginning. Based on the severity of PD, we have proposed a machine-learning-based system to detect PD.

Various studies [7, 8, 9, 10, 11, 12, 13, and 14] were conducted in the literature to check PD in the early stages. This paper focused on a machine learning-based method to detect PD using feature selection techniques. Our experiment uses the feature extraction technique to optimize features and then apply those features to predict PD using machine learning techniques. Using our proposed system, we have gained greater accuracy with the least feature space than existing methods.

The writing is constructed as follows: related works provides in Section 2. Section 3 includes a description of the UCI database [6] and our proposed approach. Section 4 presents the empirical results and a comparison of results with existing techniques. Finally, in Section 5 conclusion is provided.

II. Related Works

Many researchers classified Parkinson's disease through various methods. Some recent studies have proposed new feature selection methods to reduce the feature vector size used for classification purposes. Marziye Keshavarz Shahsavari *et al.* [7] established the extreme learning machine and hybrid particle swarm optimization to detect Parkinson's disease. They optimized at most 12 features from 22 using Hybrid Particle Swarm Optimization and Extreme Learning Machine and achieved 88.72% accuracy using their proposed method. Akshaya Dinesh *et al.* [8] proposed

2021 4th International Conference on Computing and Communications Technologies (ICCCT) | 978-1-6654-1447-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCT53315.2021.9711843

an ML approach to detect Parkinson's disease from sound recordings. They used various learning methods and filter-based instance selection schemes. They selected the top 10 features from 22 and achieved 91.21% accuracy by using Boosted Decision Tree. Contrastive research of numerous nature-inspired algorithms designed by P.Shrivastava *et al.* [9] to estimates the performance of various evolutionary methods by determining optimal characteristics for diagnosing PD. Najmeh Fayyazifar *et al.* [10] suggested a new design to identify Parkinson's disease. They adopted a Genetic Algorithm to optimize features and Ensemble Techniques to validate the performance. They selected at most six features and achieved 96.55% accuracy, but the best result was by selecting seven features and achieving 98.28% accuracy. Md. Inzamam-Ul-Hossain *et al.* [11] employed an ensemble method using the PASW benchmark to recognize people living with Parkinson's disease from healthful people. They used three classifiers named C&R Tree, Bayes Net, and C5.0 to form the ensemble arrangement. Their arrangement presented fewer mis-sorted occurrences than single classifiers used to construct the model. They ended up gaining 96.875% accuracy by using 22 features. Sachin Shetty *et al.* [12] used SVM Based ML approach to recognize Parkinson's Disease. At first, they picked up feature vectors using a host of statistical tools and selected the best seven features. Finally, they applied SVM to measure the performance. They obtain good overall accuracy of 83.33% with seven selected instances and SVM.

III. Feature Selection

Variable extraction techniques are dedicated to lessening the capacity of the characteristics term to point out only the variables that influence an event supporting a state [5]. Investigating multiple pieces of instance is not usually the common proper way to decide between a harmful and a non-harmful wound. Possessing multiple instances may be wholly superfluous, and it is more practical to analyze with a picked amount of instances. Hence, instance choice is crucial to be able to disregard unnecessary instances. Feature extraction algorithms can be broadly categorized as filters, wrappers, and embedded. In this experiment sequence of popular filter and wrapper methods is used to optimize the instance set. For that purpose, five techniques were applied from the filter and wrapper method. Those are low variance and correlation-based techniques from the filter method to reduce the feature set, and for further optimization, three techniques are applied from the wrapper method: recursive feature elimination, step backward selection, or backward search and exhaustive search.

A. Low Variance Feature Selection

The low variance method [13] is a feature selection approach that keeps features with a variance higher than the given threshold. In this proposed system, considered 0.0001 thresholds to select relevant features by the low variance method.

B. Correlation-based Feature Selection (CFS)

The correlation-based Feature Selection (CFS) procedure can choose the appropriate instance from a set of instances. CFS estimates subsets of instances based on the hypothesis; highly correlated instances are considered good subsets of instances, although they are not correlated to any additional [14]. The scheme to obtain a subgroup of data significantly associated with the classification is separated into two sections. The first portion is to assess a metric used to specify a connection between instances and analysis decisions. The metric can estimate employing equation (1).

$$P_Z = l*\overline{w_{\mathrm{mn}}}/\sqrt{l+l(l-1)\overline{w_{\mathrm{nn}}}} \tag{1}$$

Here, P_Z is a metric of subgroup, which has l proportions. $\overline{w_{mn}}$ is an average estimation of the connection between instances and conditions (n subgroup of Z). $\overline{w_{nn}}$ is an average amount of the connection between instances and other instances.

In our recommended approach, 0.95 is employed as a threshold to get the highly correlated instances by the CFS method.

C. Recursive Feature Elimination (RFE)

Recursive feature elimination (RFE) [15] is an instance choosing purpose that fits into a model and eliminates the most vulnerable instance (or instances) until it picks the particularized number of instances. The model's instances were classified using coef_ or feature_importances_ attributes and recursively removed a small number of instances per cycle. Ultimately, that scheme is recursively iterated on the pruned set until the sought number of instances is chosen. RFE tries to exclude territories and co-linearity that may subsist in the paradigm.

D. Step Backward Selection or Backward Search

Step backward selection [16] is an iterative method in which we start with having all instances in the model. In the beginning, step-backward instance selection removes one feature round-robin fashion from the instance set, and the classifier's performance is estimated. The instance set that yields the best performance is held. Next phase, another instance was removed in a round-robin fashion and estimated the performance of all the combinations of features without the two instances. This process proceeds until the defined number of instances settle in the dataset.

E. Exhaustive Search

The exhaustive instance determination algorithm [17] is a wrapper approach for brute-force evaluation of instance subsets; the best subset determines by optimizing a particularized achievement metric assigned an arbitrary regressor or classifier. The exhaustive search starts with examining the best one-component instance subset from the input instance, the same in the forward selection algorithm; then, it decides the best two-component instance subset, which may consist of any pairs of the input instances. Afterward, it proceeds to find the best triple out of all the sequences of any three input features. Finally, it picks the best-optimized instances.

F. Hybrid Feature Selection (HFS)

Figure 1 shows the hybrid feature selection technique that we have used in our proposed method by combining filter and wrapper models. Two filter methods choose to eliminate the usual repetitive or unrelated instances. At first, utilize the Low variance and Correlation-based techniques; These pairs resulted in instance sets are combined as the pre-processed feature set for fine-tuning. This step is denominated in the aggregate model. Li-Yeh Chuang et al. [18] stated that unwanted instances could be removed by applying traditional Recursive Feature Elimination, Step Backward Search, and the Exhaustive Search methods for the wrapper model. However, after the first filter model determination is implemented, it should be applied to achieve final features for classification more efficiently. These wrapper systems frequently operate until arriving at the stopping criterion. For the ending criterion, when the test result begins to get poorer, the procedure ends. This aggregation of filter and wrapper methods is called hybrid filter-wrapper feature selection methods or hybrid feature selection methods.

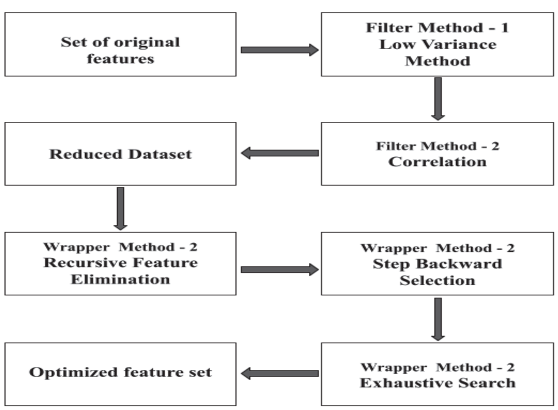

Fig. 1. Flow chart of our proposed method.

IV. Methodology

A. Proposed Method

The proposed method object is to achieve a minimal number of instances from many instances from a dataset containing relevant instances using the popular filter and wrapper methods. At first, a combination of techniques from filter methods is applied to reduce the number of instances set then a combination of techniques from wrapper methods is applied to reach an optimal subset of instances. In this writing, we suggested a new method to lessen the instance set by combining filter and wrapper methods. Filter approaches are independent of the learning of the machine and also faster in comparison with wrapper approaches. So, to make this procedure further useful from a computational point of view, we use wrapper methods followed by filter methods; this proposed technique is a hybrid instance selection technique. Figure 1 illustrated the diagram of the introduced technique used to extract the optimal number of instances. We apply several instance selection techniques to the primary dataset to select the optimized instance set for our stated approach. We develop a hybrid instance selection system by combining filter and wrapper methods to optimize the instance set. We use five techniques (two techniques from the filter and three techniques from the wrapper). The primary dataset fits inside a low variance method to remove some of the original instance set at the initial process. A threshold is fixed to conduct our experiment, and the threshold is 0.0001.

In this process, some of the instances (4 instances) were removed. After completing the low variance method, we fit the data into a correlation-based system to reduce the instance set again, and here we also fixed a threshold of 0.95. Here some of the instance(6 instances) is removed. The training ratio and the testing dataset we used in our proposed method are 70% and 30%, respectively. Figure 2 represents the correlation heatmap.

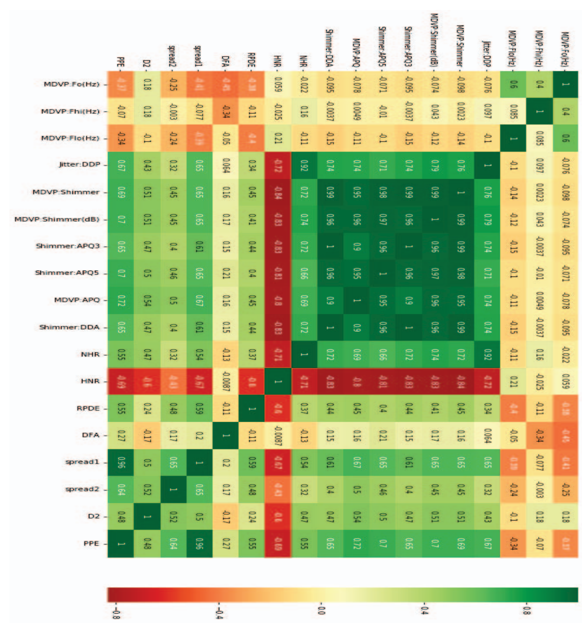

Fig. 2. Correlation heatmap after low variance method.

Here, the filter method provides reduced features. The reduced feature set is fit into the first wrapper method, which is recursive feature elimination (RFE). In RFE, we removed some features and kept the rest (9 features);

after completing RFE, a step backward selection method is applied to reduce the feature set further.

Step backward method work like a round-robin fashion. It removes one feature at the first iteration and continues until the specified features remain in the dataset. We keep seven features and fit them into the following method to get the optimized feature set. Finally, the exhaustive search is applied to get the optimized feature set. Exhaustive search is like a brute force feature selection system. At first, it combines the possible subset of all the features and then keeps the best possible subset. By this process, we get five features, and that is the optimized feature set. Figure 3 represents the optimized feature set. The horizontal axis of Figure 3 represents the features name, and the vertical axis represents the weighted value.

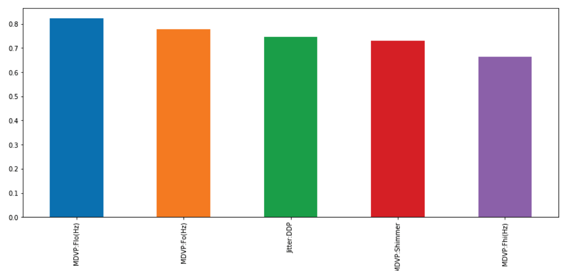

Fig. 3. Optimize feature set of PD.

Several types of classifiers are used to test our model. The classifiers employed in this analysis are Random Forest, Logistic Regression, AdaBoost, XGBoost, and Gradient Tree Boosting. The primary purpose of using these classifiers is to measure the performance of our proposed method using a hybrid feature selection technique.

Figure 4 illustrated our introduce system. At first, we select the optimal feature using HFS (Figure 1) technique. After selecting the optimal feature, we use five classifiers for performance measuring purposes. Then we analyze the result to see which classifier is best for our experiment, and we saw that Gradient Tree Boosting is the best classifier for our experiment. Finally, we compare our results with existing methods.

We used python as a programming language and Jupiter notebook as an environment to implement our proposed method. A multiple machine learning library is also used. The system where the program executes has a core i3 processor, 4 GB RAM, and 1TB of the hard disk.

V. Experimental Results

In this paper, Parkinson's disease dataset is used that is collected from the UCI Machine learning repository [6]. The collected dataset was in dot data (.data) format. For our experimental purpose, the dataset is converted into CSV format. This dataset contains a range of biomedical voice frequencies of 31 people; among them, 23 with PD, and

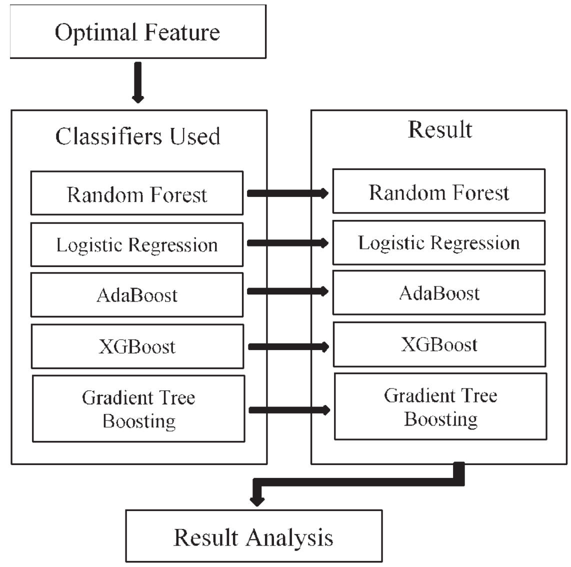

Fig. 4. Flow chart of our proposed method.

eight are healthful. Each vertical line in the table is a distinct voice example, and each horizontal line resembles one of 195 voice recordings from certain people. The foremost aim of the data is to distinguish wholesome people from those with PD. The reporting is performed at a sampling frequency of 44.1 kHz with a 16-bit resolution. The dataset is partitioned into two categories depending on its "status" column, set to 0 for wholesome subjects and 1 for PD. Figure 5 represents the speech signal of healthy and Parkinson's persons.

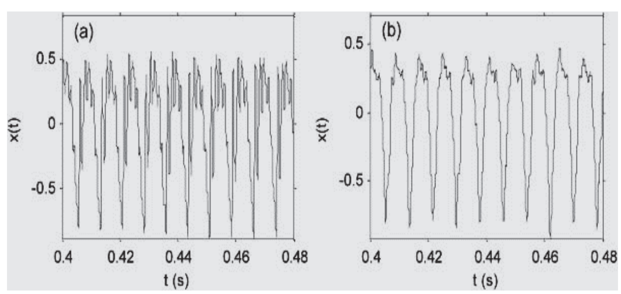

Fig. 5. Two chosen samples of speech signals: (a) wholesome, (b) case with PD. The horizontal axis is time in seconds; the vertical axis is signal amplitude (no units).

In this introduced system, the principal objective is to lessen the feature set to generate better performance; for that goal, the hybrid feature selection procedure is implemented. Eventually, we decrease the feature set from 22 to 5. Table I shows the brief description of selected features by applying the hybrid feature selection technique.

For measuring performance, Receiver Operating

TABLE I
SELECTED FEATURES OF THE PROPOSED METHOD

Attribute	Description
MDVP: Flo(Hz)	least oral elementary density
MDVP: Fo(Hz)	Mean vocal elementary density
Jitter: DDP	The estimate of variety in elementary density
DVP: Shimmer	The estimate of deviation in amplitude
MDVP: Fhi(Hz)	Topmost oral elementary density

Characteristics (ROC), accuracy, and F1 Measure are used.

The area under the ROC curve (ROC-AUC) is expected to rank an arbitrary determined positive instance higher than a randomly chosen negative instance. Accuracy is measured by equation (2):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

The usual F-measure is the symphonious average of precision and recall that is calculated using equation (3):

$$F = \frac{(recall^{-1}+precision^{-1})^{-1}}{2} = 2\frac{recall.precision}{recall+precision} \tag{3}$$

Here precision (P) is a standard of result relevancy, while recall (R) is a test of how many genuinely appropriate results are delivered that are calculated using equations (4) and (5), individually.

Here precision (P) is a test of decision relevancy, while recall (R) is a test of how many genuinely relevant results are delivered that are estimated using equations (4) and (5), respectively.

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP+TN}{TP+FN} \tag{5}$$

Here, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Table II shows the obtained results of five classifiers using optimal features obtained from the hybrid feature selection technique. We can observe that gradient tree boosting achieves the best accuracy of 98.31% out of the five classifiers that we have used.

Figure 6 represents the gradient boosting ROC curve. From that curve, we can see that the performance improves continuously and is steady for some time. After that, it started progressing over and swiftly reaches steady-state, and that is the best ROC-AUC result we received. In Table III, we analyze our stated approach with earlier developed systems in the literature. Table III shows that our stated approach can pick the most relevant and fewer instances than earlier investigations, significantly impacting classification accuracy.

Figure 7 presents the outcome comparison of our stated approach with earlier approaches using the histogram.

TABLE II
EXPERIMENTAL RESULTS USING VARIOUS CLASSIFIRES AND SELECTED FEATURES

Classifiers	Num. of Selected Fields	Accuracy (%)	ROC	F1-Measure
Random Forest	5	94.92	0.9883	0.967
Logistic Regression	5	88.14	0.9181	0.9247
AdaBoost	5	94.92	0.9933	0.967
XGBoost	5	93.22	0.99	0.9804
Gradient Tree Boosting	5	98.31	0.9866	0.9892

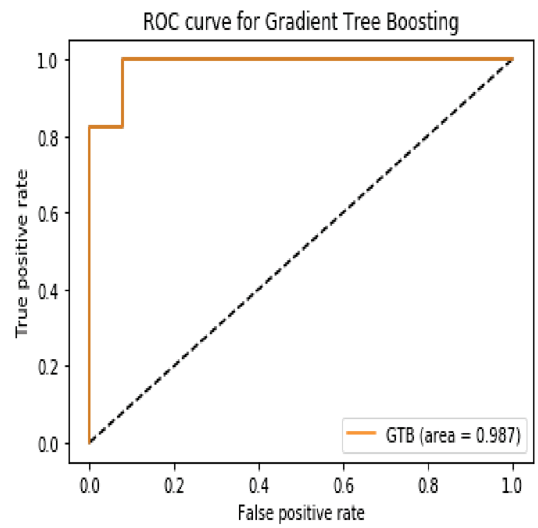

Fig. 6. Gradient Tree Boosting ROC curve.

TABLE III
EXPERIMENTAL RESULTS USING VARIOUS CLASSIFIRES AND SELECTED FEATURES

Method	Num. of Selected Features	Accuracy (%)
, Hybrid PSO-FS [7]	12	88.72
ML (Boosted Decision Tree) [8]	10	91.21
Neural network + Binary Bat algorithm	6	93.60
Genetic Algorithm + AdaBoost [10]	6	96.55
Genetic Algorithm + Bagging [10]	7	98.28
Proposed Method (HFS + Gradient Tree Boosting)	**5**	**98.31**

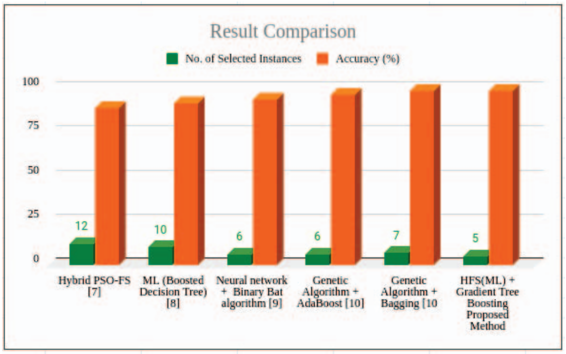

Fig. 7. Result comparison histogram.

VI. Conclusion

An effective approach has been proposed to generate an accurate predictive model for predicting Parkinson's disease using Machine Learning-based Hybrid Feature Selection technique. Here the aggregation of filter and wrapper systems is used to obtain a hybrid system. For monitoring the model performance, we use multiple classifiers. The prediction models are formed using Random Forest, Logistic regression, XGBoost, AdaBoost, and Gradient Tree Boosting. It is noted that Gradient Tree Boosting acheives the best performance for our design, which is 98.31% accuracy for the testing dataset with five classifiers. Compared with other methods, the proposed method uses less instances for classification and achieves greater performance.

References

[1] Encyclopedia Britannica. Bioinformatics | science. [online] Available at: https://www.britannica.com/science/bioinformatics [Accessed 15 Jul. 2019].

[2] C. M. Bronikowski, Angela Weng, Jacob D. Furst, Daniela S. Raicu, "Prediction of chronic fatigue syndrome using decision tree-based ensemble methods," 2011 International Conference on Artificial Intelligence, pp. 49-54, July 18-21,2011.

[3] Vu TC, Nutt JG, Holford NH, "Progression of Motor and NonMotor Features of Parkinson's Disease and Their Response to Treatment". British journal of clinical pharmacology, 74(2):267–83, 2012.

[4] Parkinson's Disease in Bangladesh. [online] Available at: https://www.worldlifeexpectancy.com/bangladesh-parkinson-disease [Accessed 15 Jul. 2019].

[5] S. Catani, M. Vannucci, M. Vannocci, and V. Colla, "Variable Selection and Feature Extraction through Artificial Intelligence Techniques", Multivariate Analysis in Management, Engineering and the Science, chapter 6, pp.103-118, 2012.

[6] "UCI Machine Learning Repository: Data Set", https://archive.ics.uci.edu/ml/datasets/Parknsons, [Accessed: 05- March- 2018].

[7] Marziye Keshavarz Shahsavari, Hasan Rashidi, Hajar Rajabi Bakhsh, "Efficient Classification of Parkinson's Disease Using Extreme Learning Machine and Hybrid Particle Swarm Optimization", 4th International Conference on Control, Instrumentation, and Automation (ICCIA), 2016.

[8] Akshaya Dinesh, Jennifer He, "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recordings", IEEE MIT Undergraduate Research Technology Conference (URTC), 2017.

[9] P.Shrivastava, A. Shukla, P.Vepakomma, N. Bhansali, K. Verma, "A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease", COMPUT METH PROG BIO, vol.139, pp. 171-179, 2017.

[10] Najmeh Fayyazifar; Najmeh Samadiani, "Parkinson's disease detection using ensemble techniques and genetic algorithm" 2017 Artificial Intelligence and Signal Processing Conference (AISP).

[11] Md. Inzamam-Ul-Hossain; Lachlan MacKinnon; Md. Rafiqul Islam, "Parkinson disease detection using ensemble method in PASW benchmark" 2015 IEEE International Advance Computing Conference (IACC).

[12] Sachin Shetty; Y. S. Rao, "SVM Based Machine Learning Approach to Identify Parkinson's Disease Using Gait Analysis", 2016 International Conference on Inventive Computation Technologies (ICICT)

[13] "scikit-learn - Low-Variance Feature Removal", https://riptutorial.com/scikit-learn/example/17328/low-variance-feature-removal, [Accessed: 15- May- 2018].

[14] M. Hall 1999, Correlation-based Feature Selection for Machine Learning.

[15] "Recursive Feature Elimination", https://www.scikit-yb.org/en/latest/api/features/rfecv.html,, [Accessed: 15- May- 2018].

[16] "Applying Wrapper Methods in Python for Feature Selection", https://stackabuse.com/applying-wrapper-methods-in-python-for-feature-selection/ [Accessed: 16- May- 2018].

[17] "Exhaustive Feature Selector", https://bit.ly/2Wq2u1J [Accessed: 20- May- 2018].

[18] Li-Yeh Chuang, Chao-Hsuan Ke, and Cheng-Hong Yang, "A Hybrid Both Filter and Wrapper Feature Selection Method for Microarray Classification", - arXiv preprint arXiv:1612.08669, 2016 - arxiv.org.