



# Parkinson's Disease Diagnosis: Detecting the Effect of Attributes Selection and Discretization of Parkinson's Disease Dataset on the Performance of Classifier Algorithms

Gamal Saad Mohamed

Computer Science Department, Faculty of Science, North Border University, Ar'ar, Saudi Arabia

Email: jamalziena@gmail.com

**How to cite this paper:** Mohamed, G.S. (2016) Parkinson's Disease Diagnosis: Detecting the Effect of Attributes Selection and Discretization of Parkinson's Disease Dataset on the Performance of Classifier Algorithms. *Open Access Library Journal*, 3: e3139. <http://dx.doi.org/10.4236/oalib.1103139>

**Received:** October 13, 2016

**Accepted:** November 12, 2016

**Published:** November 16, 2016

Copyright © 2016 by author and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Precise detection of PD is important in its early stages. Precise result can be achieved through data mining, classification techniques such as Naïve Bayes, support vector machine (SVM), multilayer perceptron neural network (MLP) and decision tree. In this paper, four types of classifiers based on Naïve Bayes, SVM, MLP neural network, and decision tree (j48) are used to classify the PD dataset and the performances of these classifiers are examined when they are implemented upon the actual PD dataset, discretized PD dataset, and selected set of attributes from PD dataset. The dataset used in this study comprises a range of voice signals from 31 people: 23 with PD and 8 healthy people. The result shows that Naïve Bayes and decision tree (j48) yield better accuracy when performed upon the discretized PD dataset with cross-validation test mode without applying any attributes selection algorithms. SVM gives high accuracy of 70% for training and 30% for the test when implemented on a discretized PD dataset and a splitting dataset. The MLP neural network gives the highest accuracy when used to classify actual PD dataset without discretization, attribute selection, or changing test mode.

## Subject Areas

Artificial Intelligence

## Keywords

PD, SVM, MLP, Decision Tree, Naïve Bayes, Classifier

## 1. Introduction

Parkinson's disease (PD) is a progressive, neurodegenerative disease that belongs to the group of conditions called motor system disorders. Parkinson's disease sufferers get worse over time as the normal bodily functions, including breathing, balance, movement, and heart function worsen [1].

Other neurodegenerative disorders include Alzheimer's disease, Huntington's disease, and amyotrophic lateral sclerosis or Lou Gehrig's disease. An estimated seven to 10 million people worldwide are suffering from Parkinson's disease. Occurrence of Parkinson's increases with age, but an estimated four percent of people with PD are diagnosed before the age of 50 [2] [3]. There is no cure or prevention for PD. However, the disease can be controlled in early stage. Hence data mining techniques can play effective role in early detection and diagnosis.

Data mining techniques in medicine is a research area that combines sophisticated representational and computing techniques with the insights of expert physicians to produce tools for improving healthcare. Data mining is a computational process to find hidden patterns in datasets by building predictive or classification models that can be learnt from past experience and applied to future cases. With the vast amount of medical data available to hospitals, medical centers, and medical research organizations, the field of medicine supported by data mining techniques can increase healthcare quality and can help physicians make decisions about their patients' care. There are various techniques for classification such as support vector machine (SVM), neural networks, decision tree, and Naïve Bayes. The objective of the study is to analyze and compare four of the abovementioned classification techniques' performances upon Parkinson's diagnosis. First, we compare the classifiers' performance on actual and discretized PD dataset and then compare their performance using the attributes selection algorithm.

## 2. Related Work

Several researches have focused on using data mining techniques for the automatic identification of Parkinson's disease.

Mohammad S. Islam *et al.* [4] conducted a comparative analysis for effective detection on Parkinson's disease using Random tree (RT), SVM and Feedforward back propagation neural network (FBANN). A 10-fold cross validation analysis has been carried out for all classification. The proposed model achieved 97.37%

Aprajita Sharma and Ram Nivas [5] evaluated the performance of the model build using artificial neural networks (ANN), K-nearest neighbor (KNN), and SVM with radial basis function. The models offered a high accuracy of ~85.29%.

Shian Wu and Jiannjong Guo [6] applied factor analysis, logistic regression, decision tree, and ANN to analyze whether voice features can discriminate a PD patient from a healthy one. They stated that among all three methods, decision tree has the lowest classification error and logistic regression model has second lowest, while ANN has a higher classification error.

Geetha Ramani and G. Sivagami [7] provide a survey of data mining techniques that

are in use today for classification. They concluded by showing that the random tree algorithm classified the Parkinson's disease dataset accurately and offer 100% accuracy. The linear discriminant analysis C4.5, C5-MC4, and KNN yields the accuracy result of above 90%.

A. H. Hadjahamadi and Taiebeh J. Askari [8] compared the classification methods (Bayesian Network, C5.1, SVM, ANN, and C&R (Classification and Regression)). C&R has an accuracy of 93.75% whereas ANN and SVM are the next best classifiers. They indicated that the computing of the variable (feature) importance is an important issue in many applied problems complementing variable selection by interpretation issues.

Yahia Alemami and Laiali Almazaydeh [9] developed and validated classification algorithms based on Naïve Bayes and KNN; their results show that the automated classification algorithm, Naïve Bayes, and KNN obtained a high degree of accuracy around 93.3%.

Rashidah *et al.* [10] proposed a model in early detection and diagnosis of PD by using the Multilayer Feedforward Neural Network (MLFNN) with Back-propagation (BP) algorithm. The output of the network is classified into healthy or PD by using K-Means Clustering algorithm. The result shows that the model can be used in diagnosis and detection of PD due to the good performance, which is 83.3% for sensitivity, 63.6% for specificity, and 80% for accuracy.

### 3. Parkinson Dataset

We conduct an analysis on real world PD data, where the disease is diagnosed using several features extracted from human voice [11]. The dataset contains 22 features extracted from 31 people, of which 23 suffered from PD. As shown in **Table 1**, each column denotes a particular voice feature, and each row corresponds to one of 195 voice recordings from these individuals. The dataset was created by Max Little of the University of Oxford in collaboration with the National Centre for Voice and Speech, Denver, Colorado.

These extracted features of human voices are used to diagnose PD and to determine who had actually entered the stages of the disease and who were healthy.

### 4. Method

This study applies several classification methods, including Naïve Bayes, SVM, and decision tree (j48) on the PD dataset. The goals of this study are as follows:

- 1) Examine which of the above classifiers give better performance, when applied to the actual PD dataset.
- 2) Examine the effects of attributes selection for PD dataset on the performances of the mentioned classifiers.
- 3) Examine the effects of discretizing PD dataset on the performances of the classifiers.

Attribute selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features [12]. The discretization of

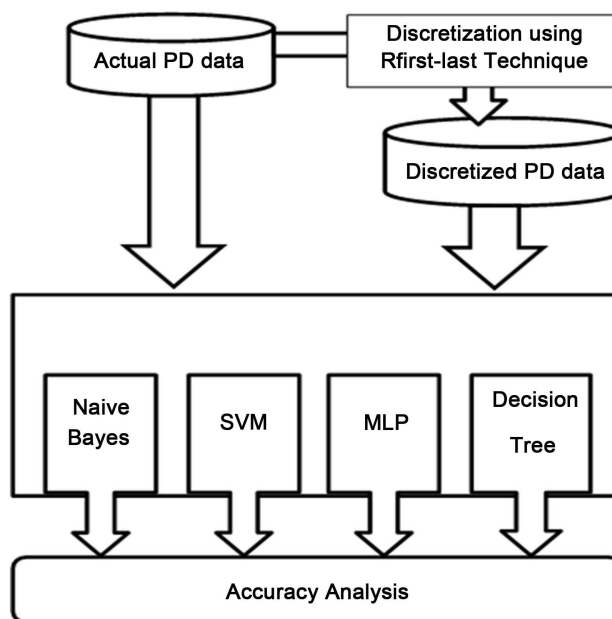
**Table 1.** Parkinson disease features.

No.	Feature name		Meaning
	MDVP: Fo (Hz)	F1	Average vocal fundamental frequency
	MDVP: Fhi (Hz)	F2	Maximum vocal fundamental frequency
	MDVP: Flo (Hz)	F3	Minimum vocal fundamental frequency
	MDVP: Jitter (%)	F4	Jitter as percentage
	MDVP:Jitter(Abs)	F5	Absolute jitter in micro second
	MDVP: RAP	F6	Amplitude perturbation
	MDVP: PPQ	F7	Period perturbation quotient
	Jitter DDP	F8	Average absolute differences between cycle, divided by the average period.
	MDVP: Shimmer	F9	Local shimmer
	Shimmer dB	F10	Local shimmer in decibels
	Shimmer APQ3	F11	3 points amplitude perturbation quotient
	Shimmer APQ5	F12	5 points amplitude perturbation quotient
	MDVP: APQ	F13	11 points amplitude perturbation quotient
	Shimmer DDA	F14	Absolute differences between the amplitude of consecutive periods.
	NHR	F15	Noise-to-harmonic ratio
	HNR	F16	Harmonic-to-noise ratio
	DFA	F17	Signal fractal scaling exponent
	PPE	F18	Pitch period entropy
	spread1	F19	Three nonlinear measures of fundamental frequency variation
	spread2	F20	
	RPDE	F21	
	D2	F22	Recurrence period density analysis
	Status	Health status of the subject (one) Parkinson's, (zero) healthy	

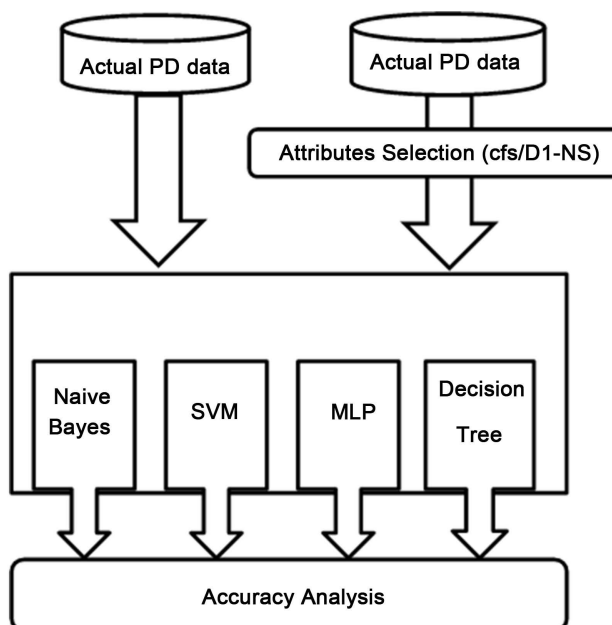
a continuous-valued attribute consists of transforming it into a finite number of intervals and to re-encode, for all instances, each value of this attribute by associating it with its corresponding interval. There are many ways to realize this process [13]. In this study, we implement Weka 3.6.11 software; the Weka tools used for attribute selection and discretization are CfsSubsetEval-BestFirst-D1-N5 and Discretize-R first-last, respectively (Figure 1 and Figure 2).

#### 4.1. Naïve Bayes

Naïve Bayes classifier is used in supervised learning method and it is based on “probability” concept to classify new entities. It assigns a new observation to the most probable class. The classification process comprises two stages as follows [14]:



**Figure 1.** Compute and analysis accuracy for both actual and discretized PD datasets.



**Figure 2.** Compute and analysis accuracy of classifiers for actual PD data with and without attributes selection.

1) Training stage: Using the training samples, the method computes the probability distribution of that sample.

2) Prediction stage: For test sample, the method computes the posterior probability of that unknown instance. The posterior is predicting that the sample belonging to each class according to the largest posterior probability, which is called Maximum A Posterior (MAP).

## 4.2. SVM

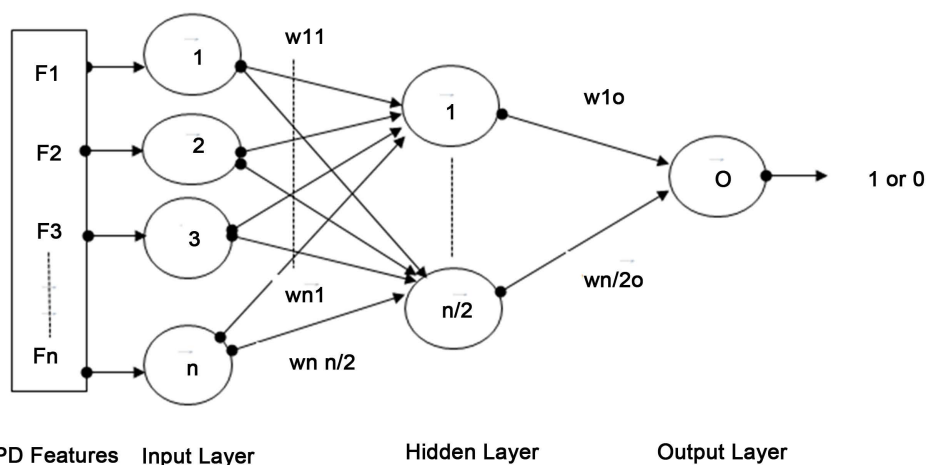
It is used in supervised learning models with associated learning algorithms that analyze data and recognize patterns used for classification. Given a set of training samples, each marked as belonging to one of two classes, an SVM training algorithm builds a model that assigns new examples into one class or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate classes are divided by a clear gap that is as wide as possible. New examples are then mapped into that space and are predicted to belong to a class based on which side of the gap they fall in [15].

## 4.3. Multilayer Perceptron (MLP)

MLP network comprises three layers. A three-layer MLP network is an entirely linked feed forward neural network consisting of an input layer, which is not calculated because its neurons are only for demonstration and therefore do no processing. In addition, a hidden layer and an output layer (PD or healthy), which correspond to the categorization result [16] [17]. **Figure 3** shows the Architecture of the Multilayer Perceptron Neural Network used. Each neuron in the input and hidden layers is linked to all neurons in the subsequent layer through weighted connections. These neurons calculate the weighted sums of their inputs and add a threshold. These sums are used to calculate the neurons' actions by applying a sigmoid activation function. The MLP network utilizes the backpropagation algorithm which is a gradient descent method for weight adjustment. The backpropagation MLP is a supervised ANN. This means the network is presented with input example in addition to the resulting desired output.

## 4.4. Decision Tree (j48)

Decision trees represent a supervised approach to classification. A decision tree is a simple structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. j48 is modified C4.5. The C4.5 algorithm



**Figure 3.** Architecture of the multilayer perceptron neural network.

generates a classification decision tree for the given dataset by recursive partitioning of data. The decision is grown using depth-first search strategy. The algorithm considers all the possible tests that can split the data set and selects a test that gives the best information gain. For each discrete attribute, one test with outcomes as many as the number of distinct values of the attribute is considered. For each continuous attribute, binary tests involving every distinct value of the attribute are considered. In order to gather the entropy gain of all these binary tests efficiently, the training data set belonging to the considered node is sorted for the values of the continuous attribute. Further, the entropy gains of the binary cut based on each distinct value are calculated in a single pass of the sorted data. This process is repeated for each continuous attributes [18].

## 5. Accuracy Analysis

The supervised learning algorithms are applied one after the other. The confusion matrix is a useful tool that determines how well the classifier classifies the instances of different classes. This also shows values such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The classifier accuracy is calculated and a comparative study is done to retrieve the best classifier algorithm.

## 6. Experimental Results

The PD dataset was divided as follows: 70% for training and 30% for testing. The experiment was performed on the abovementioned algorithms as follows:

- Apply the abovementioned algorithms one by one on the actual PD dataset without applying filter algorithm. The Naïve Bayes Algorithm classifies the PD dataset and provides 58.6% accuracy. The SVM yields 86% accuracy. The MLP neural network offers 94.8% accuracy. The decision tree (j48) provides 74% accuracy. **Table 2** shows the accuracy obtained and the value of TP, TN, FP, and FN for each algorithm.
- Using attributes selection algorithm CfsSubsetEval-BestFirst-D1-N5 to the filter PD dataset, the attributes selected were MDVP: Fo (Hz), MDVP: Fhi (Hz), MDVP: Flo (Hz), MDVP: RAP, MDVP:APQ, NHR, Spread1, Spread2, and D2. The accuracy obtained for this case are: Naïve Bayes, 72.4%; MLP neural network, 91.3%; SVM, 86.2%; and decision tree (j48), 82.7%. **Table 3** shows the accuracy obtained when applying attribute selection algorithm and the value of TP, TN, FP, and FN for each algorithm.

**Table 2.** Shows the accuracy obtained when applying a classifier upon the actual PD dataset

Classifier	Number of tested Instances	Confusion Results				Accuracy
		FN	FP	TN	TP	
Naïve bayes	58 (30%)	23	1	10	24	58.6%
SVM	58 (30%)	1	7	4	46	86.2%
MLP	58 (30%)	1	2	9	46	94.8%
Decision tree	58 (30%)	6	9	5	38	74%

**Table 3.** Shows the accuracy obtained when applying attribute selection algorithm

Classifier	Number of tested Instances	Confusion Results				Accuracy
		FN	FP	TN	TP	
Naïve bayes	58 (30%)	14	2	9	33	72.4%
SVM	58 (30%)	1	7	4	46	86.2%
MLP	58 (30%)	0	5	6	47	91.3%
Decision tree	58 (30%)	7	3	8	40	82.7%

- Applying the classifiers on discretized PD dataset, we obtained different values of accuracy: Naïve Bayes, 79.3%; MLP, 94.8%; SVM, 96.5%; and decision tree (j48), 89.6%. **Table 4** shows the accuracy and confusion result for classifiers with the discretized PD dataset.
- When test mode is changed, the classifiers give different values of accuracy. Using cross validation test mode instead of presenting the split of dataset between training and test set, lead to significant change in the accuracy of some classifiers while others showed no change. **Table 5** shows the changes in classifiers' accuracy upon changing the test mode.

As a result, we conclude the following:

- Naïve Bayes gives better performance when it implemented on the discretized PD dataset with cross-validation test mode, yielding 84.6%, which is the best accuracy obtained compared with its performance when implemented on the actual PD data and on selected attributes from PD data.
- SVM yields 96.5%, which is a high accuracy when implemented on discretized PD data and percentage spilt test mode (70% training, 30% test).
- Decision Tree (j48) gives better performance when implemented on discretized PD data yielding 89.6%. Its performance can be enhanced using cross-validation test mode, through which it yields 92.3%.

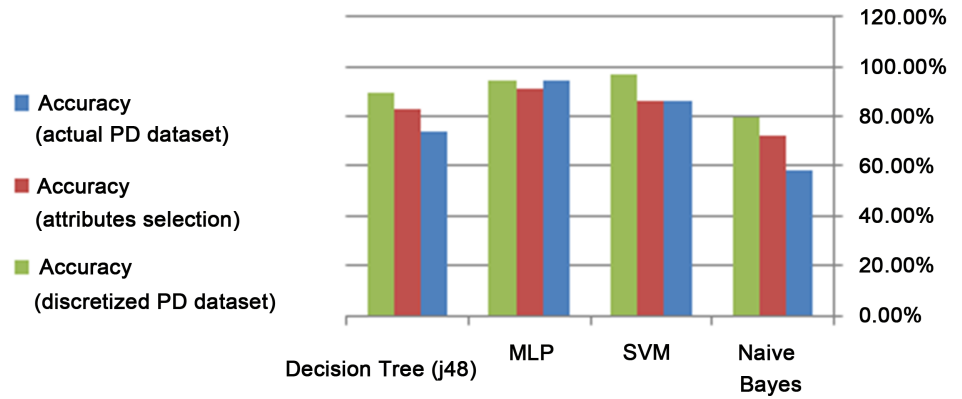
The results show that the best performance can be obtained by MLP neural network for both actual and discretized PD data, *i.e.*, 94.8%. Moreover, the attributes selection algorithm and cross-validation test model had no significant effect on MLP performance when it is used in PDclassification (**Figure 4** and **Figure 5**).

## 7. Conclusions

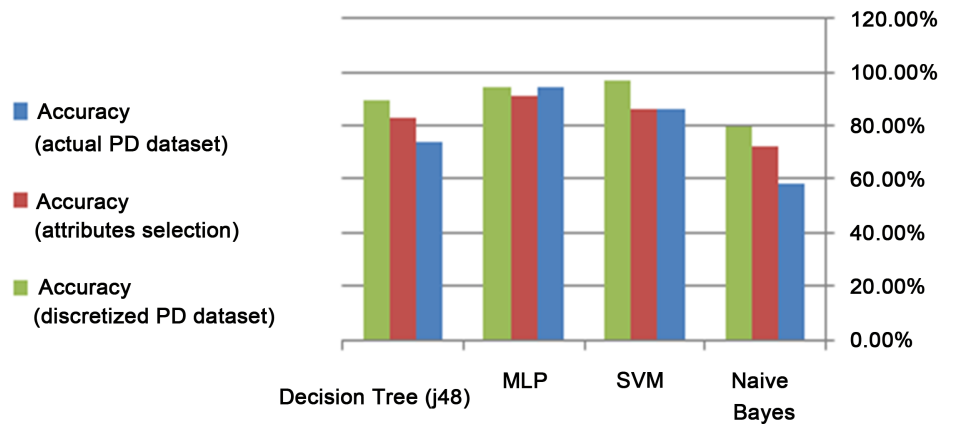
The aim of this study was to recognize how different classifiers would perform when implemented across the PD dataset and to evaluate their performance and examine the effectiveness of attribute selection, discretization, and test mode on the selected classifier performance when implemented on the PD dataset. A comparative study of Naïve Bayes, SVM, MLP, and decision tree (j48) classifiers on PD dataset is performed. This is done by implementing the classifiers upon the following datasets:

- Actual PD dataset.
- Discretized PD dataset.





**Figure 4.** Comparison of classification accuracy for classifiers when implemented on the actual PD dataset, discretized PD dataset, and the selected attributes from the PD dataset.



**Figure 5.** Comparison of classification accuracy for classifiers when performed upon the actual PD dataset, discretized PD dataset, and the selected attributes from the PD dataset using 10-fold cross-validation test mode.

**Table 4.** Shows the accuracy and confusion result for classifiers with discretized PD dataset.

Classifier	Number of tested Instances	Confusion Results				Accuracy
		FN	FP	TN	TP	
Naïve bayes	58 (30%)	9	3	8	38	79.3%
SVM	58 (30%)	1	1	10	46	96.5%
MLP	58 (30%)	3	1	11	44	94.8%
Decision tree	58 (30%)	4	2	9	43	89.6%

**Table 5.** Accuracy when using cross-validation testing mode.

Classifiers	Testing mode: 10-fold cross-validation		
	Actual PD dataset	Applying discretization	Applying attributes selection
Naïve Bayes	69%	84.6	77.9
SVM	87.6%	93.8%	87.1%
Neural network	91%	91%	90%
Decision Tree	85.6%	92.3%	87.6%

- Selected set of attributes from PD dataset.
- Shifting between percentage split and 10-fold cross validation test modes.

From the experimental result, we conclude that Naïve Bayes and decision tree (j48) yield better accuracy when implemented upon the discretized PD dataset with cross-validation test mode without applying any attributes selection algorithms. SVM gives high accuracy when implemented on discretized PD dataset and splitting dataset (70% for training and 30% for test). The MLP neural network gives the highest accuracy when used to classify actual PD dataset without discretization, attribute selection, or by changing test mode.

In conclusion, data discretization enhanced the performance of all classifiers except MLP. Attribute selection algorithm increases only the performance of Naive Bayes and Decision Tree (j48). The training methods had no significant impact on all classifiers performances.

## References

- [1] Shen-Yang, L., Puvanarajah, S.D. and Ibrahim, N.M. (2011) Parkinson's Disease: Information for People Living with Parkinson's. Novartis Corporation (Malaysia) Sdn. Bhd. and Orient Europharma (M) Sdn. Bhd.
- [2] [http://www.pdf.org/en/parkinson\\_statistics](http://www.pdf.org/en/parkinson_statistics)
- [3] [http://www.ninds.nih.gov/disorders/parkinsons\\_disease/parkinsons\\_disease\\_backgrounder.htm](http://www.ninds.nih.gov/disorders/parkinsons_disease/parkinsons_disease_backgrounder.htm)
- [4] Islam, M.S., Parvez, I., Deng, H. and Goswami, P. (2014) Performance Comparison of Heterogeneous Classifiers for Detection of Parkinson's Disease Using Voice Disorder (Dysphonia). *3rd International Conference on Informatics, Electronics & Vision*.
- [5] Sharma, A. and Giri, R.N. (2014) Automatic Recognition of Parkinson Disease via Artificial Neural Network and Support Vicyor Machine. *IJITEE*, **4**, 35-41.
- [6] Wu, S. and Guo, J. (2011) A Data Mining Analysis of Parkinson's Disease. Scientific Research. <http://www.SciRp.org/journal>
- [7] Ramani, G. and Sivagami, G. (2011) Parkinson Disease Classification Using Data Mining Algorithms. *International Journal of Computer Application*, **32**.
- [8] Hadjahamadi, A.H. and Askari, T.J. (2012) A Detection Support System for Parkinson's Disease Diagnosis Using Classification and Regression Tree. *Journal of Mathematics and Computer Science*, **4**, 257-263.
- [9] Alemami, Y. and Almazaydeh, L. (2014) Detecting of Parkinson Disease through Voice Signal Features. *Journal of American Science*, **10**.
- [10] Olanrewaju, R.F., Sahari, N.S., Musa, A.A. and Hakiem, N. (2014) Application of Neural Networks in Early Detection and Diagnosis of Parkinson's Disease. *International Conference on Cyber and IT Service Management*.
- [11] Little, M.A., McSharry, P.E., Hunter, E.J. and Ramig, L.O. (2008), Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, **56**, 1015-1022. <http://dx.doi.org/10.1109/TBME.2008.2005954>
- [12] Muhlenbach, F. and Rakotomalala, R. (2005) Discretization of Continuous Attributes. In: Wang, J., Ed., *Encyclopedia of Data Warehousing and Mining*, Idea Group Reference, 397-402.
- [13] Fayyad, U.M. and Irony, K.B. (1993) Multi-Interval Discretization of Continuous Valued

Attributes for Classification Learning. *13th International Joint Conference on Artificial Intelligence*, 1022-1027.

- [14] <http://www.mathworks.com/help/stats/compactclassificationnaivebayesclass.html?searchHighlight=naive%20bayes%20documentation>
- [15] Cristianini, N. and Shawe-Taylor, J. (2000) An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge.  
<http://dx.doi.org/10.1017/CBO9780511801389>
- [16] Haykin, S. (1998) Neural Networks: A Comprehensive Foundation. Prentice-Hall, Englewood Cliffs.
- [17] CM Bishop. (1995) Neural Networks for Pattern, Recognition. Oxford University Press, Oxford.
- [18] Zhao, Y.H. and Zhang, Y.X. (2008) Comparison of Decision Tree Methods for Finding Active Objects. *Advances in Space Research*, **41**, 1955-1959.  
<http://dx.doi.org/10.1016/j.asr.2007.07.020>



Open Access Library

**Submit or recommend next manuscript to OALib Journal and we will provide best service for you:**

- Publication frequency: Monthly
- 9 [subject areas](#) of science, technology and medicine
- Fair and rigorous peer-review system
- Fast publication process
- Article promotion in various social networking sites (LinkedIn, Facebook, Twitter, etc.)
- Maximum dissemination of your research work

Submit Your Paper Online: [Click Here to Submit](#)

Or Contact [service@oalib.com](mailto:service@oalib.com)