International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 03 – Issue 06 || June 2018 || PP. 57-60

Feature reduction techniques an efficient tool for Parkinson's Dataset

Vinitha Dominic
Presidency University, Bnagalore

Abstract: The goal of this paper is to analyse and identify the best Machine Learning technique that can be applied to predict Parkinson's disease which is a challenging problem for researchers and doctors.Intial step of pre-processing of data is done. Then feature selection techniques are applied on the pre-processed datasets, this reduces the features and improves the efficiency of the classifiers.

These reduced set of features are then used for classification. These features are then validated in terms of their accuracy for different classifiers. It is observed that these reduced set of features are anatomically relevant. Thus we can conclude thst feature selection technique help in better accuracy for prediction

Keywords: Featuren Subset selection, Classifier

Introduction

Information technology has transformed the way health care is carried out and documented. The health care domain generates exchanges and stores a lot of patient-specific data. Machine Learning techniques can be applied to healthcare domain to catalyze and support goals like bypassing clinical trails, finding adverse drug reactions, reducing hospital acquired infections, and rooting out fraud. There are many machine learning techniques that can be applied for exploration of clinical data [1]. There are some real life scenarios where these techniques have proved to be useful.[1].

The Clinical data has various domains like Parkinson's disease, cancer, diabetes and drugs. The focus of study is Parkinson's disease. A lot of research has been carried out on how effectively machine learning techniques can be used effectively for prediction of Parkinson's disease. Study has also been done to compare the various machine learning techniques especially classification for prediction of Parkinson's disease [2-3]. Another important highlight from the past work is introducing feature selection.

The datasets that is considered for study have a total of 74 attributes, but from the past work we observe that all the study is done using only 13 attributes. These attributes which are understood even by common man appear to be significant for a diagnosis of Parkinson's disease.

A brief discussion on these attributes will be done in the further sections of this paper. The past study has tried to reduce these 13 attributes by applying feature selection techniques. Genetic search was one of the feature selection technique that was and a reduction of features from 13 to 6 was observed [4]. Furthermore effort was made to reduce these 6 attributes to 4 attributes and significant improvement in the accuracy was observed [5].Various computational intelligence techniques like Multilayer Perceptron, Neural Networks were used in effective detection of Parkinson's disease [6-8].

The main focus in this paper is to study the Parkinson's disease datasets with attributes, performance and relevance of these attributes. Next, step is to apply two feature selection techniques on these features to get a reduced set of features. And finally we validate these reduced set of features on a set of classifiers in terms of accuracy.

Material and Method

Dataset Description

Here, the dataset was created by the authors of the paper [9] Max little University Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. This dataset is composed of a range of biomedical voice measurements from 31 people,23 with Parkinson's disease (PD). Each attribute is a particular voice measure, and each tuple correspond to one of 195 voice recording from the individuals.

The main goal of the dataset is to distinguish healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. There are various attributed extracted that are defined as follows:

MDVP: Flo (Hz) Minimum vocal fundamental frequency
Number MDVP: Fo (Hz) Average vocal fundamental frequency
MDVP: Fhi (Hz) Maximum vocal fundamental frequency
MDVP: Jitter (%), MDVP: Jitter (Abs), MDVP: RAP,

MDVP: PPQ,
Jitter: DDP SeveralMeasures of variation in fundamental frequency MDVP: Shimmer, MDVP: Shimmer (dB),Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA,several measures of variation in amplitude
NHR, HNR: Two measures of ratio of noise to tonal components in the voice

Feature Selection Techniques

Two feature selection techniques Genetic Search. Genetic search introduces the principle of evolution and genetics into search among possible solutions to given problem. This is done by the creation of individuals represented by chromosomes. A fitness function is derived for evaluation of the individuals. This is then followed by the process of crossover and mutation. The above process is repeated for determined number of generations, and results are analyzed [14].

Results and Discussion

The analysis is done by applying feature reduction technique ,reduced set of attributes are tested using classifiers. The attribute Num is the Class attribute i.e. if 0 absence of Parkinson’s disease else 1 presence of Parkinson’s disease. A brief overview of the procedure followed is shown in Figure.

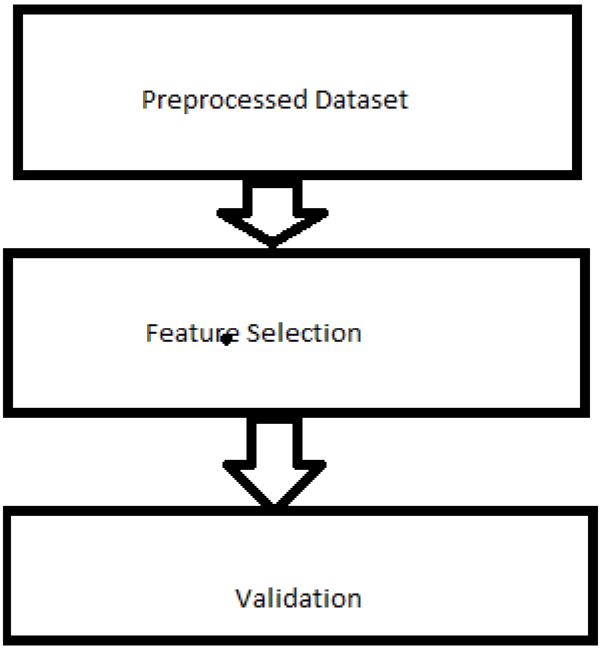

Figure 1. Steps of Analysis

Result Analysis

We have applied the feature selection technique Repeated iterations of this is done, till maximum accuracy is achieved for all the classifiers.

A slight improvement in performance is observed in accuracy for attributes after feature selection. In Figure 2 and Figure 3 we observed a slight improvement in performance compared to the performance of the 14 attributes. The blue color line depicts the reduced features and red color is the baseline performance

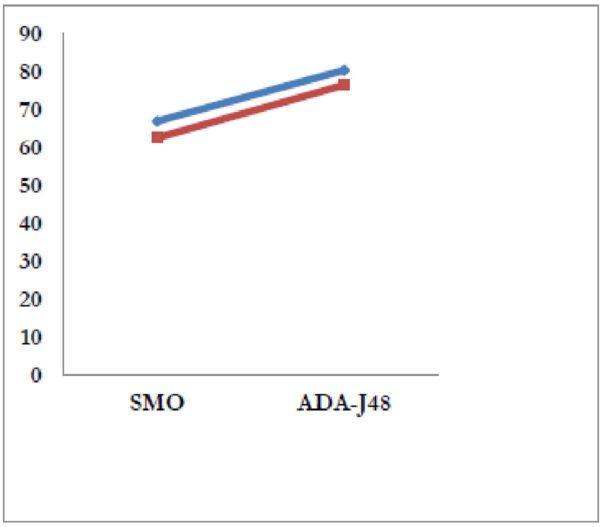

Figure 2 Accuracy of attributes after feature selection

In the second stage of analysis, test data sample is extracted and tested with the reduced set of features.

The performance is observed to be remarkable. Thus feature reduction techniques can be applied to

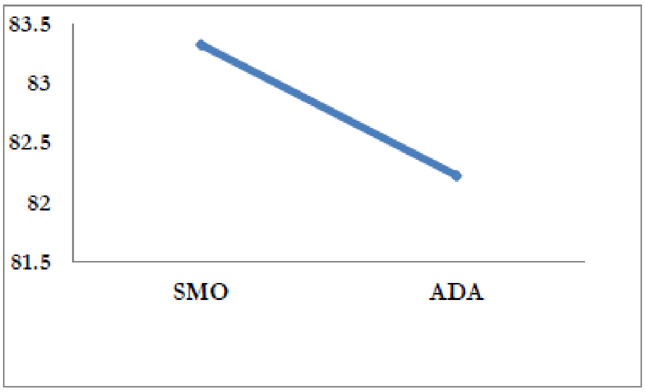

Figure 3 Performance of reduced attributes with test dataset

Conclusion

The datasets were pre processed and two feature selection method Genetic Search was applied. This method have given an improvement in the accuracy of the classifiers. The feature selection gives better results and can be proved more effective based on the data.

The features obtained are validated to be significant in the diagnosis of Parkinson’s disease by a medical practioner. Thus we can conclude that feature selection is an effective tool to derive knowledge from clinical data. Based on the availability of quality data it can be proved to be more effective.

Future Work

This analysis can be carried out on better quality data to get improved results. Then another change is to give only two classes instead of four i.e class **0** absence of Parkinson’s disease and **1** presence of Parkinson’s disease. Many more feature selection methods can be analysed and thus improve the accuracy. This analysis can be done on various other domains of healthcare.. Effective decision making systems can be introduced to provide better healthcare for the society.

References

[1]. Illhoi Yoo & Patricia Alafaireet & Miroslav Marinov & Keila Pena-Hernandez & Rajitha Gopidi & Jia-Fu Chang & Lei Hua *Data Mining in Healthcare and Biomedicine: A Survey of the Literature.* Springer Science February 2011

[2]. K.Sudhakar, Dr. M. Manimekalai, *Study of Parkinson's disease Prediction using Data Mining,* International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014.

[3]. Duraisamy.K, Haridass.K, *An Effective Comparison of SVM and CN2Rule Using Heart Dataset: A Survey*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014.

[4]. Shruti Ratnakar, K. Rajeshwari, Rose Jacob, *Prediction of Parkinson's disease Using Genetic Algorithm For Selection of Optimal Reduced Set Of Attributes,* International Journal of Advanced Computational Engineering and Networking, ISSN (p): 2320-2106, Volume-1, Issue-2, April-2013

[5]. Nidhi Bhatla Kiran Jyoti A Novel Approach for Parkinson's disease Diagnosis using Data Mining and Fuzzy Logic . *International Journal of Computer Applications (0975 – 8887) Volume 54– No.17, September 2012*

[6]. Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle , Yi-Ping Phoebe Chen *Computational intelligence for Parkinson's disease diagnosis: A medical knowledge driven approach,* Expert Systems with Applications 40 (2013) .

[7]. Ref: Miss. Chaitrali S. Dangare1, Dr. Mrs. Sulabha S. Apte, *A Data Mining Approach For Prediction of Parkinson's disease Using Neural Networks,* INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY(IJCET),Volume 3, Issue 3, October - December (2012), pp. 30-40.

[8]. Dimple, *Classification of Data for Parkinson's disease Prediction System Using MLP ,* International Journal in Multidisciplinary and Academic Research (SSIJMAR) Vol. 2, No. 5, August – September 2013 (ISSN 2278 – 5973).

[9]. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/cleveland.data

[10]. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/hungarian.data

[11]. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/long-beach-va.data

[12]. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/switzerland.data