ParkDiag: A Tool to Predict Parkinson Disease using Data Mining Techniques from Voice Data

Tarigoppula V.S. Sriram[1], M. Venkateswara Rao[2],
G.V. Satya Narayana[3] and D.S.V.G.K. Kaladhar[4]

1 CSE, Raghu Engineering College, Visakhapatnam, India
2 IT, GITAM University, Rushikonda, Visakhapatnam, India
3 IT, Raghu Institute of Technology, Visakhapatnam, India
4 Dept. of Microbiology and Bioinformatics, Bilaspur University, Bilaspur, India

Abstract - In the present decade there is more concentration on ageing disease. Some diseases are going to through their symptoms at early stage. The problem with aging disease like cancer, diabetes Alzheimer and Parkinson are if they are going to un-cover at advanced stage it is very difficult to cure them totally , instated doctors can able to stop the growth of the disease up that stage where it was un covered. But it is highly difficult to them to cure the disease up to the root level. Parkinson disease is the one of the most painful and non-curable disease. It is going to occur from the age of 40 plus people. Few detailed clinico-pathological correlations of Parkinson's disease have been published. The pathological findings in 100 patients diagnosed prospectively by a group of consultant neurologists as having idiopathic Parkinson's disease are reported. There are some symptoms to identify the disease; the symptoms like to find the Parkinson Disease (PD). The dataset for the disease is collected from the people who are already having Parkinson. However, genetic causes of Parkinson's are rare, only in approximately 6–8% of all cases. The details are collected from PD patients especially the voice data of PD patient were collated by recording their voice by pronouncing the ovals A,E,I,O,U repeatedly minimum 10 times form each PD patient. The results were analyzed by using data mining techniques, Bayes Net shows 70% , naïve bayes shows 83% , KStar and ADTree shows 100% and Random Forest classification results towards PD.

Key words: *PD, Voice recognition, Data mining, ParkDiag, Parkinson disease.*

I. INTRODUCTION

In the present decade there is an interest in designing tool for the diagnosis of ageing diseases. There is a little work that has been done to diagnose on ageing diseases. The present research focus on the design of ParkDiag tool for the diagnosis of Parkinson disease. Parkinson is the major health problem in human beings especially during middle age who crosses 40 years are the sufferers of the disease . The present decade focus is to understand the factors responsible for Parkinson disease. The voice recognition is also one of the solutions that can be used to diagnose the PD. The data mining algorithms also provide major solutions to understand the factors responsible for the cause of PD.

The symptom's that cause disease are metabolic, mainly control of genes (mutations). Change in the environmental factors and clutter conditions are the factors responsible for this disease[3]. .These mutational factors at genetic level provides change at the phenotypic level (changes on function of body)[9]. The changes occur in the body can be predicted as external factors like voice changes and changes in color of skin, color of hair etc.

Biometric recognition refers to an automatic recognition of individual based on a feature derived either from physiological or behavioral characteristics or both. Biometric recognition system should provide a reliable personal recognition schemes either to confirm or to determine the identity of an individual [6].

This paper reviews the voice characteristics and identification techniques used in recognized PD people by their voices[7]. A discussion of inherent performance limitations along with a review of the performance achieved by listening, visual examination of spectrograms , automatic computer techniques, attempts to provide a perspective which evaluate the potential of speaker recognition and productive directions for research and application of speaker recognition technology[2]. In this paper, brief view of biometric methods both single and multimodal, their advantages and disadvantages, are be presented. All the people of PD don't need to have same symptoms . The new symptoms may develop or change with time as the disease progresses'. People will experience both motor and non-motor symptoms. The motor symptoms of PD appear on one side of the body and may extend to other side as the time lapses [5].Research work is going on whether a person when exposed to herbicides, pesticides and toxins may attack PD in future [1].

II. METHODOLOGY

Voice Analysis is the First Step to predict an un-known person will get the disease or not. For this conformation, data will be collected from diseased and healthy persons who crossed 40 years of age. Generally the disease occurs after 40 years. Comparative classification studies on different data-sets in an entity have been applied for accuracy analysis and the time taken to execute the data-set in order to find the best collected classification rule. The data of healthy people and those with Parkinson can be correctly classified by using machine learning and data mining system techniques.

Bayesian theory is a mathematical model in calculating the degree of prediction. BayesNet and NaiveBayes are the most practical learning methods that have a random sequence model within the class. System Classification summarizes a sequence of classification methods using algorithms. Logistic function can be derived from simple classification problems, measuring from minus infinity to plus infinity. Simple logistic regression is used to explore associations between dichotomous outcome with continuous, ordinal or categorical exposure variables. Lazy classification scheme uses Hierarchical SVM's to select a subset of candidate classes for each test instance in order to determine the overall best performer. K-Star is an Instance-Based learner. The class of a test instance is based upon the class of those determined by similarity function with similar training instances. It is different from the other instance-based machine learners using entropy-based distance function.

Meta-classification makes its binary decision by classifying synthesized feature vectors and one meta-classifier for each class was built for each. Bagging (or Bootstrap aggregating) is an algorithm to improve machine learning of statistical classification and regression models in terms of stability and classification accuracy. A statistical method of distance based classification with the best matching rule can be explained by NNGE (Non-Nested Generalized Exemplars).

The classification is used to follow the path dictated by the successive tests placed along the tree until it finds a leaf containing the class to assign a new attribute. ADTree (An alternating decision tree).

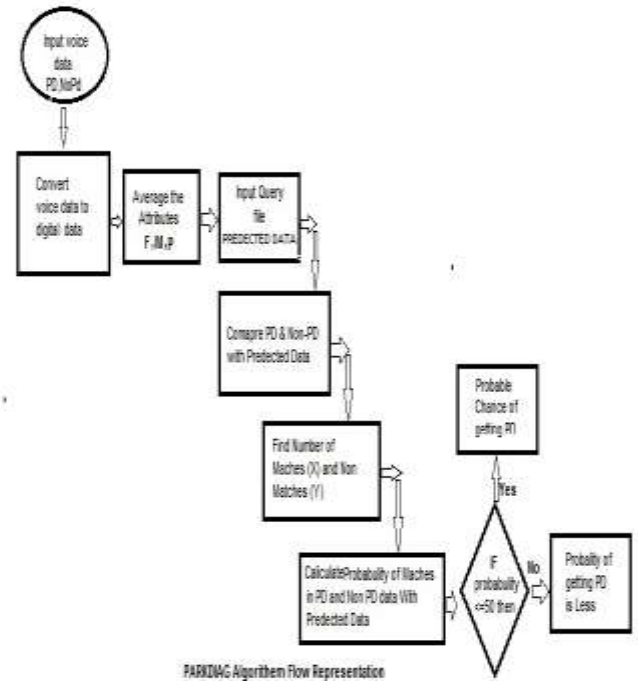

Fig1. Shows the flow of the ParkDiag tool.

The ParkDiag Algorithm is one of the detection based algorithm for the diagnosis of PD. The algorithm has been given below.

Step1 : Input Voice Data.
Step2 : Convert Output Into .CSV File.
Step3 : Average The Attributes (Frequency , Modulation , Phase) .
Step4 : Input The Query File And Run Step 1,2,3.
Step5 : Compare The Data With Tested Data.
Step6 : Predict Based On P(X)=P(X/)P(X)+P(Y)
X= Match With Tested
Y=Miss-Match With Non Tested
Step7 : Output For Probable Occurrence Of Parkinson Disease.
Step8 : Print Predicted Output.

The analysis was done by using one of the data mining algorithm navies Bayesian and Bayes networks. The results were analyzed in order to predict PD with the collected voice of Non-PD people.

III. RESULTS AND DISCUSSIONS

Following is the voice data taken from a PD patient. The vowels were pronounced repeatedly 10 times each vowel. The voice is analysed and noice data was removed from the data. Each vowel is analysed and the voice data was converted in to numaric form considering Magnitude, Frequency and Phase as key attributes.

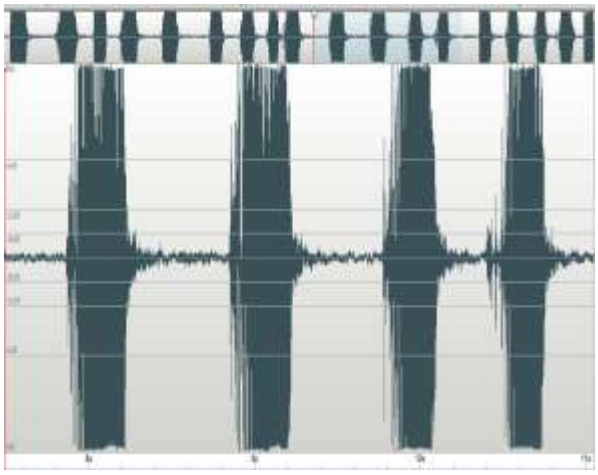

Fig.2. Voice data of a vowel A's.

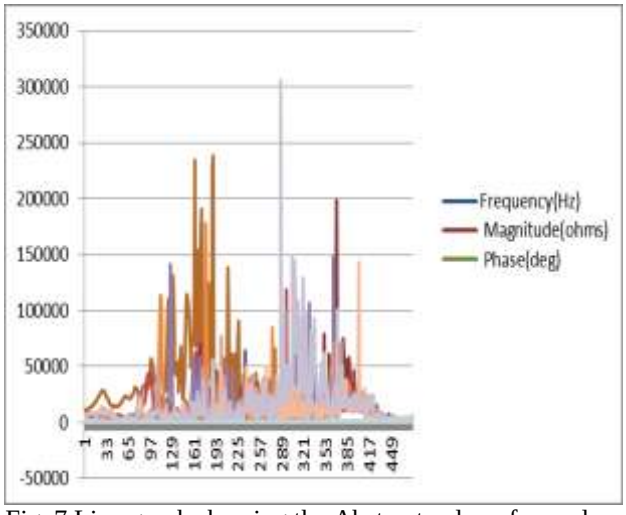

Fig. 7 Line graph showing the Abstract value of vowel "A" of Non-PD persons

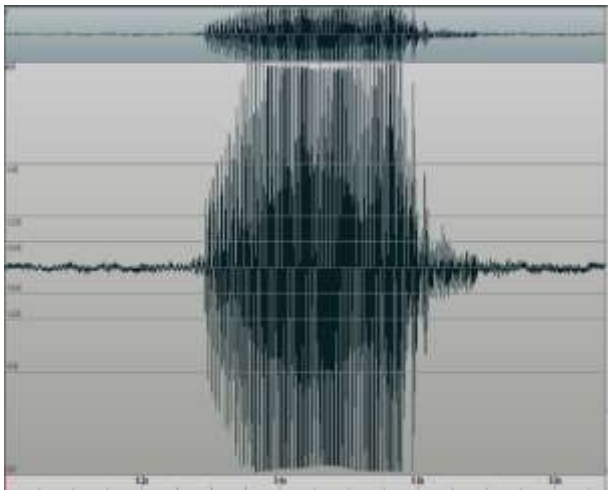

Fig.3. Magnified vowel A's.

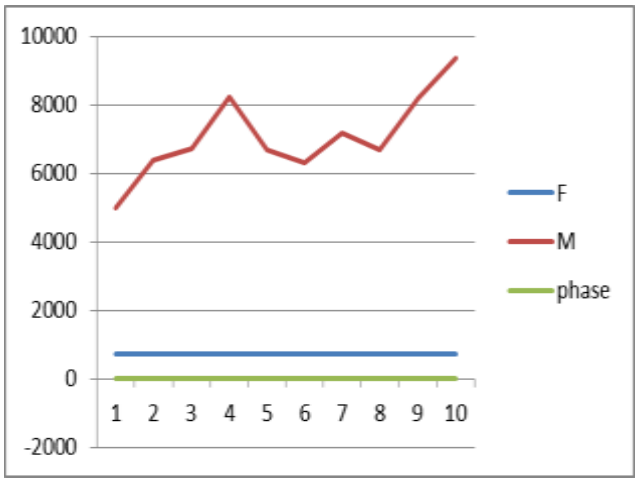

Fig. 8 Predicted aggregate for vowel "A".

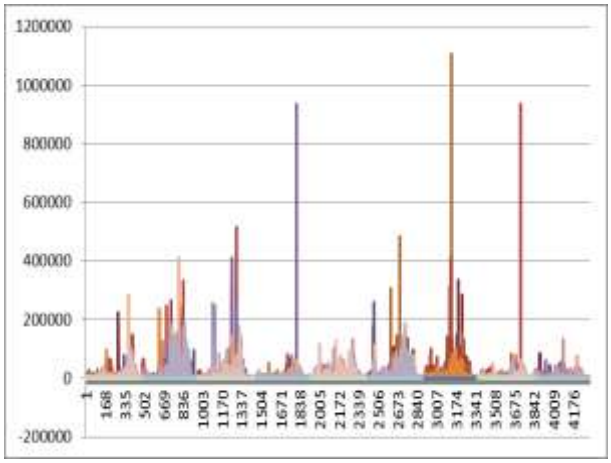

Fig..6 Line graph showing the abstract of the vowel "A" of PD persons.

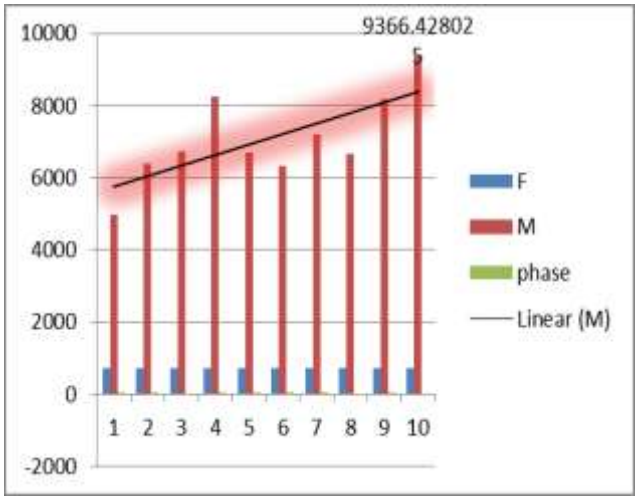

Fig. 9 Bar graph Predicted aggregate for vowel "A".

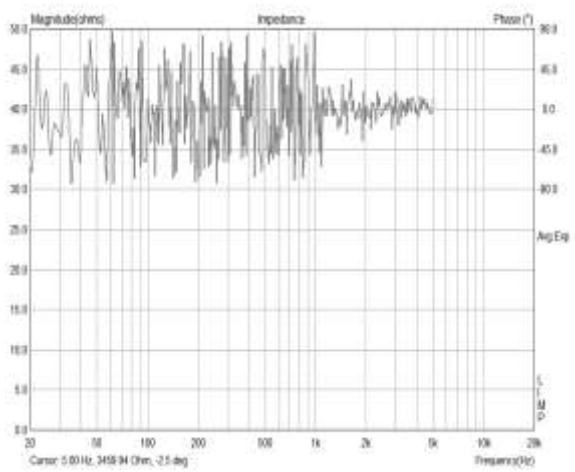

Fig. 4 Frequency, Magnitude and Phase analyzed for vowel "A".

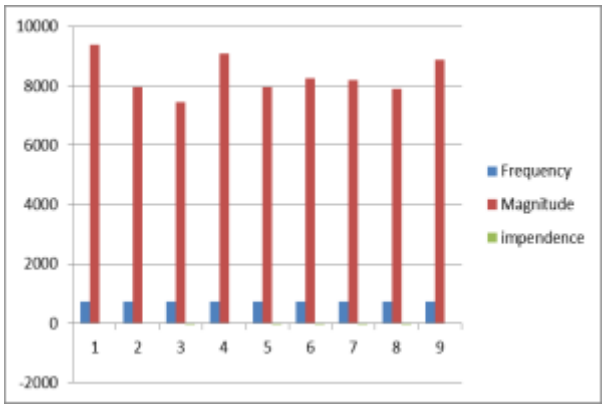

Fig.5 Graphical representation of the Frequency, Magnitude and Phase of vowel "A".

Consider the table with the collected values from PD patients as shown in tabular format.

Attributes	A	E	I	O	U
Frequency	726.8	726.8	726.8	726.8	726.8
Magnitude	9984.2	13013.0	14254.9	18477.5	14401.3
Phase	-0.6	0.2	0.9	0.2	-0.6

Table 1. Aggregate value of all vowel shown in form of a table of some of the PD patients

Probability :

P(X) = P(X) / P(X) + P(Y)
X=5
Y=3
P(5) = 5/(5+3) =5/8 = 0.625
Most probably query shows Parkinson Disease.

Result: **The person may get PD 62.5% , that is more than 50%.**

Attributes	A	E	I	O	U
Frequency	726.8	726.8	726.8	726.8	726.8
Magnitude	12867.0	11996.0	21149.0	4487.6	4126.8
Phase	0.0	0.5	1.0	-0.2	-1.4

Table 2. Aggregate value of all vowel shown in form of a table of some of the Non-PD patient.

Probability:

P(X) = P(X) / P(X) + P(Y)
X=4
Y=5
P(4) = 4/(4+5) = 4/9 = 0.44

Result: **The person may not get PD 44% , that is less than 50%.**

Algorithm	**Correctly Classified Instances**
Bayes Net	70%
Naïve Bayes	83%
Logistics	100%
Simple Logistics	50%
KStar	100%
ADTree	100%
J48	92%
Random Forest	83.33%

The collected data for correctly classified was analyzed in various Data mining Algorithms , the results are shown in the following table.

Table.4 Correctly classified results.

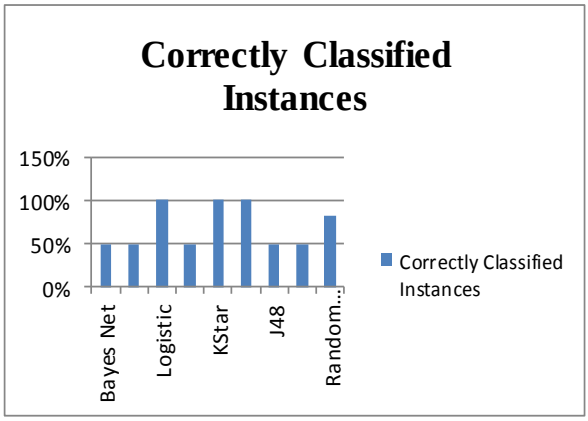

Figure.9 Bar graph showing correctly classified Instances.

IV. CONCLUSION

The PARKDIAG algorithm is the novel model for the diagnosis and prediction of PD. The tool is so simple and no special instruments are needed to analyze the data.

ACKNOWLEDGMENTS

The authors would like to thank GITAM University and Raghu Engineering College for providing computational competence and access to e-Journals to bear out this research.

REFERENCES

[1],T.V.S. Sriram ,M.Venkateswara Rao,.V. D.S.V.G.K. Kaladhar (2013) .Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms, IJEIT Volume 3, Issue 3, September 2014.

[2] T. V.S .Sriram, Venkateswara Rao .M, V. D.S.V.G.K. Kaladhar.,(2014), Diagnosis of Parkinson Disease Using Machine Learning and Data Mining Systems from Voice Dataset, Springer International Publishing Switzerland on(vol 1 pp 151-177).

[3] A. K.Jain, A.Ross,&S.Prabhakar, (2004). An introduction to biometric recognition. Circuits and Systems for Video Technology, IEEE Transactions on, 14(1), 4-20.

[4] G. R. Doddington, (1985). Speaker recognition—Identifying people by their voices. Proceedings of the IEEE, 73(11), 1651-1664.

[6] H.Valbret,E.Moulines,& J.P.Tubach (1992, March). Voice transformation using PSOLA technique. In Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on (Vol. 1, pp. 145-148). IEEE.

[7] C.Schröder, J.Möbes, M.Schütze, F,Szymanowski, W.Nager,M.Bangert, &R.Dengler(2006). Perception of emotional speech in Parkinson's disease. Movement Disorders, 21(10), 1774-1778.

[8] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Biomedical Engineering, IEEE Transactions on, 56(4), 1015-1022.

[9] C.Champod, & D.Meuwly(2000). The inference of identity in forensic speaker recognition. Speech communication, 31(2), 193-203.

[10] R.R.Geetha, G.Sivagami: Parkinson Disease Classification using DataMining Algorithms.International Journal of Computer Applications 32(9), 17–22 (2011).

[11] D.S.V.G.K.Kaladhar, R.P.V.Nageswara , N.R.B.L.V.Ramesh Confusion matrix analysis for evaluation of speech on Parkinson disease using weka and matlab. International Journal of Engineering Science and Technology 2(7), 2734–2737 (2010).

[12]C.Tanner, R.Ottman, S.M.Goldman, J.Ellenberg, P.Chan, R.Mayeux, L.J.William: Parkinson Disease in Twins: An Etiologic Study. JAMA 281(4), 341–346 (1999).

[13] M.E.Smith, L.O.Ramig, C.Dromey, Perez K. S., Samandari, R.: Intensive voice treatment in parkinson disease: Laryngostroboscopic findings. Journal of Voice 9(4), 453–459 (1995).