Smart Intelligent Computing and Communication Technology
V.D. Ambeth Kumar et al. (Eds.)

doi:10.3233/APC210029

Discriminating Parkinson and Healthy People Using Machine Learning

S. Lakshmi [a,1], Dr.Manonmani.A[b] , Jagadeeshwaran P [c]AkashS[d], JananiK[e], AkilandeswariD[f]
[a]*Associate Professor Dept of EIE, Panimalar Engineering College, Chennai, India*
[b]*Assistant Professor, Dept of EIE, Saveetha Engineering College, Chennai, India*
[c,d,e,f]*UG Scholar, Dept of EIE, Panimalar Engineering College, Chennai, India*

Abstract. Parkinson's is a disease which affects at central nervous system of human body and it is popularly known as neuro degenerative disease. we have taken a classification method were mainly focused on speech signal for detection of parkinson's disease because speech is the only deciding factor for earlier stage of the disease identification. Here the ensemble technique is used for classification. Ensemble is method of combining multiple models like classifiers or experts to solve a particular computation .In proposed technique with help of speech signal it is easy to distinguish between diseased and non diseased person. here we are using vowels, number and words from the speech signal in classifiers then how the signal involved in that technique are discussed.

Keywords. Parkinson's disease, Neuro degenerative disease, central nervous system, ensembles technique.

1. Introduction

Dr. James Parkinson describesparkinson's disease in 1817 as a "shaking palsy.".parkinson's is a neurodegenerative disease affects at central nervous system.It affects in motor system of te body. which includes symptoms like bradykinesia, muscular rigidity and resting tremor,. PD is one of the most common neurodegenerative disorders. Risk Factor Associated With Parkinson Disease: Elevated Cholesterol, Increased body mass index and High caloric intake. Diagnosis: diagnosis for parkinson's includes physical examination and comprehensive history of diseased persons. Prediction of Parkinson's disease using speech signal with extreme learning machine[1]. MFCC and SVM are used for analyzing voice prints and detect that patients with Parkinson's disease or not[3]. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system[5]. Diagnosis of Parkinson's disease using electrovestibulography[6]. Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors[10]. Collection of multiple sound records and analyzing with Parkinson speech dataset .[11]. .Symptoms: • Bradykinesia- Slowness of movement (80% to 90% it occurs) • Rigidity, tremor at rest, Postural instability, Dysarthria, Dystonia, Anxiety, cognitive impairment,, dementia.

[1]Lakshmi.S, Associate Professor Dept of EIE, Panimalar Engineering College, Chennai, India. Email:elzie.moses@gmail.com[1]

2. Existing System

In the dataset, it will consider sentence, words, vowels, etc., and also it considers numbers; starts from 0 to 10. Here we have used a different type of method to identify the Parkinson disease. Classifiers are used to differentiate between disease peoples and non- disease people. the feed forward technique, one of the most important of technique for the classification.
Flowchart of Existing System:

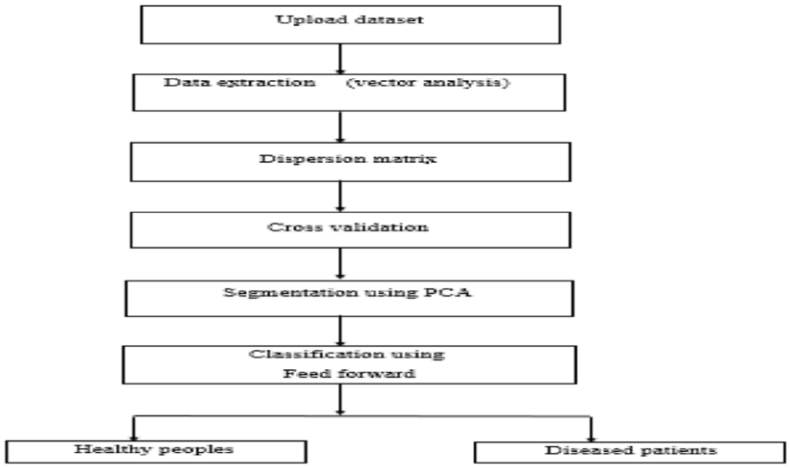


Figure 1. Proposed System

Based on the Speech Signals:
In this experiment we use different types of voice records such as jitter, shimmer, mean pitch, medium pitch and maximum pitch, etc. Consider 26 attributes are generated from the dataset.Jitter percent is also known as jitter factor. Jitter can be calculated using Boken and Orlikff equation, [Σ|Xi-Xi-1|/(n-1) /[Σ(Xi) /(n)]*100. For the normal waveform of jitter percent is 0.509% and dysphonic waveform of jitter percent is 4.673%. This can be considered as dysphonic waveform is greater than normal wave form. Shimmer is also called amplitude perturbation. It refers the intensity of the amplitude. Shimmer can be calculated using Boken and Orlikff equation i.e. [Σ|20log (Yi /Yi-1)|] /n-1, For the normal waveform of shimmer is 0.102dB. For the dysphonic waveform of shimmer is 2.005dB. Therefore dysphonic waveform is greater than normal waveform. Quality factor A can be calculated as maximal and minimal sound pressure i.e. amplitude.Outcome of Existing System: This technique works well on the dataset. In 40% non-diseased patients and 60% diseased patients we got a most amazing F measure from the dataset. In the existing system the feed forward technique with the basic concepts of digital image processing leads to a better outcome with more accurate result. A feed forward technique is introduced in existing system.
Drawbacks of Existing System: Accuracy is highly commendable, Binary attribution is only used, Risk of over-fitting the data, Not superior for arithmetic and small calculations, Explanation are not in Neural networks
Problem Identification: In the existing system accuracy using voice analysis reached is only 92% which has to be increased.

3. Proposed System

Prediction in existing system is weak due to a single classifier system. To overcome the limitation factor (Accuracy) in proposed system Ensemble Technique is used. Ensemble method combines several machine learning techniques into one predictive model in order to decrease variance (bagging), improve predictions (stacking) or bias (boosting).ensemble method is a kind of Meta algorithms means several algorithm or classifiers are combined where the prediction of each classifier is taken into an account and averaged to have final prediction. In proposed system, Naive Baye's, Cross Validation, F1 Score combination of algorithms is used. Ensemble learning is generally used to improve the (classification, prediction, function approximation, etc.) performance of a model, or reduce the likelihood of an unfortunate selection of a poor.
Data Extraction of Proposed System: X-axis = Time , Y-axis = Frequency

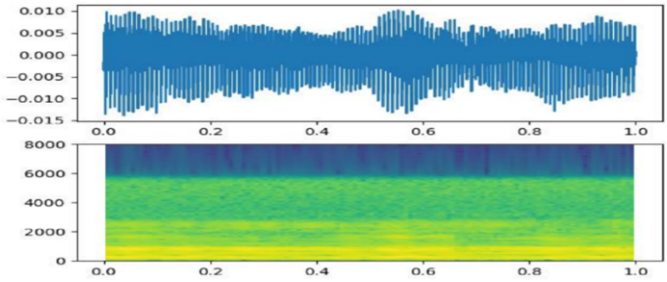
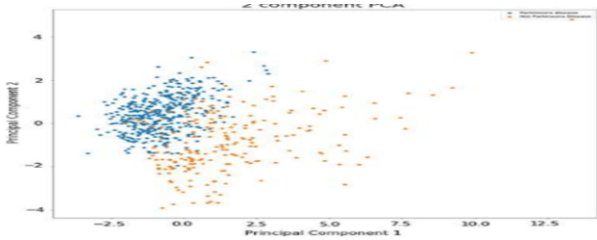


Figure 2. Data Extraction of Proposed System

Table 1. Dispersion Matrix of Proposed System

JITTER LOCAL	JITTER LOCAL ABSOLUTE	SHIMMER LOCAL	SHIMMER LOCAL ABSOLUTE	MEAN PITCH	MEDIUM PITCH	MAX PITCH
0.078	0.135	7.3E-06	0.067	0.078	0.202	7.3E-6
0.081	0.143	7.1E-06	0.73	0.081	0.219	7.1E-6
0.089	0.162	0.000008	0.087	0.089	0.26	.00008
0.089	0.14	6.9E-06	0.075	0.089	0.24	6.9E-6
0.098	0.15	7.2E-06	0.8	0.097	0.23	7.2E-6

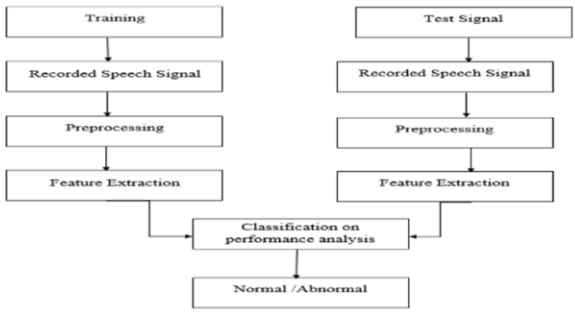
PCA Segmentation of Proposed System:
.yellow dots - Non –Parkinson Person .blue dots – Parkinson Person.



Steps of Proposed System:

1: In this step we insert the input dataset which has the details about speech signals. Based on the speech signals we identify the Parkinson disease.2: In the second step dataset is converted into a Waveform analysis i.e. data extraction.3: In this step Dispersion matrix is created. Dispersion matrix is nothing but read the data from the data source and created in a matrix form. 4: In this step cross validation done. Once the matrix is created, it will be cross validating the matrix.5: In this step segmentation using PCA is done. PCA is principle component analysis on that basis there are four parameter values are used. We calculate each row in the database using the four parameters values. That's why PCA is used here. Segmentation is nothing but separate, separately calculating the values.6: In this step using feed forward classification is done based on speech signals. Ensembler technique identifies the diseased patients and non-diseased patients.

Flowchart of Proposed System:



Outcome of Proposed System: In this technique we identify the maximum F measure value. The technique works well on the dataset. In 40% non-diseased patients and 60% diseased patients we got a most amazing F-measure from the dataset. Ensembler technique is introduced in this proposed system. In addition to this, it is used for based on the speech signals. Ensembler technique classifies the diseased patients and non-diseased patients. From this project we check the efficiency from the respective dataset.

Advantage: 1. Accuracy is increased compared to existing system. 2. Possibility of over-fitting the data. This essentially means that the model has the ability to large dataset. 3.they can be used with less data to dive the models.4. Good for arithmetic and precise calculations.

Steps To Implement:

1. Collect the sample voice of a person. 2. Convert audio file of the voice sample (mp3) into wave format (.wav). 3. Convert the wave format file (.wav) into CSV file. 4. Use the CSV file as input to the appropriate python code and run the program. 5. If the output shows '0', the person have Parkinson Disease. 6. If the output shows '1', the person doesn't have Parkinson Disease.

Results and Discussions:

Audio Waveform of Healthy People:

Here the audio file of the voice sample collected from the person is converted into wave format (.wav) in which the frequency of the voice sample is represented with respect to time. X-axis = Time, Y-axis = Frequency .time. The below graph represents the waveform of Healthy and PD persons.

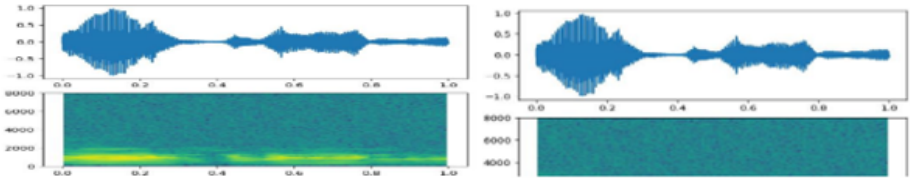


Figure 3. Representation of Normal Person's and PD Person's Frequency Waveform

*Representation of Output :*In proposed system, the accuracy reached is 98%. The F1 score, precision, recall factors are calculated for given input voice sample through python program. As we use ensemble algorithm the strength of the prediction is increased. Thus we overcome the limitation of existing system**.**

4. Conclusion and Future Enhancement

An efficient classifier combination method is developed with Ensembler technique for addressing PD diagnosis problem. In The proposed method the core component is the Ensembler classifier, whose key parameters are explored in detail. With the aid of the feature selection techniques, especially the performance of Ensembler classifier is ameliorated with much smaller features. from the obtained PD dataset has proven that the proposed classifier combination method can distinguish well enough between patients with PD and healthy persons. the highest classification accuracy of 98% is observed from Ensembler method that it achieves. Based on the empirical analysis, it is concluded safely that, the diagnosis method developed can assist the physicians to make accurate diagnostic decision. The future investigation will pay much attention to evaluating the proposed method in other medical diagnosis problems. In this system, binary attribute (1- diseased patients, 0-non-diseased patients) is used for patient's classification. In future, we will use different types of attributes for patient's classification and also identify the various stages of Parkinson disease.

References

[1] Agarwal, A., Chandrayan, S., &Sahu, S. S. (2016, March). Prediction of Parkinson's disease using speech signal with extreme learning machine. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 3776- 3779). IEEE

[2] Aich, S., Sain, M., Park, J., Choi, K. W., & Kim, H. C. (2017, November). A mixed classification approach for the prediction of Parkinson's disease using nonlinear feature selection technique based on the voice recording. In 2017 International Conference on Inventive Computing and Informatics (ICICI)(pp. 959-962). IEEE

[3] Benba, A., Jilbab, A., Hammouch, A., &Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In 2015 International conference on electrical and information technologies (ICEIT) (pp. 300-304). IEEE

[4] Chandrayan, S., Agarwal, A., Arif, M., &Sahu, S. S. (2017, March). Selection of dominant voice features for accurate detection of Parkinson's disease. In 2017 Third International Conference on Biosignals, Images and Instrumentation (ICBSII) (pp. 1-4). IEEE.

[5] Dimauro, G., Di Nicola, V., Bevilacqua, V., Caivano, D., &Girardi, F. (2017). Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. IEEE Access, 5, 22199-22208.

[6] Dastgheib, Z. A., Lithgow, B., &Moussavi, Z. (2012). Diagnosis of Parkinson's disease using electrovestibulography. Medical & biological engineering & computing, 50(5), 483-491. 60

[7] Dexter, D. T., & Jenner, P. (2013). Parkinson disease: from pathology to molecular disease mechanisms. Free Radical Biology and Medicine, 62, 132-144

[8] Mirarchi, D., Vizza, P., Tradigo, G., Lombardo, N., Arabia, G., &Veltri, P. (2017, August). Signal Analysis for Voice Evaluation in Parkinson's Disease. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 530-535). IEEE.

[9] Orozco-Arroyave, J. R., Belalcazar-Bolanos, E. A., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Skodda, S., Rusz, J., ... &Nöth, E. (2015). Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases. IEEE journal of biomedical and health informatics, 19(6), 1820-1828.