International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.01, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Statistical Analysis of Parkinson's Disease Using Principle Component Analysis and Artificial Neural Network: A Critical Review

[1]Ankita Tyagi,[2]Aditya Kishore Saxena,[3]Sabranti Maji,
[4]Nafees Akhter Farooqui

[1]Research Scholar,[2]Assistant Professor,[3]Assistant Professor,[4] Research Scholar
[1]Deptt of Computer Applications,
[1]DIT University, Dehradun,(India)

ABSTRACT

Parkinson disease (PD) is a protracted neurological disorder; it is a voluntary movement prominently characterized by slowness (i.e. bradykinesia). There are various reasons for Parkinson disease that are slowly progressive in the Parkinson's patient. In this paper, our main goal is to discriminate between healthy people and people with Parkinson's disease. Parkinson's disease is a condition in which parts of the brain become progressively damaged over many years.Principal Component Analysis (PCA) method is used to discriminate between the healthy people and Parkinson's disease affected people. PCA method has been successfully applied in the context of disease diagnosis, transition prediction, and treatment prognosis, using both structural and functional neuroimaging data. Positron emission tomography (PET) and single photon emission tomography (SPECT) have been used to monitoring the change in dopaminergic function in Parkinson's disease and these methods show decreasing of dopamine neurons in the striatum of the brain.There are many Statistical methods that are used to measure the disease and their detection at an early age. The combination of Principles component analysis (PCA) and Artificial Neural Networks (ANN) distinguishes healthily and PD patients using speech signals data set that gives classification rates as high and significant. PCA is a method to estimate the relationship between data points.

Keywords: Artificial Neural Network, Dopamine Neurons, ParkinsonDisease Detection, Principal Component Analysis, Positron Emission Tomography.

I. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder. Parkinson's primarily affects neurons in area of the brain called the substantia nigra. There are different types of symptoms of Parkinson's disease patients that accounts for a variety of motor and non-motor deficits which are the result of the loss of dopamine-producing brain cells. Disease onset was believed to be due to dopaminergic neuron reduction in the basal ganglia, but it is now recognized that PD is also characterized by the degeneration of numerous non-dopaminergic pathways. The four main symptoms of Parkinson's are tremor, which means shaking or trembling. Tremor may affect your hands, arms, legs, or head: stiff muscles, slow movement and problems with balance or

IJARSE
ISSN: 2319-8354

walking. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions [1-3]. Positron emission tomography (PET) and single photon emission tomography (SPECT) have been used to monetarize changing in dopaminergic function in PD and these methods show decreasing of dopamine neurons in the striatum of brain. Parkinson's disease measurement tool was using the Unified Parkinson Disease Rating Scale (UPDRS) to assist the diagnose [4- 7]. The disease can be difficult to diagnose accurately, particularly in the early stages of the disease when symptoms resemble other medical conditions, and misdiagnosis occurs occasionally. Current research programs are trying to respond how the disease progresses and to develop new drug therapies. Scientists looking for the cause of PD continue to search for possible environmental factors, such as toxins, that may trigger the disorder, and study genetic factors to determine how defective genes play a role. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD, and the prognosis depends on the patient's age and symptoms. The diagnosis is based on the medical history and neurological examination conducted by interviewing and observing the patient. Brain scans or laboratory tests may be used to help doctors exclude other medical conditions that produce symptoms like those of Parkinson's disease. This paper deals with the application of Neural Networks with back propagation together with Principal Component Analysis to a medical dataset concerning PD with the aim of automatically classify patients in PD or non-PD depending on their medical attributes. To test the performance andefficiency of the proposed method, the classification accuracy, sensitivity and specificity were used.

II.METHODS

A. Parkinson data set

Principal component analysis is an approach to factor analysis that considers the total variance in the data, which is unlike common factor analysis, and transforms the original variables into a smaller set of linear combinations. When we conduct principal component analysis, we get a well versed with terms such as standard deviations and eigenvalues. The eigenvalues refer to the total variance explained by each factor. The standard deviation measures the variability of the data. The task of principal component analysis is to identify the patterns in the data and to direct the data by highlighting their similarities and differences. While reviewing the papers we found the dataset consists of 180 sustained vowels phonation's from 45 female and male subjects, of which 19 were diagnosed with PD. The time since diagnoses ranged from 0 to 30 years and the ages of the subject ranged from 30 to 75 years (mean 57.5, standard deviation). The attributes information consists of the following details:

TABLE 1. List of Extracted Features and Their description

Features	Description
MDVP: FO(Hz)	Average vocal Fundamental Frequency
MDVP: FHI (Hz)	Maximum vocal Fundamental Frequency
MDVP: Flo(Hz)	Minimum vocal Fundamental Frequency
Jitter (%), Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP	Several measures of variation in fundamental frequency

Shimmer, 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
RPDE, D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent

Some prominent features are explained below:
jitter (%): it is the average absolute difference between consecutive periods of fundamental frequency, divided by the average period (expressed as a percentage)

$$Jitter(\%) = \frac{\frac{1}{N}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

Where Ti is the period of fundamental frequencies of window Number “i” and N is the total number of windows. Jitter (ABS): It is the average absolute difference between Consecutive periods of fundamental frequency, inMicroseconds (µs)

$$Jitter(ABS) = \frac{1}{N}\sum_{i=1}^{N-1}|T_i - T_{i-1}|$$

jitter (rap): the relative average perturbation, the average absolute difference between a period of fundamental frequency and the average of it and its two neighbors, divided by the average period. jitter (ddp): the average absolute difference between consecutive differences between consecutive periods, divided by the average period shimmer: the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude

$$Shimmer = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$$

Shimmer (APQ3): This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude. Shimmer (APQ5): This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closes neighbors, divided by the average amplitude. Shimmer (DDA): This is the average absolute difference between consecutive differences between the amplitudes of consecutive periods

B.Feature Selection-Principal Component Analysis

Principal Components Analysis (PCA) is a useful statistical technique that has found application in fields such as face recognition and image compression and is a common technique for finding patterns in data of high dimension. PCA is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, principal component analysis is used for data reduction, by compressing the data or mapping the data into a lower dimensional space. In general, reduction of dimensionality will be accompanied by a loss of some of the information. Thus, the main goal in dimensionality reduction is to preserve as much of the relevant information as possible.ANN is more efficient PCA method used before in the classification of data. PCA method is used for data reduction of feature vector [12,13]. It describes the data set in the sense of its variance. Each principal component identifies a percentage of the total variance of a data set and elaborates loadings or weights that each variate contributes to this variance. One of the main advantage of PCA technique is to find patterns inthe data, and data has been compressed by reducing the number of dimensions, without much loss of information [8,9]. Principal component analysis is useful if we obtained data on many variables (possibly many variables), and we believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another. Because of this redundancy, it should be possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables.In this paper, PCA is used as characterization and feature selection of medical data set. Principal components are calculated using eigenvectors and eigenvalues of covariance matrixes or correlation matrix.

C.Artificial Neural Networks

Brain is highly complex, nonlinear and capable to perform different computations (e.g., pattern recognition, perception, and motor control) many times faster than the fastest digital computer in existence today. A neural network is an artificial representation of the human brain that tries to simulate its information processing. It is an interconnected group of artificial neurons which may share some properties of biological neural networks. A neural network derives its computing power through its parallel distributed structure and its ability to learn and generalize. These two capabilities make it possible for neural networks to solve complex problems by decomposing problem into many relatively simple tasks, and neural networks are assigned a subset of the tasks that match their inherent capabilities. It is important to recognize, however, that we have a long way to go (if ever) before we can build a computer architecture that mimics a human brain. A neural network consists of a certain number of layers, and each layer contains a certain number of unit there is an input layer, an output layer, and one or more hidden layers between the input and the output layer. In general, we may identify three different classes of network architectures [10].

Recurrent networks unlike others, has at least one feedback loop. For example, a recurrent network may consist of a single layer of neurons with each neuron feeding its output signal back to the inputs of all the other neurons.

The presence of the feedback loops has impact on the learning capability of the network and on its performance. Single-Layer Feedforward networks is the simplest form of a layered network, where we have an input layer of source nodes that projects onto an output layer of neurons, but not vice versa. In other words, this network is strictly a feedforward, and single refers to one output layer of computation nodes. Multilayer Feedforward Networks has one or more hidden layers. The function of hidden neurons is to intervene between the external input and the network output in some useful manner. In this kind of networks, there are no connections from any of the units to the inputs of the previous layers (no feedback information) or to other units in the same layer, nor to unit's more than one layer ahead. Every unit only acts as an input to the immediate next layer. Obviously, this class of networks is easier to analyze theoretically than other general topologies because their outputs can be represented with explicit functions of the inputs and the weights. The architectural graph in Figure1 illustrates the layout of a multilayer feedforward neural network for the case of a two-hidden layer.

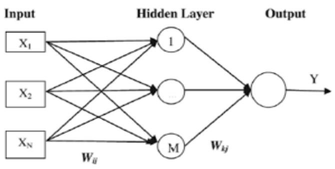

Fig1. Feedforward neural network with two hidden layers

III.COMMITTEE MACHINES

When a task is too complex, the best thing to do is to divide it into smaller and simpler tasks and combine solutions to solve the whole task. In supervised learning, computational simplicity is achieved by distributing the learning task among many experts, which in turn divides the input space into a set of subspaces. The combination of experts is said to constitute a committee machine. Basically, it fuses knowledge acquired by experts to arrive at an overall decision that is supposedly superior to that attainable by any one of them acting alone. So, committee machines are expected to produce better results than using any expert individually, because they combine knowledge from several experts to reach a decision. Committee machines can be built in two different ways, using static and dynamic structures. [10]. In dynamic structures input data is involved in each expert output combination mechanism to generate the global output. This category includes 2 methods: mixture of experts, where answers from experts are nonlinearly linked by only one gating network; hierarchical mixture of experts, where answers from experts are nonlinearly linked by several gating networks arranged in a hierarchical fashion. In static structures, combination mechanism between experts does not depend on input data. This category can also be classified in: ensemble averaging, where global output is a result of linear combination of each specialist outputs; boosting, where a weak learning algorithm can learn how to reach a higher accuracy. In boosting machine, the experts are trained on data sets with entirely different distributions. Boosting can be implemented in three different ways: boosting by filtering, subsampling and reweighting. In this paper we used boosting by filtering technique.

IJARSE
ISSN: 2319-8354

D. Boosting by filtering

In boosting by filtering, the committee machine consists of three experts, arbitrarily labelled as first, second and third. The three experts are individually trained as follows: the first expert is trained on a set consisting of N1 examples. The trained first expert is used to filter another set of examples by proceeding in the following manner: flip a fair coin to simulate a random guess. If the result is heads, pass new patterns through the first expert and discard correctly classified patterns until a pattern is misclassified. That misclassified pattern is added to the training set for the second expert. If the result is tails, do the opposite. Specifically, pass new patterns through the first expert and discard incorrectly classified patterns until a pattern is classified correctly. That correctly classified pattern is added to the training set for the second expert. Continue this process until total N1 examples has been filtered by the first expert. This set of filtered examples constitutes the training set for the second expert. Once the second expert has beentrained in the usual way, a third training set is formed for the third expert by proceeding in the following manner: pass a new pattern through both the first and second experts. If the two experts agree in their decision, discard that pattern. If, on the other hand, they disagree, the pattern is added to the training set for the third expert. Continue with this process until a total of N1 examples have been filtered jointly by the first and second experts. This set of jointly filtered examples constitutes the training set for the third expert. The third expert is then trained in the usual way, and the training of the entire committee machine is thereby completed. The three-point filtering procedure is illustrated in Fig2.

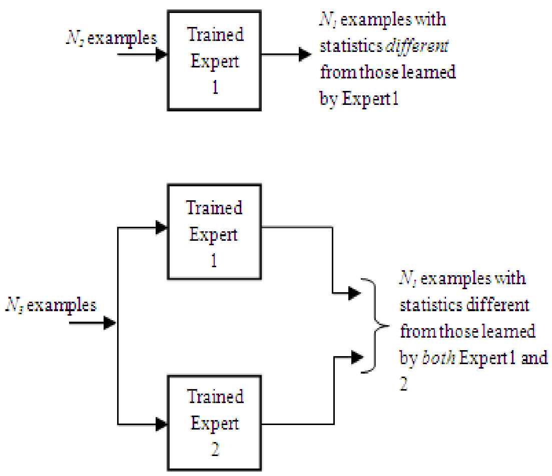

Fig2. Illustration of boosting by filtering, (a) Filtering of examples performed by Expert 1, (b) Filtering of examples performed by Expert 2 and 3

Let N2 denote the number of examples that must be filtered by the first expert to obtain the training set of N1 examples for the second expert. Let N3 denote the number of examples that must be jointly filtered by the first and second experts to obtain the training set of N1 examples for the third expert. With N1 examples also needed to train the first expert, the total size of data needed to train the entire committee machine is N4=N1+N2+N3. However, the computational cost is based on 3N1 examples because N1 is the number of examples used to train each of the three experts. We may therefore say that the boosting algorithm described here is indeed smart in the sense that the committee machine requires a large set of examples for its operation, but only a subset of that data set is used to perform the actual training. To evaluate the performance of the committee machine on test patterns, simple voting scheme was used in this paper. If the first and second experts in the committee machine agree in their respective decision, that class label is used. Otherwise, the class label discovered by the third expert is used. Before the classification of Parkinson dataset, PCA is used to reduce the dimensionality of the input. After using PCA, the input dataset was randomly partitioned into train and test dataset. For neural networks classifier, the following adjustments were carried out: the back-propagation learning algorithm has been used in the feed-forward, four hidden layer neural network [11]. A tangent sigmoid transfer function has been used for both the hidden layers and the output layer

TABLE 2. Samples Used at The Stage of Testing Processes

Patients	X1	X2	X3	Test results
sickness	120	75	57	84
healthy	40	24	14	26

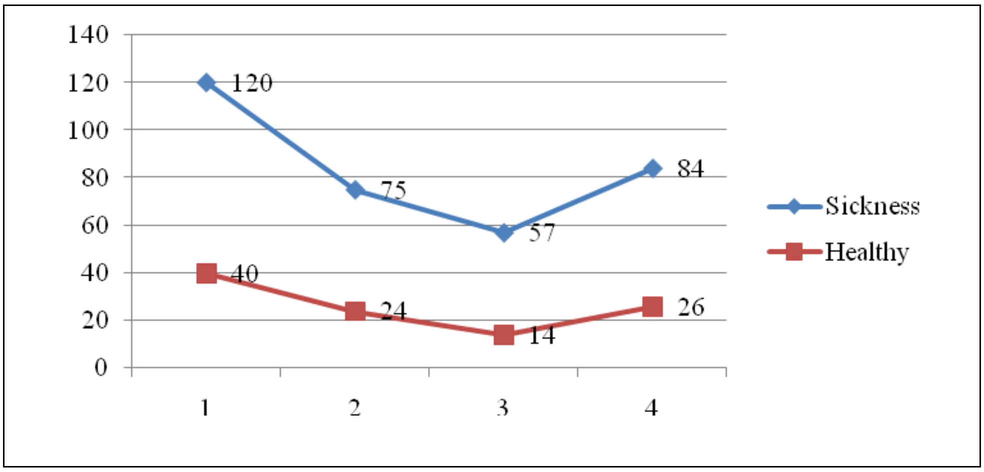

Fig3. Comparative analysis between healthy and PD affected persons

IV.CONCLUSION

In this research we aimed to discriminate between healthy and Parkinson’s disease people.In this paper, a method based on PCA and artificial neural network used for diagnosing and classification on healthy people and people with Parkinson was investigated. We have proposed a system for classification of Parkinson’s disease based on a neural network. Artificial Neural Network can be looked upon as a parallel computing system comprised of some number of rather simple processing units (neurons) and their interconnections.The focus was on static structure of committee machine, known as boosting by filtering, and the total seven committee machines were used. In future we use different machine learning approaches such as MLP (multi-layer perceptron) used for classifications and prediction of (PD) by adapting radial basis function neural network(RBFNN) to detect at early stage of the disease and for research purposes in the PD detection problem, is another possible future work.

V.ACKNOWLEDGMENT

The author thanks the DIT University for providing the research grant to support this research work. The corresponding author wishes to thank Prof K.K. Raina and Prof S.K. Gupta for the great cooperation and motivation for this research.

REFERENCES

[1] Parkinson's UK, http://www.parkinsons.org.uk/about_parkinsons/what_is_parkinsons.asp

[2] National Parkinson Foundation, http://www.parkinson.org/parkinson-sdisease.

[3] Webmd, http://www.webmd.com/parkinsonsdisease/tc/parkinsonsdisease

[4] Schrag, A., Ben-Schlomo, Y., Quinn, N (2002). How valid is the clinical diagnosis of Parkinson’s disease in the community? Journal of Neurology, Neurosurgery Psychiatry.73.529-535.

[5] Booij, J., Tissingh, G., Winogrodzka, A., Royen, E A (1999). Imaging of the dopaminergic neurotransmission system using singlephoton

[6] Emission tomography and positron emission tomography in patients with Parkinsonism, Eur. J. Nucl. Med. 26. 171–82.

[7] Martínez-Martín, P. et all “Unified Parkinson’s disease rating scale characteristics and structure” Movement Disorders, vol9(1), pp.76–83,2004

[8] Smith, L.I. (2002). A tutorial on PrincipalComponents Analysis, http://kybele.psych.cornell.edu/~edelman/Psych-465-Spring- 2003/PCA-tutorial. Accessed 26 February 2002

[9] Mutihac, R., Van Hulle, M. M (2004). Comparison of Principal Component Analysis and Independent Component Analysis for Blind Source Separation”, Romanian Reports In: Physics, 56 (I) 20-32.

[10] S. Haykin, Neural Networks, A Comprehensive Foundation, Second Edition, Pearson Prentice Hall, 1999.

[11] M. Erne, Neural network-based approach to discriminate healthy people from those with Parkinson's disease, Math Comp Sci Ser, 35:112-116,2008

[12] I. Cohen, Q. Tian X.S. Zhou, T.S. Huang, Feature selection using Principal Feature Analysis, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.8481,(2013, 05,12)

[13] S. lakhina, S. Joseph and B. verma Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection NSL-KDD, J Eng Sci Dgn, 2: 1790-1799, 2010