Neural Network based Technique for Parkinson's Disease Classification using ANOVA as Feature Selection Model

Poornima K.M[1]

2nd yearM.Tech Student, Dept of Computer Science and Engineering, BTL Institute of Technology, Bangalore, India

Smt. T Jayakumari[2]

Assistant Professor, Dept of Computer Science and Engineering, BTL Institute of Technology, Bangalore, India

Abstract: **Parkinson's disease (PD) caused due to the loss of dopamine in the brains thalamic region that effect in involuntary or oscillatory movement in the body. In this study, depicted the analysis performed based on two training test algorithms which are Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) of the Multi-Layer Perceptrons (MLP) with Back-Propagation learning algorithm, that are applied to classify for effective diagnosis of PD with Analysis of Variance (ANOVA) as a feature extraction.**

Keywords - Parkinson's disease (PD), Multilayer Perceptrons (MLPs) Neural Network, Levenberg-Marquardt (LM) , Scaled Conjugate Gradient (SCG).

I. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative affliction after Alzheimer's disease (AD)[1].PD is a progressive neurological disorder qualified by tremor, rigidity, and slowness of movements. It is associated with progressive neuronal loss in the substantia nigra and other brain structures [2]. Non-motor features, such as dementia and dysautonomia, happen frequently, particularly in advanced stages of the disease. Diagnosis depending on the presence of two or more cardinal motor features such as rest tremor, bradykinesia, or rigidity[3].Having so many factors to examine to diagnose PD, specialist generally makes decisions by measuring the current test results of their patients. Furthermore, the previous outcomes made on other patients with a similar condition are also done by them.

These are complex procedures, especially when the number of factors that the specialist has to measure is high (high quantity and variety of these data). For these reasons, PD diagnosis needs experience and highly skilled specialists. Classification systems can help in increasing accuracy and relevance of diagnoses and minimizing potential errors, as well as making the diagnoses more time effectively [4].

This paper depicted experimental analysis of MPLs neural network based on two training test algorithms which are Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) in diagnosing PD. The experiment comprises of two parts .The first part of the experiment, the attribute acted as inputs to MLP without ANOVA as feature Extraction and the second part of the experiment, the significant of ANOVA as feature extraction is measured. The results attained from this paper can be used for other diseases such as breast cancer, heart problem and etc. Much cost can be reduced by applying this technique because the analysis process is fully based software Approach by utilizing real database of the disease.

II. RELATED WORK

There are several techniques and analysis that have been done to diagnose Parkinson's disease (PD), for example artificial neural network classifier for the diagnosis of Parkinson's disease using [99mTc] TROD AT-1 and SPECT.Besides, there are another diseases have been detected using Artificial Neural Network. The brief descriptions on the detection disease using several methods are shown below:

"Artificial Neural Network Classifier for the Diagnosis of Parkinson's disease" using [99mtc] TRODAT-1 and SPECT According to Acton PD and Newberg A in their research [5], imaging the dopaminergic neurotransmitter system with positron emission tomography (PET) or single photon emission tomography (SPECT) is a powerful tool for diagnosis of Parkinson's disease. It has been hypothesized that an artificial neural network (ANN), which can mimic the pattern recognition skills of human observers, may provide similar results. A set of patients with PD, and normal healthy control subjects, were studied using the dopamine transporter tracer[99mTc]. TRODAT-1 and SPECT. The sample was comprised of 81 patients (mean age ± SD: 63.4 ± 10.4years;age;range:39.0-84.2years)and94healty cont

range: 39.0-84.2 years) and 94 healthy controls (mean age ± SD: 61.8 ± 11.0 years; age range: 40.9-83.3 years). The images were processed to extract the striatum and the striatal pixel values were used as inputs to a three layer ANN. However, by using this technique, it was difficult to interpret precisely what triggers in the images were being detected by the network.

III. DATA SET

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders [6]. The attributes are tabulated as in TABLE I.

ATTRIBUTE	DESCRIPTION
MDVP:Fo(Hz)	Average vocal fundamental Frequency
MDVP:Fhi(Hz)	Maximum vocal fundamental Frequency
MDVP:Flo(Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%) MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP	Several measures 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
STATUS	The status of the patient (1) - Parkinson's Disease (0)-Healthy

Table I : Feature information for Parkinson dataset

The data set is comprises of a range of biomedical voice measurement with 195 samples with 16 attributes where 147 samples were diagnosed with PD. The main objective of study is to recognize healthy people from those with PD using MLP. According to "status" column which is set to 0 for healthy and 1 for PD.

IV. METHODOLOGY

This section will depict the methodology as described in Fig. 1 for classifying PD using MLPs neural network based on two training test algorithms which are LM algorithm and SCG Algorithm. ANOVA is applied as the feature Extraction for the PD data set.

A. Multilayer Perceptrons Neural Network

Multilayer is feed-forward neural networks trained with the standard back-propagation algorithm. It is supervised networks so they need a desired response to be trained [7].
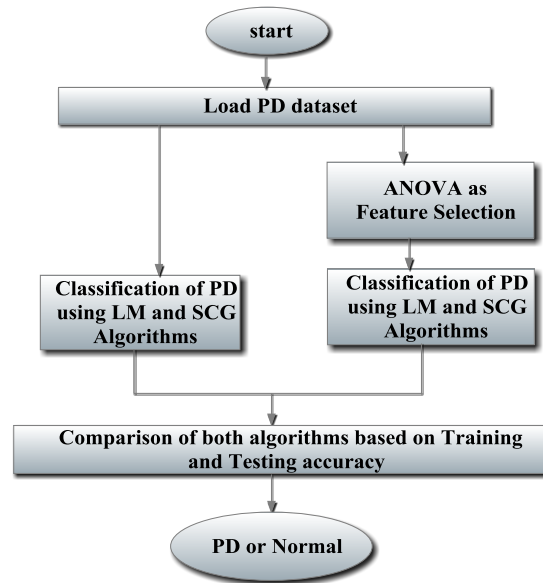

Fig. 1. Overall Block diagram

It learns how to transform input data in to a desirable response, so they are widely applied for pattern classification. With one or two hidden layers, they can estimated almost input-output map. It has been presented to estimate the performance of optimal statistical classifiers in difficult problems. The most popular static network in the multilayer [8]. The multilayer is trained with error correction learning, which is appropriate here because the desired multilayer response is the arteriographic result and as such known. Error correction learning works in the following way from the system response at neuron j at iteration t, $y_j(t)$, and the desired response $d_j(t)$ for given input pattern an instantaneous error $e_j(t)$ is defined by

$$e_j(t) = d_j(t) - y_j(t) \qquad (1)$$

Using the theory of gradient descent learning, each weight in the network can be adapted by correcting the present value of the weight with a term that is proportional to the present input and error at the weight, i.e.

$$w_{jk}(t+1) = w_{jk}(t) + n\delta_j(t)x_k(t) \qquad (2)$$

The $n(t)$ is the learning-rate patameter.The $w_{jk}(t)$ is input neuron j at iteration t.The local error $\delta_j(t)$

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRTS-2015 Conference Proceedings

can be computed as a weighted sum of errors at the internal neurons.

B. ANOVA (Analysis of Variance)

Analysis of Variance is a statistical test used to determine whether three or more data sets (means) are statistically significantly different. Selection of proper ANOVA depends on the data to be examined. First criterion involving selection of ANOVA test represents to distribution of the data sets. Because the data analyzed do not follow normal distribution, nonparametric version of the classical one-way ANOVA should be used. There are two suitable tests implemented in MATLAB that can be used KruskalWallis test and Friedman's test. Both the tests examine the ranks of the data rather than their original numeric values. Ranks are obtained by ordering the data from smallest to largest across all groups, and taking the numeric index of this ordering [9].

Kruskal-Wallis test is a nonparametric test that comparing three or more unpaired groups of data. Moreover, MATLAB implementation of the test allows analysis of data where number of recordings is not equally distributed into individual groups to be examined. The Kruskal-Wallis test measures the hypothesis that all samples come from populations that have the same median, against the alternative that the medians are not all the same. As a result, p-value for the null hypothesis that all samples are drawn from the same population is obtained[10].

ANOVA can be used as feature extraction technique [11], this are the Steps involved in ANOVA is given below:

Step1: Calculation of Total Sum of Squares (SST):

There are three possible sums of squares-between-group some of squares (SSB), within-group or error sum of squares (SSW), and total sum of squares (SST). Total sum of squares can be partitioned into between sum of squares and within sum of squares, representing the variation due to the independent variable and variation due to individual differences in the score respectively:

$$SST = SSB + SSW$$

Sum of squares between-groups examines the differences among the group means by calculating the variation of each mean $(\bar{y})$ around the grand mean $(\bar{\bar{y}})$.

$$SSB = n \sum (\bar{y}_i - \bar{\bar{y}})^2$$

n is the number of observations in each group.

Sum of squares within-groups examines error variation or variation of individual scores around each group mean. This is variation in the scores that is not due to the independent variable.

$$SSW = \sum \sum (y - \bar{y}_i)^2$$

The total sum of squares can be computed by adding the SSB and the SSW.

Step2: Calculation of Degree of Freedom (DF)

Degree of freedom (DF) represents the number of independent values in a calculation, minus the number of estimated parameter

DF within the group= $(K - N)$

DF between the group=$(K - 1)$

$$DF = \frac{DF_{WITHIN}}{DF_{BETWEEN}} = (N - 1)$$

Where N is number of sample, K is number of groups.

Step3: Calculation of Mean Squares Total (MST)

Mean squares (MS) are estimates of variance across groups[12]. Mean squares are used in analysis of variance and are calculated as a sum of squares divided by its appropriate degrees of freedom. Let N equal the total number of samples in a survey, and K the number of groups, then the:

Mean Square Total (MST) is an estimate of total variance against the grand mean (mean of all samples).

$$MST = \frac{SST}{N - 1}$$

Square Between groups compare the means of groups to the grand mean:

$MSB = \frac{SSB}{K-1}$ Mean Square within (MSW) groups calculates the variance within each individual group:

$$MSW = \frac{SSW}{N - K}$$

Step4: Calculation of F statistic or also known as F ratio:

A value used in determining whether the difference between two variables is statistically significant or stable. A larger variance is divided by a smaller variance, both of which are the results of analysis of variance procedures. Mean Square Between (MSB) and Mean Square Within (MSW) are used to calculate the F-ratio:

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRTS-2015 Conference Proceedings

$$F\ ratio = \frac{MSB}{MSW}$$

Step5: Calculation of P-value

To find the P- value from an F distribution you must know the numerator (MSW) and denominator (MSB) degrees of freedom, along with the significance level. P-value has df1 and df2 degrees of freedom, where df1 is the numerator degrees of freedom equal to $K-1$ and df2 is the denominator degrees of freedom equal to $N-K$.

Step 6: Decision Rule

Reject the null hypothesis if: F (observed value) > P-value .It means that there is a significant difference between the groups. Fail to reject null hypothesis means that there is no difference between the groups.

V. RESULTS AND DISCUSSION

As a proposed, ANOVA as feature Extraction is Implemented in MATLAB R2013b. Fig 2 showed the result of ANOVA table Analysis.

Fig. 2. ANOVA Table analysis

After using ANOVA, the attribute reduced to 4 features instead of 16 features with statistic value, F of 1068.65 as Shown in Fig. 2 and Fig. 3.The selected features are MDVP:FO(HZ), MDVP:Fhi(HZ), HNR, MDVP:Flo(HZ).The highest value of F indicates that the ANOVA analysis is effective.
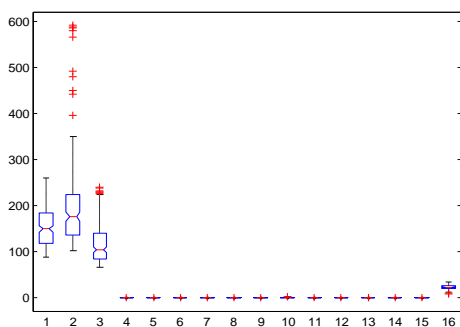

Fig. 3. ANOVA Analysis

The Fig. 4 shows the multiple-comparison of the mean. The Group 6 mean is highlighted and the comparison interval is in blue. Because the comparison intervals for the other 4 groups do not intersect with the intervals for the group 6 mean, they

are highlighted in red. This lack of interaction indicates that both means are different than group 6 mean. Select other group means to confirm that all group means are significantly different from each other.
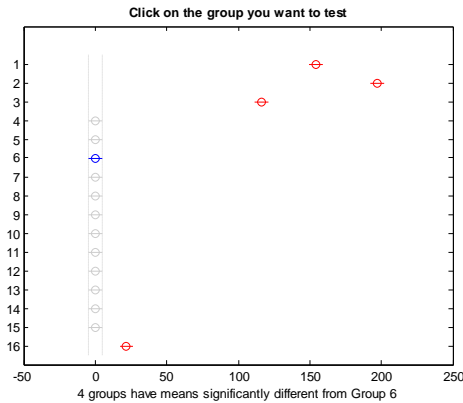

Fig. 4. Multi-comparison analysis

LM and SCG were developed using the MATLAB Neural Network Toolbox. In this section the experimental results obtained using LM and SCG algorithms in terms of Average Training Accuracy, Average Testing Accuracy, Average Iterations and Average MSE is discussed. Prior to training, the dataset was rescaled between -1 and 1 before divided into 50:20:30 as ratio for training: validation: testing.
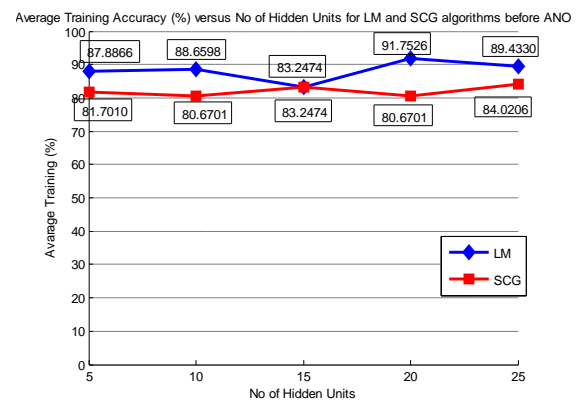

Fig. 5. Average Training Accuracy (%) versus No of Hidden Units for LM and SCG algorithms before ANOVA

Fig. 5 and Fig. 6 showed the results for the Average training accuracy for LM and SCG for both analysis; with and without ANOVA. It can be seen that the average training accuracy using LM algorithm for both analysis are higher at almost all hidden unit as compared to SCG algorithm. As for LM algorithm, highest accuracy is obtained at hidden units of 20 & 10 for both Analyses with and without ANOVA as feature Extraction that is 91.75% and 90.20% whilst SCG algorithm achieved the accuracy rate 84.02% at 25 hidden units. However, via ANOVA, SCG increased the accuracy rate to 88.14% at hidden unit 10.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
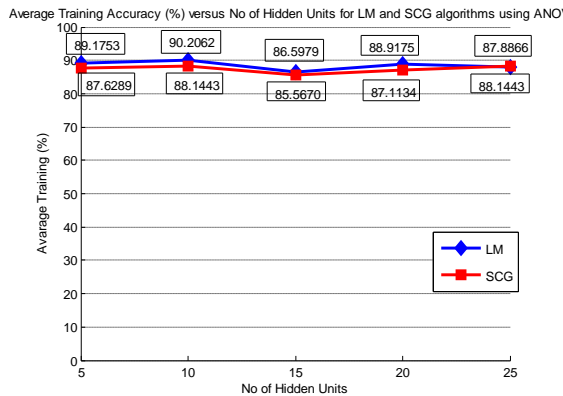**NCRTS-2015 Conference Proceedings**

Fig. 6. Average Training Accuracy (%) versus No of Hidden Units for LM and SCG algorithms after ANOVA
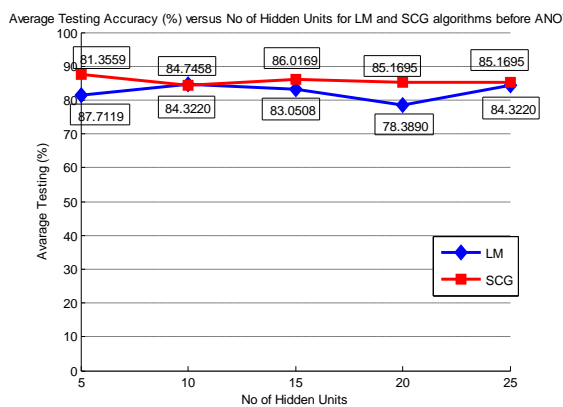

Fig. 7. Average Testing Accuracy (%) versus No of Hidden Units for LM and SCG algorithms before ANOVA
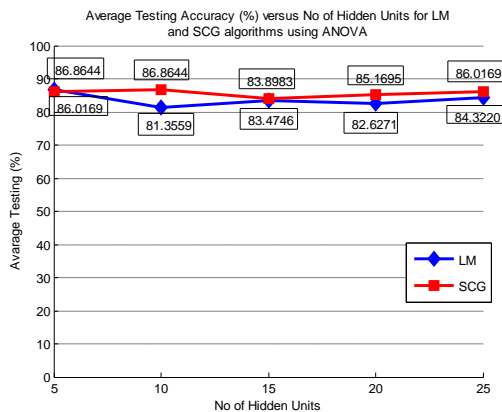

Fig. 8. Average Testing Accuracy (%) versus No of Hidden Units for LM and SCG algorithms after ANOVA

In Fig. 7 and Fig. 8 the average testing accuracy for both LM and SCG is showed. Using LM algorithm, achieved testing accuracy rate at the hidden units and are ranged between 78.38% to 84.74% while, for SCG are between 84.32% to 87.71% With ANOVA as feature selection, LM algorithm achieved testing accuracy rate at hidden unit with ranged from 81.35% to 86.86% while for SCG algorithm are ranged from 83.89% to 86.86% by using ANOVA. This confirmed that for both training and testing phase, LM performed better than SCG.

VI. CONCLUSION

As a conclusion, MLPs can be used to classify the Parkinson's disease. The experiment shows that feature selection helps to increase computational efficiency whilst improving classification accuracy Further, based on both training algorithm Measured, the classification of PD using LM algorithm of MLPs training achieved higher classification rate as compared to SCG algorithm. This is confirmed from the accuracy rate achieved as well as lower MSE obtained. Further, with ANOVA as feature selection the PD dataset is reduced to 4 features instead of 16 features that showed more than 75% reduction in dataset with above 90% accuracy rate achieved for LM while SCG achieved above 80% accuracy.

References

[1] Norlinah Mohamed Ibrahim, "Misconceptions about Parkinson's Disease", Neurology Unit, Pusat Perubatan University Kebangsaan Malaysia, November 2009.

[2] AJ Hughes, SE Daniel, L. Kilford, and AJ Lees. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. British Medical Journal, 55(3) 181184, 1992.

[3] "Deep-Brain Stimulator and Control of Parkinson's disease"proc. SPIE 5389,smart Structure And Metrials 2004;

[4] Anchana Khemphila, Veera Boonjing, "Parkinsons Disease Classification using Neural Network and Feature selection", World Academy of Science & Tech, 64, 2012.

[5] J. Tebelskis, "Speech recognition using neural networks," PhD, Carnegie Mellon University, Pittsburgh, Pennysylvania, 1995.

[6] UCI Machine Learning Repository- Center for Machine Learning and Intelligent System, http://archive.ics.uci.edu.

[7] Data Mining Techniques For Marketing,Sales, and Customer Support. John Wiley & Sons,Inc.

[8] Active media Innovation Sdn. Bhd, "Applying Neural Network with MATLAB.

[9] Lukas Zoubek, "Introduction to Educational Data Mining Using MATLAB", Department of Information and Communication Technologies, Pedagogical Faculty, University of Ostrava.

[10] Statistics Toolbox User's Guide. (September 2009),Available:http://www.mathworks.com/access/helpdesk/help/pdf_doc/stats/stats

[11] http://www.upa.pdx.edu/IOA/newsom/da1/ho_ANOVA

[12] http://www.chegg.com/homeworkhelp/definitions/mean-squares-31