Classification of Parkinson's Disease Based on Multilayer Perceptrons Neural Network

Zahari Abu Bakar, Nooritawati Md Tahir, Ihsan M. Yassin
Faculty of Electrical Engineering
Universiti Teknologi MARA
40450 Shah Alam, Selangor
Malaysia
xaharie84@gmail.com

Abstract – **Parkinson's disease (PD) is the second commonest late life neurodegenerative disease after Alzheimer's disease. It is prevalent throughout the world and predominantly affects patients above 60 years old. It is caused by progressive degeneration of dopamine containing cells (neurons) within the deep structures of the brain called the basal ganglia and substantia nigra. Therefore, accurate prediction of PD need to be done in order to assist medical or bio-informatics practitioners for initial diagnose of PD based on variety of test results. This paper described the analysis conducted based on two training algorithms namely Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) of Multilayer Perceptrons (MLPs) Neural Network in diagnosing PD. The dataset information of this project has been taken form the Parkinson Disease Data Set. Results attained confirmed that the LM performed well with accuracy rate of 92.95% while SCG obtained 78.21% accuracy.**

Keywords – **Parkinson's disease (PD), Multilayer Perceptrons (MLPs) Neural Network, Leverberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG)**

I. INTRODUCTION

PD is caused by progressive degeneration of dopamine containing cells (neurons) within the deep structures of the brain called the basal ganglia and substantia nigra [1]. Both motor and non-motor symptoms exist in PD, although the cardinal features and the diagnosis of PD are based purely on the presence of motor symptoms. Parkinson's disease (PD) is a condition of the nervous system which causes the muscles to become stiff and the body to shake. According to [1], Head of the Neurology Unit at Universiti Kebangsaan Malaysia Medical Centre (PPUKM), PD symptoms slowly creep up in patients, mostly over age 60, and this is often misinterpreted as a part of normal ageing but PD also affects patients below age 45. It is estimated that more than 10,000 people in Malaysia are suffered from PD, and the number is increasing each year.

Symptoms in PD can be divided into motor and non-motor. However, for this project, the data set used is based on the motor symptoms [1]. Typically, motor symptoms include tremor of the hands and/or legs and slowness of movement. Tremor in PD is also known as "pill-rolling" tremor as it has the characteristic appearance of someone rolling a pill in his hands [1] [3]. It usually occurs when the patient is resting and become more pronounced with mental tasks or anxiety. The initial symptoms of PD can be detected among the patient when hand-writing become more progressively smaller (micrographia) and patient may lose dexterity in carrying out daily tasks. Besides, the voice may become low-pitched and monotonous, with excessive drooling of saliva. The other symptoms are the gait may be affected where resulting in strides with lack of arm swing and giving a shuffling appearance when walking.

This paper described experimental analysis of artificial neural network based on two training algorithms specifically Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) in diagnosing PD. The dataset information of this project has been taken form the Parkinson Disease Data Set (*http://archive.ics.uci.edu/ml/datasets/Parkinsons*). The data set is composed of a range of biomedical voice measurement with 195 samples with 16 attributes where 147 samples were diagnosed with PD. The main aim of is study is to discriminate healthy people from those with PD using MLP.

II. THEORETICAL BACKGROUND

A. Neural Network

A neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaption to, or learning from, a set of training patterns [2]. Two training methods utilized in this study will be elaborated.

B. The Levenberg-Marquardt Algorithm

The LM algorithm [8] attempts to solve a sum of squares of nonlinear functions minimization problem and the application of

LM optimization to train ANNs was introduced in [9]. When searching for the minimum on the error surface, a good learning rule will logically take larger steps in flat areas to skip plateaus quickly, and takes smaller steps when it encounters a large gradient to avoid overstepping the local minima. The LM does this by combining curvature as well as gradient information based on two update rules, the Vanilla Gradient Descent (VGD), and the Gauss-Newton (GN) rule. The LM update rule is such that large steps are taken in the direction of low curvature to skip past plateaus quickly, and small steps taken in the direction of high curvature to slowly converge to minima.

C. The Scaled Conjugate Gradient Algorithm

SCG is a supervised learning algorithm which has been show to handle large-scale problems effectively [11]. Similar to the LM algorithm, it utilizes second order information (curvature) from the neural network, but has modest memory requirements due to inexpensive calculations of the gradient information [11]. Detail information of SCG algorithm can be referred in [11].

III. METHODOLOGY

This section will describe the methodology as depicted in Fig 1 for classifying PD using MLPs neural network based on two training algorithms that are LM algorithm and SCG algorithm. The PD dataset comprised of biomedical voice measurements with 195 samples with 16 attributes where 147 samples were diagnosed with PD. The main aim of the data is to discriminate healthy people from those with PD, according to target set at -1 for healthy and 1 for PD. The attributes are tabulated as in Table 1.

TABLE 1
FEATURE INFORMATION FOR PARKINSON DATASET

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

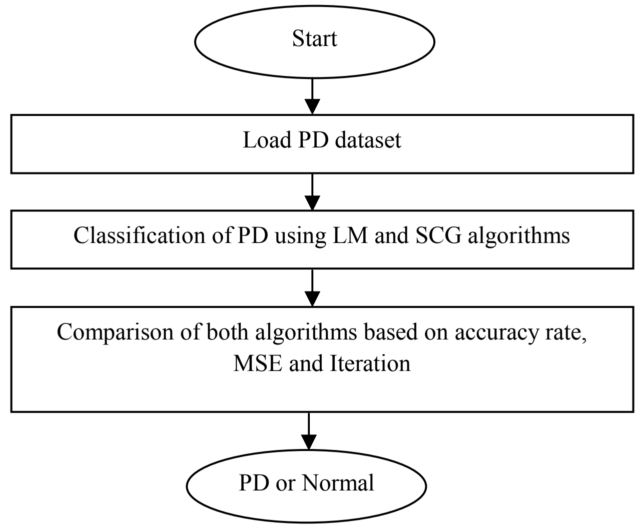

Fig. 1 Overall block diagram of PD classification using MLPs with LM and SCG algorithms

IV. RESULT AND DISCUSSION

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. 2 showed the results for the average training accuracy for LM and SCG. It can be seen that the average training accuracy using LM algorithm are higher at every hidden units as compared to SCG algorithm. As for LM algorithm, highest accuracy is achieved at hidden units of 25 with 97.86% classification rate while SCG algorithm attained the best accuracy of only 79.06% at 10 hidden units.

Further, in Fig. 3 the average testing accuracy for both LM and SCG is depicted. It can be seen from the figure that using LM algorithm, again the testing accuracy is higher than SCG at all hidden units and are ranged between 72% to 93% whilst for SCG are between 67% to 79%. This confirmed that for both training and testing phase, LM performed higher than SCG.

Next, the average MSE versus number of hidden units using LM and SCG training algorithms are shown in Fig. 4. The smallest value of MSE indicated that the residuals are small which means that the particular MLPs had fitted the data very well. It is observed the minimum MSE occurred at hidden unit of 25 for LM training algorithm and hidden unit of 10 of SCG training algorithm.

Fig. 2 Average Training Accuracy versus Number of Hidden Units using LM and SCG training algorithms

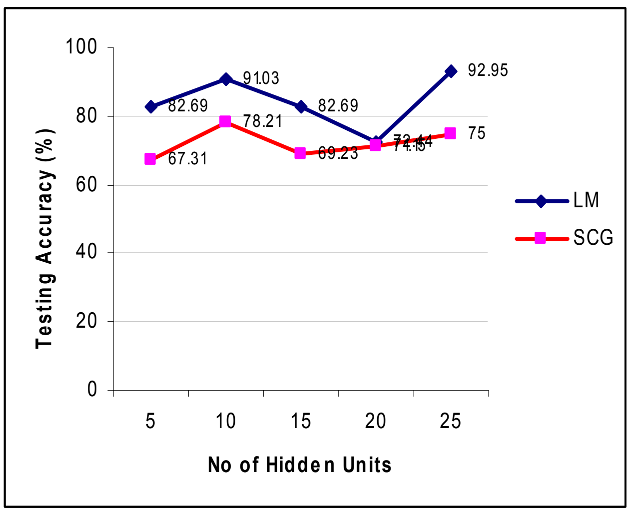

Fig. 3 Average Testing Accuracy versus Number of Hidden Units using LM and SCG training

It is also observed that the best training accuracy and testing accuracy for LM algorithm occurred at hidden unit of 25 as compared to other hidden units with 97.86% for training phase and 92.96% for testing stage. The smallest value of MSE (0.03552) also occurred at this point which means the MLPs fitted the data very well. Then, for SCG algorithm the best values for training accuracy, testing accuracy and MSE occurred at hidden point of 10 with values as 79.06%, 78.21% and 0.3872 respectively.

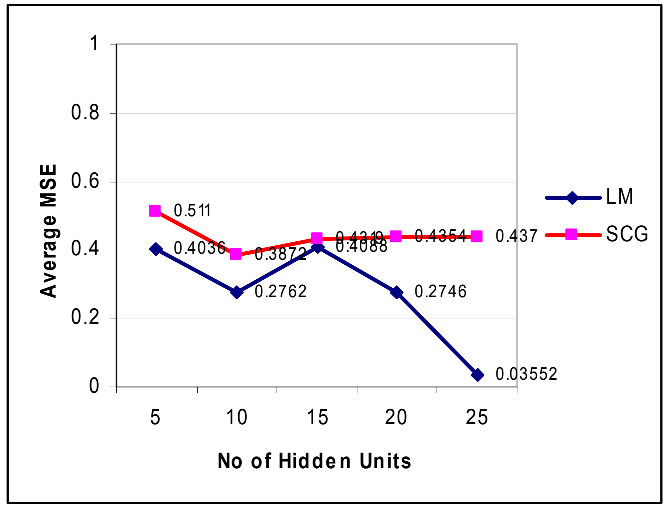

Fig. 4 Average MSE versus Number of Hidden Units using LM and SCG training algorithms

In addition, Fig. 5 showed the average training iterations versus number of hidden units using both algorithms. It is again observed from Fig. 5, at hidden unit of 25 of LM algorithm and 10 of SCG algorithm, are the points with maximum iterations achieved.

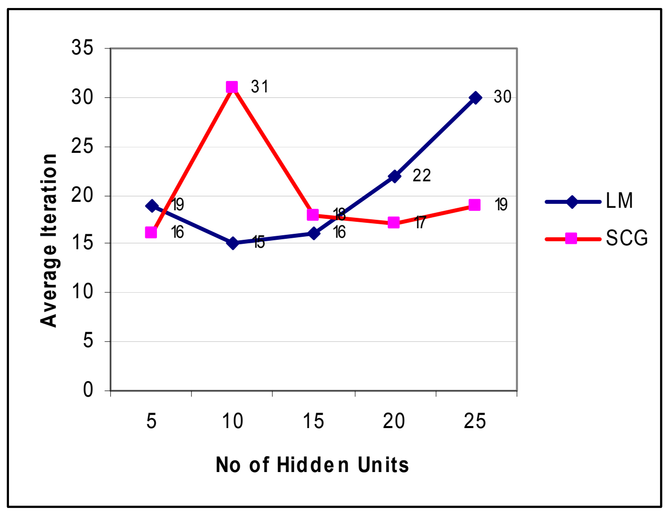

Fig. 5 Average Iteration versus Number of Hidden Units using LM and SCG training algorithms

V. CONCLUSIONS

As a conclusion, it is confirmed that MLP can be employed to classify PD dataset. Further, based on both training algorithm evaluated, the classification of PD using LM algorithm of MLPs training attained higher classification rate as compared to SCG algorithm. This is proven from the accuracy rate achieved as well as the lower MSE obtained. These initial findings can be used to assist medical practitioners and bio-informatics practitioners for identification of PD using MLPs Neural Network.

REFERENCES

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

[2] Activemedia Innovation Sdn. Bhd, "Applying Neural Network with MATLAB", 2006.

[3] Max A. Little, *Member IEEE*, Patrick E. Mc.Sharry, *Senior Member IEEE*, Eric J. Hunter, Jennifer Spielman, Lorraine O.Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease", *System Analysis, Modelling and Prediction Group*, University of Oxford, UK, September 2008.

[4] V. B. Rao, *C++ neural networks and fuzzy logic*: MTBooks, IDG Books Worldwide, Inc., 1995.

[5] D. Anderson and G. McNeill, "Artificial neural network technology," Rome Laboratory, New York1992.

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

[7] L. Pretchelt, "Early stopping - but when?," *Neural Networks: Trick of the Trade,* vol. 1524, pp. 55-69, 1996.

[8] Weisstein, Eric W. "Levenberg-Marquardt Method", http://mathworld.wolfram.com/Levenberg-MarquardtMethod.html

[9] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. on Neural Networks,* vol. 5, pp. 989 - 993, 1994.

[10] H. Demuth and M. Beale. (2005) MATLAB Neural Network Toolbox v4 User's Guide. *Mathworks Inc.*

[11] M. F. Møller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks,* vol. 6(4), pp. 525–533, 1993.