Enhanced MLP classifier performance using variant of Back Propagation on Modified Cervical Pap smear Data

[1]Dr. K. Usha Rani
Professor, Dept. of Computer Science,
Sri Padmavati Mahila Visvavidyalayam, Tirupati, India
usharanikuruba@yahoo.co.in

[2]Dr. K. Hemalatha
Assistant Professor, Dept. of MCA,
Sri Venkateswara College of Engineering, Tirupati, India
hemalathakulala@gmail.com

Abstract — Multi-Layer Perceptron (MLP), the most popular Neural Network (NN) widely used in Medical applications. Training process is an important pace which can influence the performance of NN. Learning Algorithm plays a vital role in training NN. Back Propagation (BP) algorithm is commonly used supervised Learning algorithm for training MLP. Several variants of BP algorithms are available in Literature. Selection of suitable BP algorithm is significant to enhance the performance of MLP. In this study widely used BP algorithms: Bayesian Regularization (BR), Resilient back propagation (Rprop), Scaled Conjugate Gradient (SCG) and Levenberg-Marquardt (LM) are considered to train the MLP on Modified Cervical Pap smear (MCPS) Data and the results are analyzed. The best BP variant which enhanced the MLP performance is identified.

Keywords - Neural Network; Multi-Layer Perceptron; Back Propagation algorithms; Cervical Cancer data.

I. INTRODUCTION

Neural Network (NN) is a significant Soft Computing technique which has been used widely to solve problems related to Pattern Recognition, Optimization, Classification, etc. An artificial system that can perform the tasks intelligently similar to the performance of human brain can be built using NN. NN acquire knowledge through learning and store that knowledge as synaptic weights like human brain. Because of the NN characteristic ability to learn it is implemented successfully in many research areas particularly in Computer-Aided Disease Diagnosis. Multi-Layer Perceptron (MLP) is the most commonly used NN architecture. Three-layer architecture of MLP consists of Input Layer, Hidden Layer and Output Layer [1].

MLP with single hidden layer has universal approximation capability for complex non-linear function [2] Learning/Training of NN involves the process of finding the values of all weights through which the desired output can be generated based on the input provided. In another sense it is the process of minimization of error between the output of the network and desired output of training observations. Major categories of NN learning are Supervised Learning and Unsupervised Learning. Supervised Learning takes place in the presence of a teacher and during the learning process global information may be necessary. Whereas Unsupervised Learning happens in the absence of the teacher and the learning takes places based on the local information [3].

Back Propagation (BP) algorithm is one of the widely used Supervised Learning algorithms for MLP [4] BP algorithm has the capacity to solve complex problems successfully. The main motive behind the development of BP algorithm is to reduce the overall output error gradually during the learning process of NN. There are many variants of BP algorithm. To identify the best fit BP algorithm, most popular variants of BP with MLP are experimented in this study. The variants of BP used are: Bayesian Regularization (BR), Resilient back propagation (Rprop), Scaled Conjugate Gradient (SCG) and Levenberg-Marquardt (LM). MLP is trained with Cervical Cancer data using the above mentioned learning algorithms and its performance is calculated and evaluated.

II. MATERIALS AND METHODS

This section provides the details about the dataset used and the theoretical background of the methods used in this study.

A. Dataset

In our previous studies [5] Cervical Pap smear images from publicly available Herlev dataset were preprocessed and segmented using Fuzzy Hybrid Method (FHM). The four important features: Size of the Nucleus, Size of the Cytoplasm, Grey level of the Nucleus and Grey level of the Cytoplasm are extracted and a Modified Cervical Pap Smear (MCPS) dataset with 4 features and 917 samples was generated. The generated dataset consist of both normal and abnormal cases samples. The MCPS dataset generated in our previous study is considered for experimentation in this study to train MLP with various BP Learning algorithms for classification purpose. The sample images, which are used to generate the MCPS dataset, are presented in the figure.

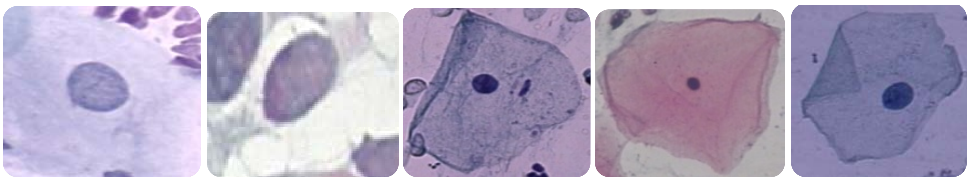

Figure 1. Sample Cervical Pap smear images used to generate MCPS dataset

B. Training MLP using Back Propagation Algorithm:

Back Propagation Algorithm is the standard Supervised Learning method commonly used to train MLP. MLP generally consist of Input layer, one or more Hidden layers and an Output layer. The normalized raw data is provided to the input layer. That data will be distributed to the Hidden layer(s) through the interconnections. Hidden layer(s) process the data using an activation function. The output of MLP network depends on: The activity of Hidden units; Weights between Hidden and Output units [6]. In the training process, MLP is learned using BP algorithm to adapt the connection weights to obtain minimum difference between MLP output and desired output.

BP algorithm updates the weights along the negative gradient direction. Initially all weights are set to small random values during the training using BP algorithm. Later inputs and desired outputs are feed to the MLP. Each trial sample of the training set presented recurrently until the network weights are stabilized [7]. Using the activation function the actual outputs are calculated. Weights are adjusted using (1) given below [8].

$$\boldsymbol{W_{ij}(t+1) = w_{ij}\,(t) + \eta\delta_j x_i'} \qquad \textbf{(1)}$$

where $W_{ij}(t)$ = weights from the hidden or an input node *j* at time *t*

$x_i^{'}$ = output of node *j*

η = gain term

δ_j = error term

C. Variants of BP

The BP algorithm was investigated many times with minor variations to improve the efficiency of MLP. Gradient Descent (GD), Scaled Conjugate Gradient (SCG), Resilient back propagation (Rprop) , Quasi-Newton (QN), One-step Secant , Levenberg- Marquardt (LM) and Bayesian Regularization (BR) are the different variants of BP algorithm. Storage and Computational requirements are different for the above mentioned variants. Rprop and LM are good for Pattern Recognition and QN is suitable for multiface detection [18]. Particular variants are suitable for some types of applications only. It is difficult to find the best BP variant suitable for all applications under all situations [9]. Hence, four most popularly used variants of BP : BR, Rprop, SCG and LM are considered [10] in this study to train MLP.

1. *Bayesian Regularization (BR)*

Bayesian Regularization (BR) is one of the most robust BP algorithms that can reduce lengthy cross-validation. BR framework is based on the probabilistic analysis of network parameters such as weights. In traditional BP algorithm optimal set of weights are elected by minimizing an error function. But BR involves in the probability distribution of network weights. Generally in training process a common performance function is used to calculate the difference between real and predicted data. BR adds an additional term to penalize larger weights to obtain smoother mapping between layers [13]. The performance function of the training process using BR is presented in (2).

$$F = \boldsymbol{\beta E_D}(\mathbf{D}|\boldsymbol{\omega}, \mathbf{M}) + \boldsymbol{\alpha E_W}(\boldsymbol{\omega}|\mathbf{M}) \qquad (2)$$

where α,β are the hyper parameters necessary to estimate the function parameters. E_W (ω|M) is the sum of squares of the weights. If the parameter α is less than the parameter β then the algorithm make small errors. The weights of MLP are updated using Baye's rule [14].

2. Resilient back propagation (Rprop)

Rprop is a first order algorithm. Time and memory are required for Rprop training depends on its parameters. Even though it has more number of adjustable parameters, majority of them can be set by default values. The learning principle of Rprop eliminates the harmful effects arises during the calculation of weights using partial derivatives of magnitude. Instead of magnitude of the derivative, the sign of the magnitude is considered to determine the direction of updated weight [15]. If the weight has the same sign for two successive iterations the updated value of each weight and bias is increased by a factor η. If the sign changes from previous iteration, the updated value will be decreased by η.

3. Scaled Conjugate Gradient (SCG):

Several thousands of weights should be adjusted in the Learning process of NN in realistic applications. Therefore, optimization methods are necessary for the learning of NN for large-scale problems. Conjugate Gradients methods are a class of optimization methods. SCG algorithm belongs to the class of Conjugate Gradients methods. SCG is based on the optimization strategy to choose the search direction and step size using the second order approximation. SCG avoids the time consuming process for the line-search per learning iteration. Hence, SCG is the faster algorithm than any other second order algorithms [16].

4. *Levenberg- Marquardt (LM)*

LM is the combination of Steepest Descent and Gauss-Newton methods. It switches between these two algorithms during the training process. The formula for the weight update is presented in the equation 3.

$$\boldsymbol{W_{k+1} = W_k - (J_K^T + \mu I)^{-1} J_k e_k} \qquad \mathbf{(3)}$$

where W is the weight vector, e is the error vector, J is the Jacobian matrix, I is the identity matrix and μ is the combination coefficient.

If the combination coefficient μ value is nearer to zero then Gauss-Newton algorithm will be applied. When the μ value is large Steepest Descent algorithm will be applied. Therefore LM becomes more powerful training algorithm than conventional training algorithms [17].

Among the variants mentioned in this section, BR algorithm minimizes the linear combination of squared errors and weights. Rprop algorithm works faster in Pattern Recognition problems. SCG uses a step size scaling mechanism which avoids a time consuming line search per learning iteration. Hence, SCG perform well over a wide variety of problems, especially for Network with large number of weights. LM algorithm has fastest convergence and it provides very accurate training. LM performance depends on the number of weights in the network.

III. RESULTS AND DISCUSSION

A. Experimental Setup:

The Constructed MCPS dataset is used for experimentation. The extracted Features: Size of the Nucleus, Size of the Cytoplasm, Gray Level of the Nucleus and Gray Level of the Cytoplasm are given as inputs to the network. The four input nodes are represented as I1, I2, I3 and I4. Figure 2 displays the structure of the MLP constructed for the experimentation in this study.

Four input nodes of Input Layer are connected to the Hidden Layer node with default neurons (i.e.10 neurons).

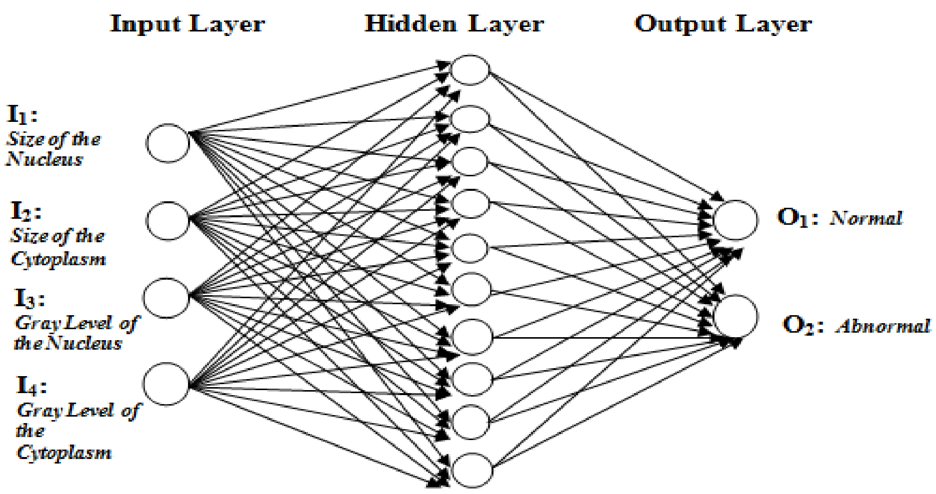

Figure 2. Structure of the MLP constructed in this study

Four input nodes of Input Layer are connected to the Hidden Layer node with default neurons (i.e.10 neurons). Therefore total number of neurons in input layer is 40. The constructed MLP contains one Hidden Layer. Based on the intermediate calculations carried by the hidden layer the output will be generated. As the target classes are Normal and Abnormal, the output layer consists of 2 nodes as O1 and O2. After construction of MLP, it is initialized with random weights. 70% samples of MCPS dataset are used to teach the MLP network and 15% samples are used for the validation purpose. Remaining 15% samples are used to test the accuracy of the output values. Then MLP is trained with BR, Rprop, SCG and LM algorithms respectively. The training of four models of MLP stops when any one of these conditions occurred.

- The maximum number of epochs reached.
- Performance is minimized to the goal.

After the completion of the training these MLP models are validated and tested. For the experimentation Mat lab R2015a tool is used.

B. Results:

The performance measures such as Classification Accuracy, Recall, Precision, and Mean Square Error (MSE) of MLP with various BP variants i.e, BR, Rprop, SCG, and LM are calculated and the results are presented in the table 1.

TABLE I. NETWORK PERFORMANCE MEASURES OF MLP MODELS

MLP Models	Accuracy (%)	Recall (%)	Precision (%)	MSE
MLP-BR	91.2	76	90	0.0617
MLP-RP	89.6	68	89	0.0833
MLP-SCG	91.5	76	90	0.0629
MLP-LM	***92.03***	***78***	***90***	***0.0616***

By analyzing the results it is clear that MLP-LM is the best suitable model with highest accuracy 92.03% and minimum MSE 0.0616 on Pap smear data than other models.

The analysis of MLP models performances are represented graphically in figure 3 and figure 4.

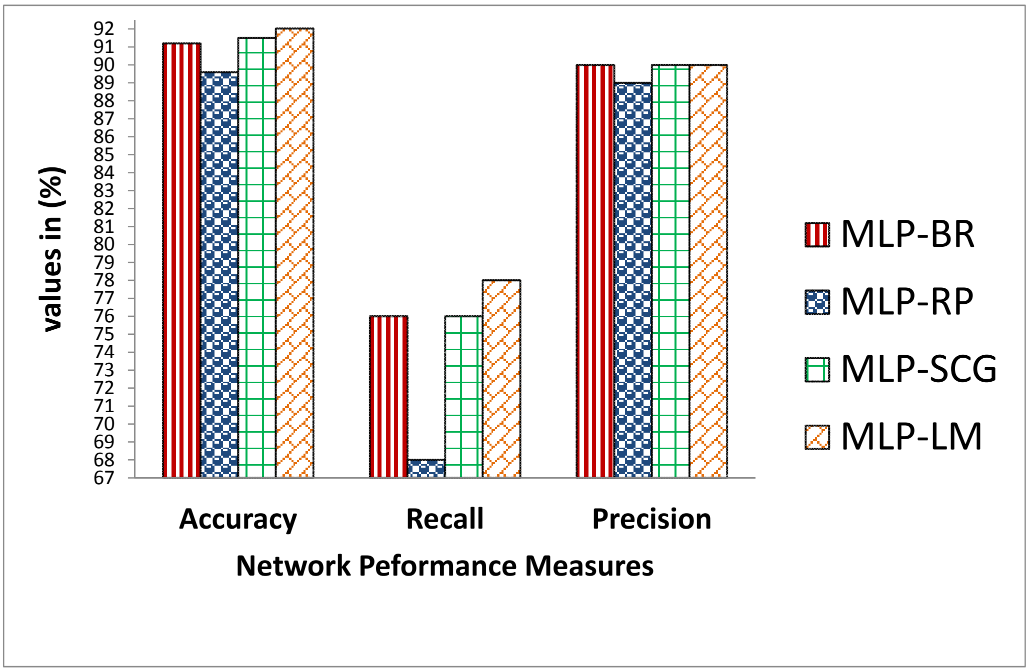

Figure 3. Performance Analysis of Accuracy, Recall and Precision of MLP models

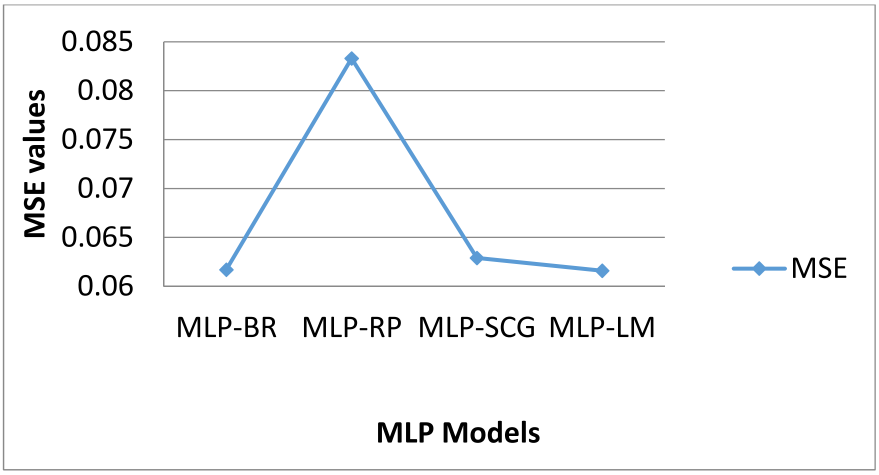

Figure 4. Analysis of MSE of MLP Models

Figure 3 shows that MLP-LM model has the highest Accuracy whereas MLP-SCG model has the second best highest accuracy. The recall value of MLP-BR and MLP-SCG are equal. MLP-LM has highest Recall value. The precision value of MLP-BR, MLP-SCG and MLP- LM are same.

From the above figure it is observed that the MSE value of MLP-LM and MLP-BR models are almost cost to each other.

However MLP-LM has the highest accuracy than MLP-BR model, Out of the four models, MLP-LM is considered as the best suitable model, because it has higher classification accuracy, Lower MSE value than other models on Cervical Pap smear data.

IV. CONCLUSION

Multi-Layer Perceptron (MLP) is a popular Feed-forward Network used for the classification of Medical Data. The performance of MLP depends on its training process. Selection of proper learning algorithm affects the training process. Back Propagation (BP) is the most preferable choice to train MLP. Hence, in this study various models of MLP using four BP variants such as Bayesian Regularization (BR), Resilient back propagation (Rprop), Scaled Conjugate Gradient (SCG) and Levenberg-Marquardt (LM) are build. Using these models MLP is trained to classify the MCPS dataset into normal and abnormal classes. Out of these MLP-LM model trained the MLP successfully and enhanced its performance. Hence, it is concluded that LM, a BP variant successfully trained MLP and enhanced its classification performance than other BP variants on Modified Cervical Pap smear Data.

REFERENCES

[1] Werner Kinnerbrock, "Neural Network", New Delhi, 1995.
[2] E. Kolay and T. Baser, “Estimating of the dry unit weight of compacted soils using general linear model and multi-layer perceptron neural networks,” Applied Soft Computing, vol. 18, 2014, pp. 223-231.
[3] Chhedaiya, Neelmani, Bhilai, Ssgi, Moyal, Vishal, “Implementation of Back Propagation Algorithm in Verilog”, International Journal of Computer Technology and Applications. 03, 2012.
[4] S. S. Haykin, “Neural networks and learning machines”, vol. 3. Pearson Upper Saddle River, NJ, USA, 2009.
[5] K. Hemalatha and K. Usha Rani, “An Efficient Fuzzy Hybrid Method for Image Segmentation and Feature Extraction of Cervical Pap smear Images”, International Journal of Imaging Science and Engineering (IJISE) ISSN: 1934 – 9955, Volume-10, Issue-1, January, (2018).
[6] Manke, Preeti & Publication IAEME, “A Performance Comparison of Different Back Propagation Neural Networks for Nitrogen Oxides Emission Prediction in Thermal Power Plant ”, International Journal of Computer Engineering & Technology (IJCET), 5, 2014, 65-73.
[7] Vallurn, B., and Hayoriva, V,” C++ Neural Network and Fuzzy Logic”, 2nd ed., USA: MIT Publishing, 1996.
[8] Aal-Yhia, Ahmad, “Enhanced Energy Backpropagation Algorithm”, Iraqi Journal of Science, 50, 2009, 553-560.
[9] Ganesh K. Venayagamoorthy, “Teaching Neural Networks Concepts and Their Learning Techniques”, Proceedings of the 2004 American Society for Engineering Education Midwest Section Conference, 2004.
[10] Anil Ahlawat, Sujata Pandey, “A Variant of Back-Propagation Algorithm for Multilayer Feed-Forward Network”, International Conference on Information Research & Applications, i.Tech, 2007.
[11] Shashank Mishra, G.S. Tripathi, “Comparison of Various Neural Network Algorithms used for Location Estimation in Wireless Communication”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 6, June, 2013.
[12] S.Esugasini, M. Y. Mashor, N. A. Mat Isa, N. H. Othman, “Performance Comparison for MLP Networks Using Various Back Propagation Algorithms For Breast Cancer Diagnosis”, Conference Paper in Lecture Notes in Computer Science, August, 2005.
[13] Titterington, D.M, “Bayesian methods for neural networks and related models”, Stat. Sci, 19, 2004, 128–139.
[14] Kayri, Murat, “Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data”, Mathematical and Computational Applications. 21, 2016, 1-11.
[15] Shashank Mishra, G.S. Tripathi, “Comparison of Various Neural Network Algorithms Used for Location Estimation in Wireless Communication”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 6, June, 2013.
[16] Moller, M.F, “A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. Neural Networks”, Neural Networks, 6, 1993, 525-533.
[17] Amrut Ranjan Jena, Raja Das, D.P.Acharjya, “Using of Three Training Algorithms for Material Remove Rate (MRR) In Electrical Discharge Machining (EDM) for ANSI D2 Steel by Neural Network: A Comparative Study”, International Journal of Pharmacy & Technology, Vol. 8, Issue No.2, 12701-12711, 2016.
[18] Zulhadi Zakaria, Nor Ashidi Mat Isa and Shahrel A. Suandi, “A Study on Neural Network Training Algorithm for Multiface Detection in Static Images”, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol: 4, No: 2, 2010