Chebyshev Multilayer Perceptron Neural Network with Levenberg Marquardt-Back Propagation Learning for Classification Tasks

Umer Iqbal[(⊠)] and Rozaida Ghazali

Faculty of Computer Science and Information Technology, Universiti Tun
Hussein Onn Malaysia, Batu Pahat, 86400 Parit Raja, Johor, Malaysia
umeriqball5@gmail.com, rozaida@uthm.edu.my

Abstract. Artificial neural network has been proved among the best tools in
data mining for classification tasks. Multilayer perceptron (MLP) neural network
commonly used due to the fast convergence and easy implementation. Mean-
while, it fails to tackle higher dimensional problems. In this paper, Chebyshev
multilayer perceptron neural network with Levenberg Marquardt back propa-
gation learning is presented for classification task. Here, Chebyshev orthogonal
polynomial is used as functional expansion for solution of higher dimension
problems. Four benchmarked datasets for classification are collected from UCI
repository. The computational results are compared with MLP trained by dif-
ferent training algorithms namely, Gradient Descent back propagation
(MLP-GD), Levenberg Marquardt back propagation (MLP-LM), Gradient
Descent back propagation with momentum (MLP-GDM), and Gradient Descent
with momentum and adaptive learning rate (MLP-GDX). The findings show
that, proposed model outperforms all compared methods in terms of accuracy,
precision and sensitivity.

Keywords: Classification · Chebyshev multilayer perceptron · Levenberg
Marquardt · Gradient descent

1 Introduction

Classification is appeared as revolutionary task in data mining. Classification problem
occurs when an object needs to be assigned to a specific class or group on the base of
their attributes that related to objects. The significance of classification can be seen in
different real life phenomena's such as diagnosis of diseases in medical [1], marketing
[2] and in stock exchange [3]. Classification task is based on two steps. First step is to
construct the model, which represents the group of precogitated classes. Second step is
the model usage, which is used for classifying the unknown objects.

Various techniques have been developed for classification, such as statistical and
neural networks techniques, which are prominent. With the passage of time, artificial
neural networks have gained much popularity as useful alternative of statistical tech-
niques and due to having the variety of applications in real life [4]. Multilayer per-
ceptron known as MLP, is a feed forward neural network that consists of input, hidden
and output nodes. Every node is interconnected with other node in next layer which

makes a connection between them. Back propagation is one of the most used and old supervised learning algorithm, which is proposed by Rumelhart, Hinton and Williams [5]. MLP has been used to find the missing values from data [6]. Due to its ability of fast learning, MLP has been tested for stock trading problems for better prediction [7]. MLP is also successfully implemented for early fault detection in gearboxes [8].

Researchers used different types of learning algorithm with back propagation to train different types of algorithms such as adaptive momentum to improve accuracy of gradient descent [9]. Some trained the artificial bee colony (ABC) with LM for classification problems [10]. Besides the advantages of MLP, there are also some disadvantaged of multilayer perceptron such as firstly, this neural network can only be used for supervised learning, and secondly due to its multilayer structure, it cause computationally expensive training that stuck in local minima [11].

In this paper, we propose Chebyshev polynomials as functional expansion with multilayer perceptron. These polynomials are used to make standard MLP more accurate for classification tasks. We have compared proposed method with four types of learning algorithms that are used as training of multilayer perceptron. The rest of the paper is organized as follows: In Sect. 2, the proposed model is presented. Some experimental setups are presented in Sect. 3. Section 4 is devoted to results and discussions. Finally, Sect. 5 outlines the conclusion.

2 Proposed Model: Chebyshev Multilayer Perceptron Neural Network

According to approximation theory, the non-linear approximation capacity of Chebyshev orthogonal polynomial is very dominant [12]. The proposed method is a combination of the characteristics of Chebyshev orthogonal polynomial and multilayer perceptron, which is, named as CMLP. This method utilizes the MLP input-output pattern with non-linear capabilities of Chebyshev orthogonal polynomial for classification. The Chebyshev multilayer perceptron is multilayer neural network. The structure consists of two parts, first one is the transformation part and learning is the second part. Transformation means that from a lower feature space to higher feature space. The approximate transformable method is implemented on input feature vector to the hidden layer in transformable part. This transformation is also known as functional expansion, where Chebyshev polynomial basis can be seen as a new input layer. Levenberg Marquardt back propagation is used for as a learning part [1]. Table 1 describes the recurrence relation to find the Chebyshev polynomials of degree n.

Where $T_0(x)$, $T_1(x)$ are the Chebyshev polynomials when $n = 0, 1$ and $T_{n+1}(x)$ is the equation for n^{th} polynomial. The reason for which we are using Chebyshev polynomials is that, when we are using polynomial of truncated power series then series represent the

Table 1. Recursive formula for Chebyshev polynomials

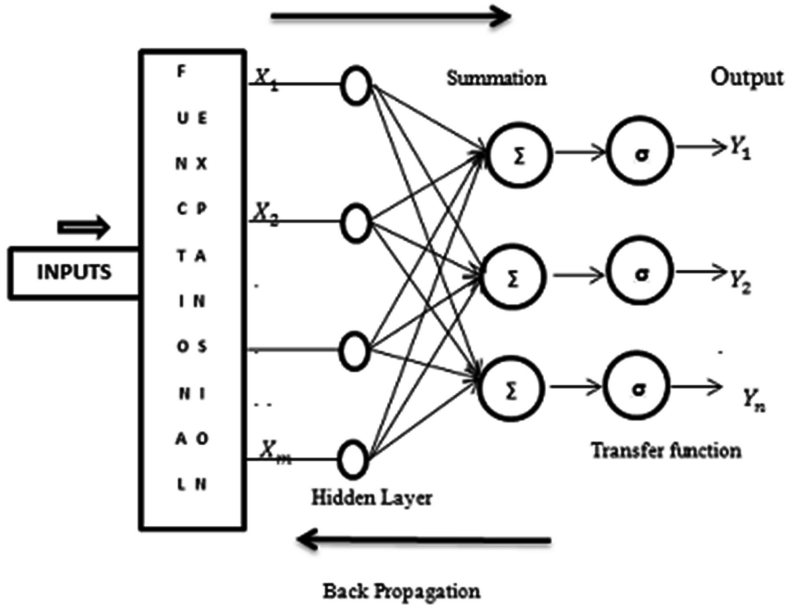|	
$T_0(x) = 1,$	(2.1)
$T_1(x) = x,$	(2.2)
$T_{n+1}(x) = 2xT_n(x) + T_{n-1}(x).$	(2.3)

Fig. 1. Chebyshev multilayer perceptron neural network for classification

function with very small error near the point of expansion and when implemented on all points then it cause increase in error. On the other hand, computational economy gained by Chebyshev increases when power series is slowly convergent [12]. This reason makes Chebyshev polynomials more useful and effective for approximation to functions as compared to other polynomials. Chebyshev polynomials are more convergent than the other expansions of polynomials. These properties make Chebyshev polynomial more selective on other polynomials. The proposed method can be seen in Fig. 1.

Where, $X_1, X_2 \ldots X_m$ are the inputs and $Y_1, Y_2, \ldots Y_n$ indicates the output of the neural network.

3   Experimental Setup

This section will illustrate about the working of proposed method, comparison techniques, data sets and evaluation measures.

3.1   Data Collection

The data set for classification analysis, which is the requisite input to the models, is obtained from UCI Repository [13]. We have collected four data sets named as, Iris, Wine, Breast Cancer and Bank Authentication for classification tasks. The dataset is divided in two parts such as; training set and testing set. The data ratio of 70% and 30% was set for training and testing respectively. The Iris dataset consists of 4 features and

150 samples. This data is categorized into three classes named as, Iris Setosa, Iris Versicolour and Iris Virginica. Wine dataset consist of 13 features and 178 samples. This data is categorized into three classes. Similarly, breast cancer data consists of 10 features and 699 samples. This data is categorized into two classes such as, benign and malignant. Respectively Bank Authentication data is consisting of 1372 samples, 5 features and 2 classes. Detail of used datasets is described in Table 2.

Table 2. Description of datasets

Datasets	Number of samples	Number of attributes	Number of classes
Iris	150	4	3
Wine	178	13	3
Breast Cancer	699	10	2
Banknote Authentication	1372	5	2

3.2   Proposed Methodology

The Fig. 2 illustrates the working of proposed method for classification. Inputs are expanded from lower dimension to higher dimension by using Chebyshev functional expansion block [14]. To train the network, we supposed the random weights; choose number of hidden nodes, outputs nodes (according to desired classes) and activation function. Levenberg Marquardt back propagation is used as learning algorithm. If we get our desired error according to the classification task then training will be finished, otherwise, we have to start the training again until the desired results followed by the testing of data based on evaluation measures.
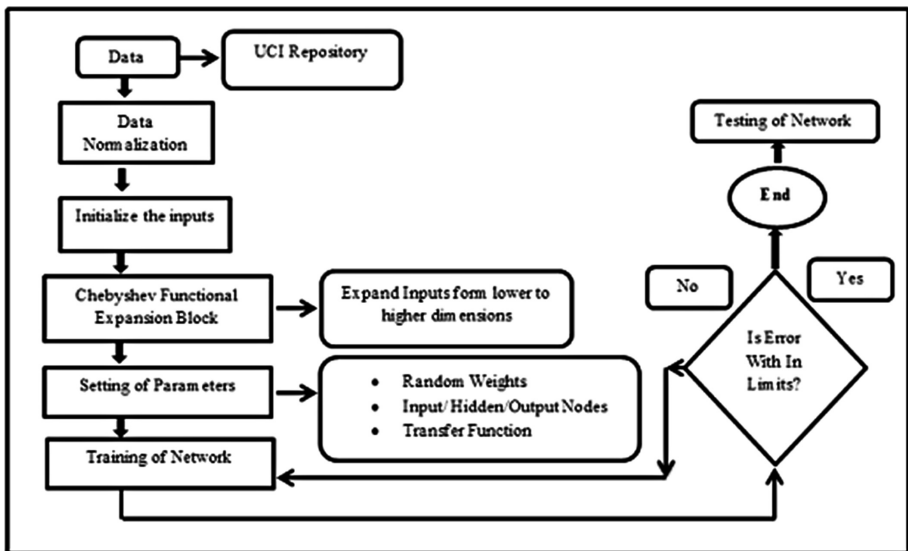
Fig. 2. Working of chebyshev multilayer perceptron neural network

3.3   Benchmarked Approaches

MLP appeared as a good competitor over the statistical methods due to its less com-
putational cost and better accuracy [15]. Many researchers implemented multilayer
perceptron neural network for different task of data mining by using different learning
algorithms. These learning algorithms proved their strength to make this network
stronger as compared to statistical methods. Researchers have used the Levenberg
Marquardt back propagation (LM-BP) [1], gradient descent back propagation (GD-BP)
[9], gradient descent with momentum back propagation (GDM-BP) [16], gradient
descent with momentum and adaptive learning rate back propagation (GDX-BP) [17]
to improve the efficiency of the neural network. We have compared our proposed
method with multilayer perceptron neural network learned with these four mentioned
learning algorithms.

3.4   Evaluation Measures

The performance of the proposed model against the MLP with four different learning
algorithms will be claimed based on accuracy, precision and sensitivity for classifi-
cation tasks. The equation to find out the accuracy is given as:

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}. \tag{3.1}$$

For classification task, precision is defined as "The number of true positives is
divided by the total number of true positives and false positives". The equation for
precision is described as:

$$Precision = \frac{Tp}{Tp + Fp} \times 100. \tag{3.2}$$

The sensitivity for the classification is calculated as:

$$Senstivity = \frac{Tp}{Tp + Fn} \times 100. \tag{3.3}$$

Where Tp, Tn, Fp and Fn are the true positive, true negative, false positive and false
negative values respectively.

4   Results and Discussion

This section describes the experimental results for the proposed and comparison
methods. Classification accuracy of four considered data sets viz. Iris, Wine, Breast
Cancer and Bank Authentication was verified. Training and testing data for experi-
ments were taken as 7:3 respectively. Results were verified by trial and error method
that was done 10 times for each experiment and then average for each experiment was

considered for further processing. The number of epochs for each experiment was same that is 1000 epochs. Performance of proposed and comparison methods were based on accuracy, precision and sensitivity. We have represented our proposed model Chebyshev multilayer perceptron neural network as CMLP-LM. Here, LM is the learning algorithm that we used in our proposed method. On the other hand, four learning algorithms that we used with multilayer perceptron neural network as comparison techniques are written as MLP-LM, MLP-GD, MLP-GDM and MLP-GDX. It is evident from the results that CMLP-LM provides much better results with classification accuracy of 98%, 99%, 78.11% and 90% on Iris, Wine Breast Cancer and Bank Authentication data sets respectively. On considering the detail of classification accuracy, MLP-LM training proves to be 89% accurate on Iris classification results but it provides best results on wine data set, that is, 94.50%. MLP-GD classification accuracy was found to be 89.22% on wine data set. Similarly, MLP-GDM and MLP-GDX perform best on Iris dataset; those are 84% and 81.50% respectively. The Fig. 3 gives comparative accuracy results, which support the proposed model being more efficient on the comparison methods.

Figure 4 represents the precision results of proposed model compared with the existing methods. It can be inferred from the simulation results that MLP-LM and MLP-GD give their best precision values on Wine dataset, that are, 95% and 90.4% respectively while MLP-GDM and MLP-GDX precision values are 83.8% and 79% respectively on Iris data set. On the other hand, CMLP-LM precision results appear to
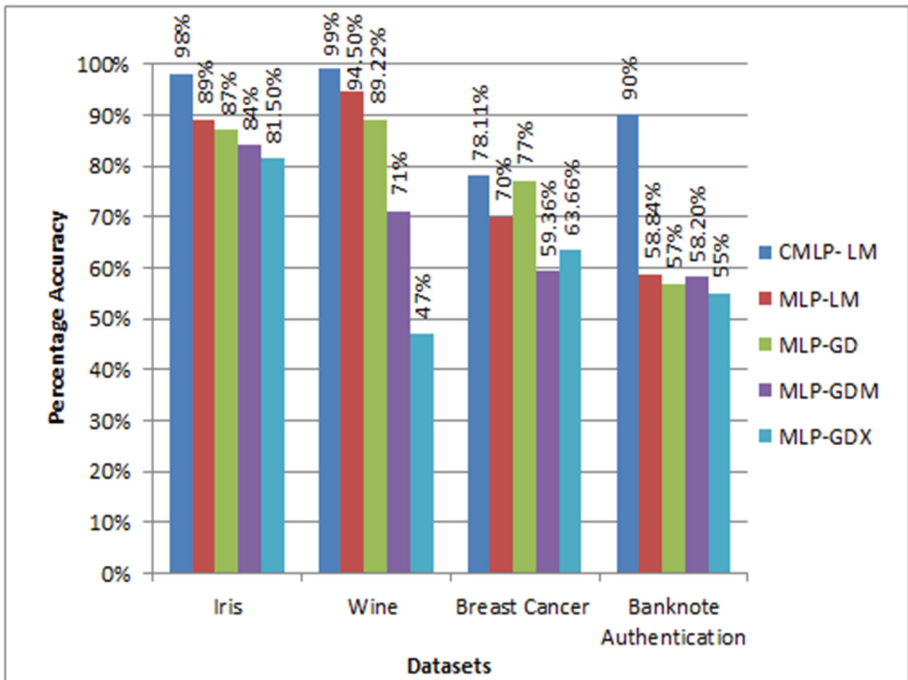

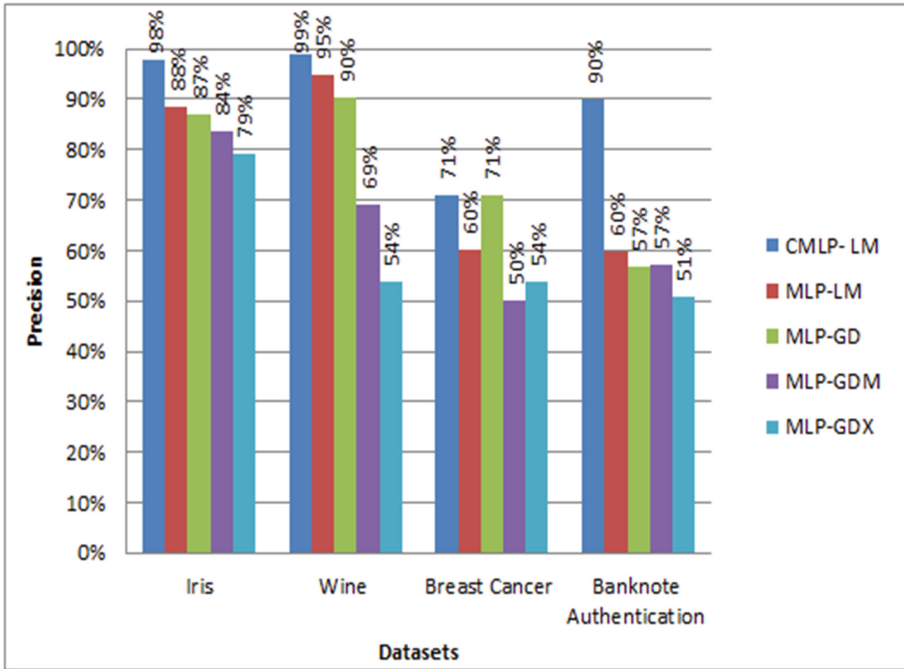**Fig. 3.** Comparison in terms of classification accuracy

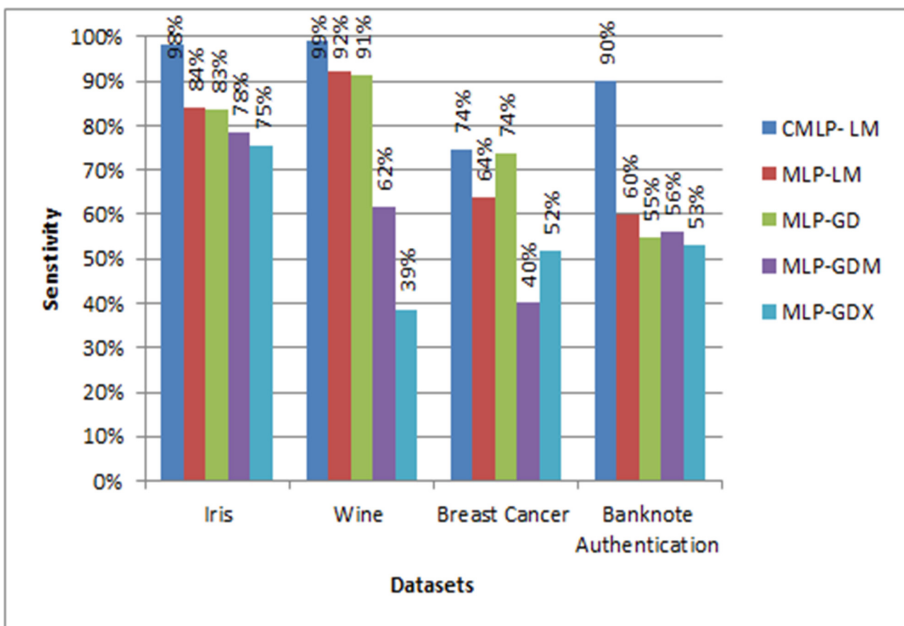**Fig. 4.** Comparison in terms of Precision

Fig. 5. Comparison in terms of Sensitivity

be more precise on all data sets, which are, 97.96% for Iris, 98.9% for Wine and 70.93% for Breast Cancer, 90% for Bank Authentication.

Comparative analysis of proposed and comparison methods based on sensitivity shows that MLP-LM and MLP-GD results are 92.29% and 91.36% on Wine dataset but MLP-GDM and MLP-GDX results for Iris are 78.44% and 75.42% respectively. Figure 5 shows the performance of CMLP-LM for Iris, that is, 98.06%; 98.99% for Wine, 74.44% for Breast Cancer and 90% for Bank Authentication.

5   Conclusion

In this paper, we have proposed the Chebyshev multilayer perceptron neural network with Levenberg Marquardt back propagation learning algorithm (CMLP-LM). This model was developed for classification task. Method was experimentally trained and tested on four benchmarked data sets which are taken from UCI repository. The performance of proposed method indicates the validity of classification task. The evaluation measures performance is showing that the CMLP-LM has better performance in terms of accuracy, sensitivity and precision over MLP-LM, MLP-GD, MLP-GDM and MLP-GDX.

References

1. Kumar, M., Singh, S., Rath, S.K.: Classification of microarray data using functional link neural network. Procedia Comput. Sci. **57**, 727–737 (2015)
2. Decker, R., Kroll, F.: Classification in marketing research by means of LEM2-generated rules. In: Advances in Data Analysis, pp. 425–432 (2007)
3. Bebarta, D.K., Biswal, B., Rout, A.K., Dash, P.K.: Forecasting and classification of Indian stocks using different polynomial functional link artificial neural networks. In: INDICON Annual IEEE, pp. 178–182 (2012)
4. Paliwal, M., Kumar, U.A.: Neural networks and statistical techniques: a review of applications. Expert Syst. Appl. **36**(1), 2–17 (2009)
5. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Cogn. Model. **5**(3), 533–536 (1988)
6. Silva-Ramírez, E.L., Pino-Mejías, R., López-Coello, M.: Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbours for monotone patterns. Appl. Soft Comput. **29**, 65–74 (2015)
7. Mabu, S., Obayashi, M., Kuremoto, T.: Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems. Appl. Soft Comput. **36**, 357–367 (2015)
8. Jedliński, Ł., Jonak, J.: Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. Appl. Soft Comput. **30**, 636–641 (2015)

 9. Rehman, M.Z., Nawi, N.M.: The effect of adaptive momentum in improving the accuracy of gradient descent back propagation algorithm on classification problems. In: Software Engineering and Computer Systems, pp. 380–390 (2011)
10. Shah, H., Ghazali, R., Nawi, N.M., Deris, M.M., Herawan, T.: Global artificial bee colony-Levenberq-Marquardt (GABC-LM) algorithm for classification. Int. J. Appl. Evol. Comput. (IJAEC) 4(3), 58–74 (2013)
11. Lee, T.T., Jeng, J.T.: The Chebyshev-polynomials-based unified model neural networks for function approximation. IEEE Trans. Syst. Man Cybern. Part B: Cybernet. 28(6), 925–935 (1998)
12. Konstantinidis, S., Karampiperis, P., Sicilia, M.A.: Enhancing the Levenberg-Marquardt method in neural network training using the direct computation of the error cost function hessian. In: Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS) (2015)
13. Blake, C., Merz, C.J.: UCI repository of machine learning databases (1998)
14. Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I.: Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, pp. 1–18 (2015)
15. Liu, H., Tian, H.Q., Liang, X.F., Li, Y.F.: Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl. Energy 157, 183–194 (2015)
16. Singh, B., De, S., Zhang, Y., Goldstein, T., Taylor, G.: Layer-specific adaptive learning rates for deep networks (2015)
17. Gates, G.W.: The reduced nearest neighbor rule. IEEE Trans. Inf. Theor. 431–435 (1972)