Comparative Analysis of FeedforwardBackpropagation and Cascade Correlation Algorithm on BUPA Liver Disorder

AlpanaJijja[#1], Dr.Dinesh Rai[#2], Priyanka Mathur[#3]

#School of Engineering & Technology, Ansal University, Gurgaon, Haryana, India.
[1]alpanajijja@ansaluniversity.edu.in
[2]dineshrai@ansaluniversity.edu.in
[3]priyankamathur@ansaluniversity.edu.in

Abstract—**Neural Network(NN) is widely used in medical research because of its cost effectiveness and easy-to-use systems. NN plays an important role in the decision support system. The sole reason to use the efficient NN model is to manage the rapidly increasing medical data effectively and use these data points for accurate prediction of diseases and thereby providing better health care to the patients. In this paper we have made use of two classification algorithmsi.e. feedforwardbackpropagation neural network(FFBPNN) and Cascade Correlationfeedforward network (CCFFN) and compared the performance of both the algorithms on the basis of sensitivity, specificity and accuracy. The experiment is done on Bupa liver dataset using FFBPNN and CCFFN model using different set of neurons which are trained with Levenberg-Marquardt training algorithm(trainlm) and ResilientBackpropagationtraining algorithm (trainrp). This paper has provided the comparative analysis among different classifiers using two different training algorithmsto find the benign and malignant patients.**

Keywords-Neural Network, FeedforwardbackpropagationNeural Network, CascadeCorrelationfeedforwardnetwork, trainlm,trainrp, BUPA Liver dataset.

I. INTRODUCTION

Artificial Neural Networks (ANN) is the mathematical Algorithms, generated by computers. ANNs train itself from standard data and capture the information contained in the data. ANN, when trained approaches the functionality of small biological neural group in a very elementary way. These are the digitized representation of biological intelligence and can detect multifaceted nonlinear relationships between dependent as well as independent variables in a data set where human brain may fail to distinguish. McCullogh and Pitts in 1951 stated the definition of the first artificial neuron [1].In parallel, with the evolution of computer technology, modeling of increasingly complicated neural functions and activity of simple neural clusters was defined. During 1982 and 1987 McLelland, Rummelhart, Hopfield and Kohonen developed the Mathematical models that were applied to practical applications[2,3].ANNs have been preferably used by many authors for modeling in medicine and clinical research. ANNs have been extensively applied in diagnosis, electronic signal analysis, medical image analysis and radiology[4].Various types of neural networks have been premeditated already and new ones are being invented but these can be described by the transfer functions of their neurons, by the learning rule, and by the connection formula. We discuss the main application fields of neural network technology in medicine. Between 1990 and 1997 applications of neural networks were introduced in near 2000 papers [5].

The largest and one of the most vitalinternalorgans in the human body is the liver with almost 4% of the body weight and blood flow of1.5 litre per minute. It receives blood supply from two main blood vessels: the hepatic artery and the portal vein. The hepatic artery supplies oxygenated blood, whereas portal vein provides 80% of the total blood supply.The normal stress in the portal vein is between 3 and 5mmHg. The liver is located at the right upper quadrant of the abdomen, completely protected by the thoracic rib cage. The liver serves as a guard between the digestive tract and other partsof the body. The liver detoxifies and accumulates metabolites. In extension, the liver has the capability of generating plasma protein for example albumin which are carried into the blood similarly metabolites which are constituents of bile. Liver diseases can be described as a defect that affects the liver. The types of liver diseases can be categorized into hepatocellular (hepatitis, heart failure and toxins), cholesteric, infiltrative diseases (tumour, sarcoid) and cirrhosis (hepatocellular loss and scarring). The liver performs over 500 jobs and does so silently and prudently. Among those jobs are the master tasks of managing cholesterol, hormones and filtering of blood. One can't withstand without any one of those processes. The liver is the largest organ in the body and when it is done doing all of that, it is the only organ that can regenerate itself. Because of the rapidly increasing medical data, it has become extremely important to manage

this data properly and use these data points for accurate prediction of disease and for providing better health care to patients. The rest of the paper is organized as follows: Section II – reflects on the literature review, Section III - describes the methodology in detail followed by Section IV, which describes our experimental setup. Section V deals with results and its discussion followed by the conclusion.

II. RELATED WORK:

In the works of Sumaiya et al.,the various classification algorithmswere implemented on different medical datasets using Rtool ver3.2.2 and evaluated result shows the performance of adaboost is better than other algorithms of classification[6]. Hyontai generated more accurate decision trees for liver disorder disease. This paper, suggests a method based on oversampling in minor classes to compensate the insufficiency of data effectively on decision tree algorithmsi.e. C4.5 and CART[7].

Bendivenkata et al., compared different classification algorithms for accuracy, precision, sensitivity and specificity in classifying liver patient dataset using NBC,C4.5,Back propagation, K-NN and SVM algorithm showing KNN,backpropagation and SVM are giving better results with all the feature set combinations[8] .Gupta et al., in their paper hadused WEKA to evaluate the performance of classification of algorithms and found that Random forest algorithm classify the various given dataset better than other four algorithms with 5 fold cross validation test[9].Kannan et al., have used classification technique using machine learning to analyze and predict, the best classification algorithm for diabetes diagnosis dataset. The different classifiers namely BayesNet,Baggingalgorithms and SVM, use Cross validation method of 10 fold and comparative analysis on execution time, accuracy and error rateswere used for analyzing results [10].Karthiyayini et al., in their paper have used the effectiveness of diverse techniques of data mining such as classification,association, regression. The survey also highlights requisite of datamining in medical field [11].

III. METHODOLOGY

A. Feed Forward Back propagation Neural Network

Neural networks area predictive model that have ability to learn, analyze, organize the data points and predict test results accordingly.Varioustype of neural network models have been proposed for a Pattern Classificationfunction approximation, and other modelling tools.Among them the class of multilayer feedforward networks is perhaps themost popular. Feedforward neural network are powerful models for solving nonlinear mapping problems[12]. The training of these networks is generally undertaken with a standard backpropagation type of training algorithm which performs gradient descent –based optimization routines in the weight space of a network with fixed topology[13]. In general this type of training is useful only when the network architecture is elected correctly.

Among several kinds of neural networks, feed forward neural network is usually deployed in medical diagnosis applications. These networks are trained by a set of patterns called training set, whose outcome is already known. In this study, that we have undertaken, Multilayer Perceptron feed forward back propagation Neural Network is trained with Levenberg Marquardt (LM) algorithm for classification. LM training algorithm does not get stuck in local minima and produces a better cost function. FFNN consists of input, hidden and an output layer, and the data operates in forward direction, the error is back propagated to update the weights at every epoch in order to reduce errors [14].Feed forward neural network is a non-parametric estimation of statistical models for extracting nonlinear relations of the input data. The training algorithm involves two phases [18].

1) *Forward Phase:*The free parameters of the networks are fixed and through the network, the input signal is propagatedduring this phase. It finishes with the computation of an error signal.

$$e_i \ = \ d_i - y_i \tag{1}$$

d$_i$ here is the desired responseand y_i is the actual output produced by the network in response to the input x_i.

2) *Backward Phase:* During this second phase, the error signal e$_i$ is propagated through the network in the backward direction, and that's how the algorithm gets its name. It is during this phase that the adjustments are applied to the free parameters of the network to minimize the error e$_i$ in a statistical sense. The back propagation learning algorithm is computationally efficient and simple to implement.The set of training examples is split into two parts:

1) Estimation subset used for the training of the model.
2) Validation subset used for evaluating the model performance.

The network, in general, is finally tuned using the entire set of training examples and then tested on test data [19].

B. Casscade –Correlation learning

Constructive learning changes the network structure as learning proceeds, producing accordingly a network with an appropriate size. In this approach one starts with an initial network of a small size and then incrementally adds new hidden units or hidden layers until some pre-specified error requirement is reached or no performance improvement can be realized. The network obtained this way is reasonably sized, for the given problem at hand. Constructive learning alters the network structure as learning proceeds, producing automatically a network with an appropriate size.

The cascade correlation is also a universal approximation.The training of the output unitminimizes the sum-squared error E:

$$Ep = \sum_p \frac{1}{2} \sum_o \left(y_{po} - t_{po} \right)^2 \qquad (2)$$

Where t_{po} is the desired output and y_{po} is the observed output of the output unit for a pattern p. The error E is minimized bygradient decent using

$$e_{po} = \left(y_{po} - t_{po} \right) f_p^I \left(net_o \right) \qquad (3)$$

where f_p^I is the derivative of an activation function of an output unit o and I_{ip} is the value of an input unit or a hidden unit i for a pattern p. w_{ip}denominates the connection between an input or hidden unit i and an output unit o.After the training phase the candidate units are adapted, so that the correlation C between the value $y_{po}$of a candidate unit and the residual error $e_{po}$of an output unit becomes maximal. The correlation is given by Fahlman'smathematical background:

$$C = \sum_o |\sum_p (y_{po} - \bar{y}_o)(e_{po} - \bar{e}_o)| \qquad (4)$$

$$C = \sum_o |\sum_p (y_{po}e_{po} - \bar{e}_o \sum_p y_{po}| \qquad (5)$$

$$C = \sum_o |\sum_p y_{po}(e_{po} - \bar{e}_o)|, \qquad (6)$$

Where $\bar{y}_o$ is the averageactivation of a candidate unit and $\bar{e}_o$is the average error of an output unit over all patterns p. The maximization of C proceeds by gradient descent using

$$\delta_p = \sum_o \sigma_o \left(e_{po} - \bar{e}_j \right) f_p^I$$

$$\frac{\partial C}{\partial w_i} = \sum_p \bar{o}_p I_{pi} \qquad (7)$$

where σ_o is the sign of the correlation between the candidate unit's output and the residual error at output o[15].

Training of cascade correlation architecture is by construction, i.e., it builds the architecture step by step, by adding one hidden layer neurons at a time. As it is a multilayer neural network framework, it has to solve the issue of how to trainhidden layer neurons without information concerning the hidden layer neuron behaviors. For comparative analysis cascade correlation network algorithm has been chosen as it has several advantages over feedforwardbackpropagation neural network. The former itself organizes and grows the hidden layer during training. Time taken for training is very fast – often 100 times as fast as a perceptron network. This makes cascade correlation networks suitable for large training sets.Thefollowing modelbest describes thecascade correlation neural network architecture, as introduced by Fahlman[15].shown in Fig 1:
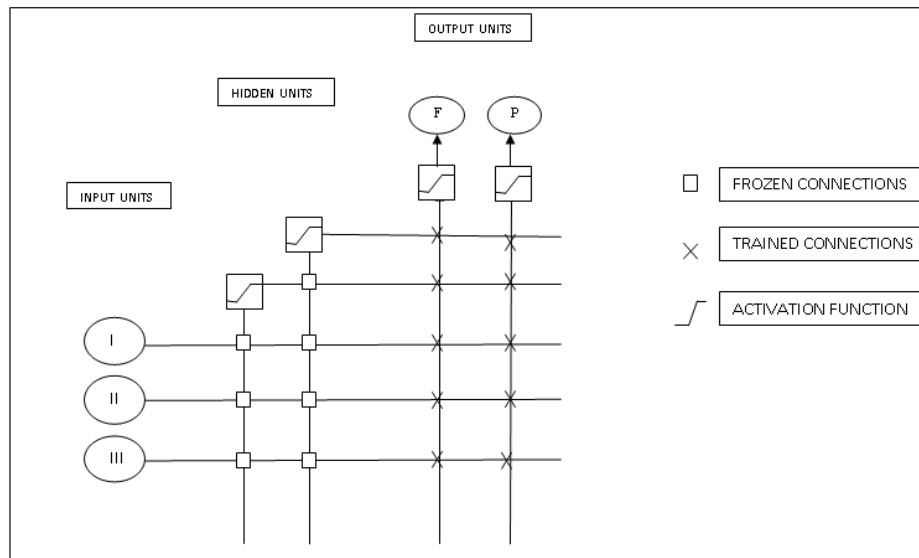
Fig. 1 Cascade Correlation Neural Network

Fig. 1 shows cascade correlation neural network architecture of two hidden units, three input units, and two output units. The hidden units essentially have nonlinear activation functions. The output units either have nonlinear activations, for classification problems; or linear activations, for regression problems. The input units all have linear activation functions while the square symbol denotes that the weights of the unit, which once obtained will be frozen. The cross symbol denotes weights which are required to be trained.

IV.   EXPERIMENTAL SETUP

Two models of neural network have been developed. These models are feedforward neural networks trained with back propagation (FFBP) and cascade correlation algorithm (CCA). The liver dataset was obtained from UCI machine learning repository created by BUPA Medical Research Limited, India[16].

This data set is trained on Levenberg –Marquardt Algorithm and Resilient Algorithm with one hidden layer of 10, 20, 30 and 40 neurons respectively. These two models of FFBP and CCA have three layer network which consists of an input that is connected with the hidden layer with aid of connection weights. The hidden layer is connected with the output layer also with the aid of connection weight.

The dataset is made up of six (6) attributes in which first five attributes are blood test which were thought to be sensitive to liver disorders that might arise from excessive consumption of alcohol. The attributes are:

1. Mean corpuscular volume (mcv)
2. Alkalinephosphotase(alkphos)
3. Alamine aminotransferase(sgpt)
4. Aspartate aminotransferase(sgot)
5. Gmma-glutamyltranspeptidase
6. Drinks (Number of half-pint equivalent of alcoholic beverages drunk per day).

A. Preprocessing

For the best performance of the classifier, it must change the attribute values into homogenous and well behaved values that give numerical stability [17].Therefore, the attribute values have to be in value ranging between 0 and 1. The output layer has 1 denoting liver cancer is present and 0 represents it is absent.The process is referred to as normalization. Normalization is obtained by dividing each sample of a particular attribute by its maximum value.

$$normalised\ sample = \frac{sample\ of\ attribute}{maximum\ value\ of\ teh\ attribute}$$

The classification experiments are conducted on the Liver disorder dataset. The BUPA Liver dataset contains 345 data points, 138 data are used for training and 238 data for testing. The results of FFBPNN and CCFFN classification for liver dataset are analyzed in the following section.

B. Performance measure

Performance accuracy assesses the overall effectiveness of the classifiers. Sensitivity and specificity are two measures that separately estimate a classifier's performance on different classes [20, 21].In terms of medical research, sensitivity is used to measure the percentage of correctly classified benign tumors while specificity is used to measure the percentage of correctly classified malignant tumors.

V.   RESULTS AND DISCUSSION

To come to an optimal neural network, neural networks that were trained with different parameters will have to be compared. We achieve this by analyzing the performance of the neural network. To get an unbiased measure of performance, the performance measure is calculated using the test data set. This set will not be used in the training process. We will express the performance of a neural network based on sensitivity, specificity and accuracy.

TABLE I. Results obtained on BUPA Liver Disorder Data Set using(trainlm)

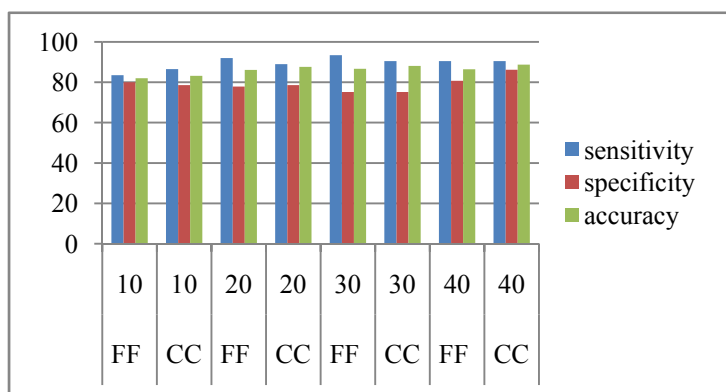| S_no	train(lm)	Neurons	Sensitivity %	Specificity%	Accuracy%
1	FF	10	83.5	80	82
2	CC	10	86.5	78.6	83.2
3	FF	20	92	77.9	86.1
4	CC	20	89	78.6	87.6
5	FF	30	93.4	75.2	86.7
6	CC	30	90.5	75.2	88.1
7	FF	40	90.5	80.7	86.4
8	CC	40	90.5	86.2	88.7

Fig.2.Comparison of classifiers using tarin(lm) algorithm.

TABLE II.Summary of the results obtained on BUPA Liver Disorder Data Set using(trainrp)

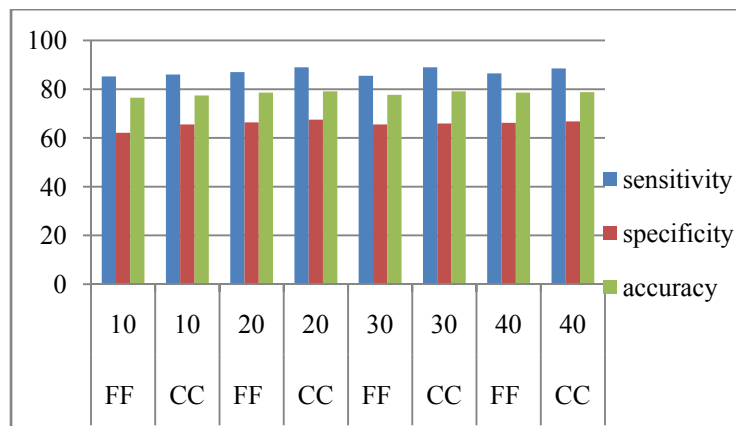| S_no	train(rp)	Neurons	Sensitivity %	Specificity%	Accuracy%
1	FF	10	85.2	62.1	76.5
2	CC	10	86	65.5	77.4
3	FF	20	87	66.4	78.6
4	CC	20	89	67.5	79.1
5	FF	30	85.5	65.5	77.7
6	CC	30	89	65.9	79.1
7	FF	40	86.5	66.2	78.6
8	CC	40	88.5	66.8	78.8

Fig. 3.Comparison of classifiers using train(rp) algorithm

VI.   CONCLUSION

In this paper we have performed comparison of selected classification algorithms on MATLAB12b using trainlm and trainrp respectively and represented the comparison graphically.In figure2.,the graphic representation shows that cascade correlation algorithm using trainlm is better than feedforwardbackpropagation algorithm. In figure 3.When training algorithm is changed to trainrp even then performance in terms of sensitivity, specificity and accuracy of cascade correlation algorithm is better. Finally it has been concluded that cascade correlation algorithm using both the training algorithms gives the better results. This work will motivate the medical experts to diagnose liver cancer.

REFERENCES

[1]   McCullogh WS, P. W. (1943). Bull math Biophys.,5, pp. 115-18.
[2]   T.Kohonen,"Self-Organized formation of topologically correct feature maps", Biological Cybermetics, vol.43, 1982,pp.59-69.
[3]   D. E. Hinton, and J. L. McClelland,"Ageneral framework for parallel distributed processing", in Parallel Distributed Processing: Explorations in the microstructureof Cognition, vol. I: Foundations, MIT Press, Cambridge, MA, pp45–76, 1986.
[4]   Tai, Z., & Wang, X. (2010). Control Automation Robotics & Vision, pp. 518 – 522.
[5]   Swati S,AnjanG,etal.,"Performance of ARTI network in the detection of breast cancer", in Int. Conf Design and Engineering,vol.49, no.19,pp 100-105,2012.
[6]   SumaiyaG and AngelinaGeetha,"Domain independent model for data prediction and visualization",(IJSRET),vol.l5,no.4, pp 241-247,2016.
[7]   HyontaiSug,"Improving the prediction accuracy of liver disorder disease with oversampling",Applied Mathematics in Electrical and Computer Engineering".
[8]   BendiV.R,Surendra P.B and N.B Venkateshwarlu,"Acritical study of selected classification algorithms for liver disease diagnosis",International Journal of Advanced research in Computer Science ,vol.3,no.2 ,May2011.
[9]   D. L Gupta, A.K Malviya and S.Singh,"Performanceanalysis of classification tree algorithms",International Journal of Computer Application,vol.55,no.6,pp39-44,2012.
[10] Kamalallannan.V and Ramyachitra.D."An experimental analysis of different classification techniques for diabetic dataset",International Journal of Advanced Research in Computer Science and Software Engineering,vol.6, no.2, February 2016.
[11] R.Karthiyayini and R.BalaSubramanian, "A survey in health care datausing data Mining Techniques ",Data Mining& Knowledge Engineering,CIIT, vol7,no.2, 2015.
[12] S.Haykiin. "Neural Networks ".A Comprehensive Foundation. New York: Macmillan.1994.
[13] R.Flectcher,"Practical methods of optimisation",2nd ed. New York:Wiley,1990.
[14] Paul R. Harper, "A review and comparison of classification algorithms for medical decision making" Health Policy,vol 71,no.3, pp.315-331, 2005.
[15] S. E. Fahlman." The cascade-correlation learning architecture" in D. S. Touretzky, editor, Advances in Neural Information Processing Systems, pp.524-532, San Mateo, CA, 1990. Morgan Kaufmann Publishers.
[16] D.J. Newman, S. Hettich, C.L. Blake, & C.J. Merz, UCI Repository of machine learning databases, (1998),http://www.ics.uci.edu/~mlearn/MLRepositor y.html, Irvine, CA: University of California, Department of Information and Computer Science, last access: 11th August 2015.
[17] Ebenezer O.Olaniyi, KhashmanAdnan,"Onset diabetes diagnosis usingArtificial neural network", International Journal of Scientific and Engineering Research ,vol. 5, pp 754-759,2014.
[18] D.E. Rumelhart, G.E. Hinton, R.J. WilliamsLearning,"Internal representations by back-propagating errors", Nature, 323 (6088), pp. 533 – 536,1986.
[19] S. Haykin,"Neural networks": A Comprehensive Foundation (second ed.)Prentice-Hall, Englewood Cliffs, NJ (1999) pp. 196–197.
[20] J. Ren, "ANN vs. SVM: which one performs better in classification of MCCs in mammogram imaging,"Knowledge-Based Systems, vol. 26, pp. 144-153, 2012.
[21] T. S. Subashini, V. Ramalingam, S. Palanivel, "Breast mass classification based on cytological patterns usingRBFNN and SVM," Expert Systems with Application, vol. 36, no.5, pp. 5284-5290, 2009.