Prediction Of Heart Disease Using Back Propagation MLP Algorithm

Durairaj M, Revathi V

Abstract: Diagnosing the presence of heart disease is actually tedious process,as it requires depth knowledge and rich experience. In general, the prediction of heart disease lies upon the traditional way of examining medical report such as ECG (The Electrocardiogram), MRI (Magnetic Resonance Imaging), Blood Pressure, Stress tests by a medical practitioner. Now days, a large volume of medical data is available in medical industry and acts as a great source of predicting useful and hidden facts in almost all medical problems. These facts would really in turn, help the practitioners to make accurate predictions. The novel techniques of Artificial Neural Network concepts have also been contributing themselves in yielding highest prediction accuracy over medical data. This paper aims to predict the existence of heart disease using Back Propagation MLP (Multilayer Perceptron) of Artificial Nerual Network. The results are compared with the existing works carried out in the same domain.

Index Terms: Minimum ECG, MRI, MLP, Nerual Network.

1 INTRODUCTION

Heart is an important organ of all living creature, which plays a vital role of pumping blood to the rest of the organs through the blood vessels of the circulatory system. Any functional problem in the heart has a direct impact on the survival of concerned human being, since it affects other parts of the body such as brain, lungs, kidney, liver etc. Heart Diseases describe a rang of conditions that affect the heart and stand as a leading cause of death all over the world. The clinical symptoms of the Heart Disease complicate the prognosis, as it is influenced by many factors like functional and pathologic appearance. This could subsequently delay the prognosis of the disease. Hence, there is a need for the invention of newer concepts to improve the prediion accuracy with short span. Disease prognosis through numerous factors or symptoms is a multi-layerd problem, even that could lead to a false assumption. Therfore, an attempt is made to bridge the knowledge and the experience of the experts and to build a system that fairly supports the diagnosing process. Hence, this paper proposes a new approach, by implementing the Multilayer Perceptron (MLP) Algoithm over a Heart Disease dataset that was taken from the popular UCI machine laearning repository. The rest of the paper is organized as follows.Related works are presented in Section 2. In Section 3, the methodology is described. Design and implementation details are presented in Section 4, Section 5, contains the experimental result and Section 6 concludes the paper.

2.REVIEW OF RELATED WORKS

In [1], medical decision support system based on the MLP Nerual Network architecture for heart disease diagnosis. Identified 40 input varibles critical to the diagnosis of the heart diseases of interest and encoded them accordingly. The system is trained by employing an improved BP algorthim. A heart diseases database consisting of 352 cases has been used in this study. Three assessment methods, cross validation, holdout and bootstrapping, have been applied to assess the generalization of the system. The results show that the proposed system can achieve very high diagnosis accuracy (<90%) and comparably small intervals (>5%), proving its usefulness in support of clinic diagnosis descicion of heart diseases. In [2], presented Heart Disease prediction system (HDPS) using data mining and Artificial Nerual Network (ANN) techniques. From the ANN, a Multilayer Perceptron Nerual Networl (MLPNN) along with Back Propagation algorithm is used to develop the system. Because MLPNN model proves the better results and helps the domain experts and even person related with the field to plan for a better diagnose and provide the patient with early diagnosis results as it performs realistically well even without retraining. The experimental result shows that using neural networks, the system predicts Herart Disease with more accuracy. In [3], the momentum term, adaptive learning rate, the forgetting mechanics and conjugate gradients method are applied to improve the basic BP (Back Propagation) algorithm for MDDSS (Medical Diagnostic Decision Support System). A case study with a heart diseases consisting of five common heart disease to assess the classification accuracy and convergence speed of the MLP network for each step improvement of learning algorithm indicates that the network the augmented BP algorithm has a quicker and higher capability of classification for Heart Diseases, and the MLP network System established shows great application prospect in supporting Heart Diseases diagnosis. In [4], based on the hypothesis of research, new SMFFNN (Supervised Multilayer Feed Forward Nerual Network) model was proposed which can get the best result in speed and accuracy by using new preprocessing technique without gradient of mean square error function and updating weights in one epoch. During experiments, the new model was implemented and analyzed using Weight Linear Analysis (WLA). The combination of data pre-processing and new pre-training techniques show that WLA generated normalized input values and potential weights. This shows that WLA serves as global mean and vectors torque formula to solve the problem. Two kinds of datasets from UCI `Repository of Machine Learning are choosen to illustrate the strength of WLA techniques. SPECTF Heart is a multivariate integer dataset and SPECT

- *M.Durairaj, Assistant Professor, School of Computer Science, Engineering And Applications, Bharathidasan University, India, Mobile No:+91 9487542202*
- *(e-mail:durairaj.bdu@gmail.com)*
- *V.Revathi, Research Scholar, School of Computer Science, Engineering And Applications, Bharathidasan University, India, Mobile No:+91 9597879303*
- *(e-mail:revathittm@gmail.com)*

Heart is a multivariate categorical-binary dataset. The result of BPN by using pre-processing techniques and new SMFFNN with application of WLA showed significant improvement in speed and accuracy. Therefore, the new model with proposed technique can solve the main problem of finding the suitable weights. The accuracies implementation on two datasets of SPECTF and SPECT Heart are computed by F-measure formula. The results show the robust and flexibility properties of new technique and new SMFFNN model for classification. In [5], highlighted that the classification is an important problem in the rapidly emerging field of data mining. As per wide range of applicability of ANN and their ability to learn complex and nonlinear relationships including noisy or less precise information, Nerual Networks are well suited to solve problems in biomedical engineering. From the analysis, it was observed that Multilayer Feed Forward Network with BackPropagation algorithm using 15 input attributes gives the highest accuracy BackPropagation algorithm with momentum and variable learning rate was used to train the networks. To analyze performances of the network various test data were given as input to the networks. To analyze performance of the network various test data were given as input to the network. This binary Heart Disease dataset classifier can be used to assist physicians to classify the dataset of Heart Disease. From the Nerual Networks design was evident that MLP NNs required a compact architecture as compared to other NNs, in terms of number of hidden nodes required for the classification. Thus the number parameters such as weights and biases required for the designing of MLP NN is suffiently lower than other. In [6], presented an approach that was based on Back Propagation Nerual Network to predict the Heart Disease. In this paper, the prediction of Heart Disease is developed using Neural Network. Totally 13 medical attributes were used in the experiment and it has shown performance improvisation compared to similar approaches of the state of the art. In [7],the MLP and improved data mining classification techniques were used to predict cardiovascular disease in patients. Independent evaluation was done using 10v fold crossValidation. Improved MLP model had more prediction power than MLP. The results showed that improved MLp algorithm gave better performance than MLP on parameters TPR, FPR, F Measure, ROC, Sensitivity and Accuracy. Results of above three simulated experiments show that in all parameters such as TRP, FPR. ROC, F Measure and Accuracy improved MLP approach out performs MLP. ROC value is also closer to 1.0. It is presented a simple and efficient way to improve execution time of MLP by increasing accuracy. So the proposed improved MLP method is feasible. Limitation of this study was that drawing all conclusions from data collected for decision making should be done with caution. Also newly identified cases still need to be validated to confirm their positivity. In [8], Heart valve diseases are Aortic Stenosis (AS), Aortic Reguragiation (AR), Mitral Stenosis (MS) and Mitral Regurgitation (MR). Initially the Heart Sounds were successfully categorized using a SVM (Suppoort Vector Machine) classifier as normal or disease-related and then the corresponding murmurs in the unhealthy cases were classified as systolic or diastolic. For the Heart Sounds diagnosed as having systolic murmur using SVM classifier for performing a more detailed classification of the having aortic stenosis or mirtal regurgitation. In [9], the work was based on a wavelet decomposition of the sounds and a Nerual Network based classifier, Heart Sounds are associated with likely underlying pathologies. Preliminary results promise a system that is both accurate and robust, while remaining simple enough to be implemented as low cost. In [10], it is demonstrated that more effective models can be created. This work used SAS based software 9.1.3 and obtained 97.4% classification accuracy from the experiments made on data set containing 215 samples. It also obtained 100% and 96% sensitivity and specificity values, respectively, in valvular Heart Disease diagnosis. In [11], the Experimental results show that recognition rate obtained by Sugeno Fuzzy integral with triangular norm is more successful than recognition rates obtained by standard discreate HMM (DHMM-Discrete Hidden Markov Model) and choquet integral based FDHMM (Fuzzy Discrete Hidden Markov Model). In addition to this the performance of the Sugeno integral based method was better than the performances of Artificial Nerual Network (ANN) and HMM based method classification systems that were used in previous studies of the authors. In [12], the Artificial Nerual Network (ANN) algorithm was used for classifying the Heart Disease based on output. Learning Vector Quantization (LVQ) is a prototype based Supervised Classification Algorithm. Accuracy was 80% Sensitivity was 85%, Sensitivity was 85% and Specificity was 70%. The performance of the prognostic model is analysed using the following criteria: If a classification model is trained the predictive model will present the best accuracy.

Accuracy= (TP+TN)/(TP+TN+FP+FN)*100%
Sensitivity=TP/(TP+FN)*100%
Specificity=TN/(FP+TN)*100%
Where TP is the number of True Positive
Tn is the number of True Negative
FP is the number of False Positive
FN is the number of False Negative

Area under Receiver-Operator Characteristics Curve (ROC). The curve that has a superior area under curve is better than the one that has a minor area under curve. The specificity is the probability that a test is negative for patients without diabetes and Heart Disease. In [13], place Heart Sound is a very informative bio-signal, since it directly encodes the mechanical activity of the Heart. It has been shown that Heart Sound has the potential to be applied in very important biomedical applications both in a hospital and in a home-care setting. For identifying the normal and different types of abnormal Heart Sound signals was applied a system based on feature extraction analysis combined with support vector machine was applied. This work a survey on diagnosis of coronary Heart Disease using ANN, Fuzzy Logic and Neuro-Fuzzy integrated approach. It has been found that although ANN and Fuzzy Logic have a lot of advantages but these techniques have some disadvantages too. It was observed from this work that, more data make the network more intelligent and gives much accurate results.

3 METHODOLOGY

Methodology is a method of accomplishing the proposed work with the proposed algorithm. In this paper the Multilayer Perceptron Back Propagation Algorithm of Nerual Network is used Back Propagation Algorithm is also called as ‘Back Propagation of Errors Algorithm’, is a popular method that produce an effective ANN (Artificial Nerual Network) training in conjunction with some optimization techniques like gradient desent. The method computes the depth of the loss function in the input data with respects to all the weights in the network.

The gradient techniques are then applied to the optimization methods to adjust the weights to minimize the loss function in the network. Hence, the algorithm requires a known and a desired output for all inputs in order to compute the gradient of loss function. Usually, the generalization of MultiLayerd Feed Forward Networks is done using delta rule which possibly makes a chain of iterative rules to compute gradients for each layer. Back Propagation Algorithm necessitates the activation function to be different between the neurons. The ongoing researches on parallel, distributed computing and computational neuroscience are currently implemented with the concepts of MultiLayer Perceptron using a Back Propagation Algorithm. MLP Back Propagation Algorithm has also gained focus in pattern recognition domain. They are so convenient in research, because of their ability in solving complex problems, and also for their fitness approximation results even with critical predictions. MLP is one of the Nerual Network models, has the same architecture of Feed-Forward back Propagation for Supervised training. In general, MLP network contains an input, one or more hidden layers and an output layer. Fig 1. Depicts the architecture of MLP network.

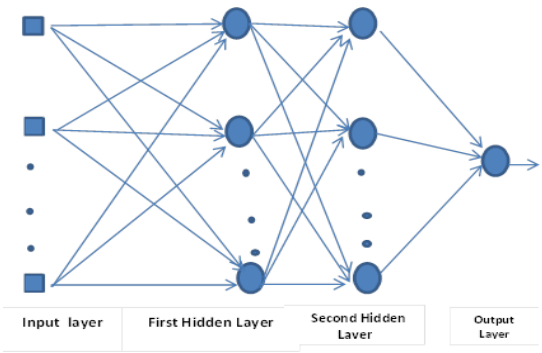

***Fig 1**. MlP Back Propagation Network Architecture*

The following algorithm explains the proposed Multi-Layer Perceptron Back Propagation Network.

Algorithm MLPBPA

Step 1: Input the Heart Diseaes dataset into MLPBPA
Step 2: Set the class attribute (num) as target value and pass onto the MLPBPA
Step 3: Call trainbr, trainlm, traingdx, trainscg with Learngdm, MSE and purelin function to train, adpt train, fine tune performance and to transfer input/output respectively
Step 4: Set the number of default epochs and goal as 10 and 0
Step 5: train the network until the target reached to desired output
Step 6: If (target! =output) reinitialize the network and train network
Step 6.2: Increase the number of neurons
Step 6.3: Increase epochs, goal, number of hidden layers, transfer function and training algorithm
Step 7: else stop the execution.

4 EXPERIMENTAION

Dataset used for experiment is comprised the information of Heart Disease which was taken from the UCI Machine Learning Repository. The data folder actually consisted of 10 different datsets with the variation in number of attributes and missing values. Among them,Cleveland dataset was seemed to be more suitable with less number of 14 attributes (out of 64) with 303 instances. The instances were decreased to 297 when the data was cleaned and preprocessed before it was submitted to the proposed algorihm.The description of the data set is illustrated in Table 1.

***Table 1:**Heart Disease Dataset*

S.No	Attribute	Description	Range
1	*Age*	Patient's Age	29-77
2	*Sex*	1=male; 0=female	0-1
3	*Cp*	Value 1:typical angina Value 2: atypical anginal Value 3: non-anginal pain Value 4: asymptotic	1-4
4	*trestbps*	Resting blood pressure(in mm Hg)	94-200
5	*Chol*	Serum cholesterol in mg/dl	126-564
6	*Fbs*	(Fasting blood sugar .120mg/dl) (1=true; 0=false)	0-1
7	*Restecg*	electrocardiography results Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST Elevation or depression of>0.05mV) Value 2: showing probable or definite left	0-2
8	*Thalach*	Maximum heart rate achieved	71-202
9	*Exang*	Exercise induced angina(1=yes;0=no)	0-1
10	*OldPeak*	ST depression induced by exercise relative to rest	0-6.2
11	*Slope*	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3:down sloping	0-2

12	*Ca*	Number of major vessels (0-3) Colored by fluoroscopy	0-3
13	*Thal*	Normal, fixed defect, reversible defect	3-7

The experiment was carried out in MATLAB 7.0. The dataset were divided into three categories of data as input, target and sample. MLP with Back Propagation Algorithm is trained using input data. All the 13 attributes were taken into the consideration for obtaining the target. Input weights are automatically adjusted using the purelin function for increasing the accuracy of the output. The error is calculated as the difference between the target output and the network output. The goal is to minimize the average of the sum of these errors. The LMSE (Least Mean Square Error) algorithm adjusts theweights and biases of the linear network so as to minimize this mean square error. Fortunately, the mean square error performance index for the linear network is a quadratic function. Thus, the performance index will either have one global minimum, a weak minimum Or no minimum, depending on th characteristics of the input vectors. Specifically, the characteristics of the input vectors determine whether or not a unique solution exists.

$$MSE = \frac{1}{Q}\sum_{k=1}^{Q}(t(k) - a(k))^2$$

In this work, the MLP with Back Propagation is used for Heart Disease prediction. It represent any input-output relationship with a finite number of discontinuities assuming that there are enough neurons in the hidden layer.

5 RESULT AND DISCUSSION

Figure 2. Shows the performances of TRAINBR training function has secured the highest accuracy with 96.30%, which outperforms all other training functions of Feed Forward Back Propagation algorithm. Table 2. Depicts the comparative analysis of Feed Forward Back Propagation Algorithm.

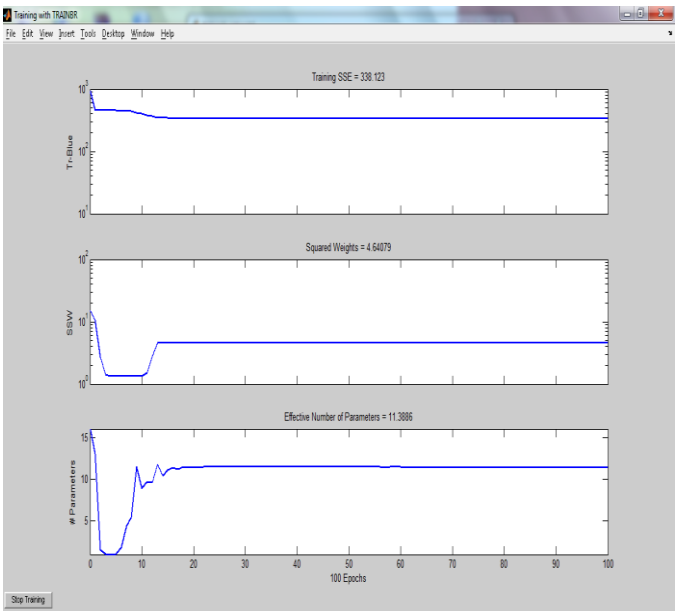

Fig 2:Target Vs Sample Output Data

Table 2: *Compartive Analysis on Feed-Forward Back Propagation Training Function*

S. No	Algorithm	Accuracy
1	TRAINBR	96.2963%
2	TRAINGDX	80.13468%
3	TRAINLM	93.26599%
4	TRAINSCG	82.15488%

6 CONCLUSION

In this paper, a novel approach based on MLP Back Propagation Nerual Network is proposed to predict the Heart Disease. The Proposed system used nearly 13 significant medical attributes for Heart Disease predictions. The experiment displays good performance of the proposed algorithm and was compared to similar approaches over the same dataset. Moreover,the accuracy of the different training functions of Multilayer Perceptron Algorithm were compared to select the best taining function. The TRAINBR algorithm gives more prediction accuracy than other algorithms. The experimental results show the encouraging results that the MLP with proper training algorithm can be an effective tool to predict Heart Disease with improved accuracy.

REFERENCES

[1] Hongmei yan, Yingtao Jiang, Jun Zheng, Chenglin Peng, Qinghui Li,"A Multilayer Perceptron-Based Descision Support System for heart Disease Diagnosis", 30(2006) 272-281

[2] Miss. Chaitrali S. Dangarw, Dr. Mrs. Sulabha S. Apte,"A Data Mining Approach For Prediction of Heart Disease Using Neural Networks" volume 3, Issue 3, October-december (2012), pp.30-40.

[3] YAN Hongmeil, PENG Chenglin1, DING Xiaojun2, XIAO Shouzhong, "Improving the accuracy of heart disease diagnosis with an augmentes back propagation algorithm", june 2003.

[4] Roya Asadi, Norwati Mustapha, Nasir Sulaiman, Nematollaah shiri,"New Supervised Multilayer Feed Forward Neural Network Model to Accelerate Classification with High Accuracy", Vol.33 No.1, pp.163-178

[5] Miss.Manjusha B. Wadhonkar, Prof.P.A. Tijare and Prof. S.N.Saqwalkar,"Classification of Heart disease Datset using Multilayer Feed Forward BackPropogation Algorithm",Volume 2 issue 4, April 2003.

[6] S.Radhimeenakshi, G.M.Nasira"Prediction of Heart Diease using Nerual Network with Back Propagation" Pages:1166-1169.

[7] Sunila,Prahat,Pandy,Nirmal Godara,"Decision Support for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron" Computer Applications (0975-8887) Volume 45-no.8, may2012.

[8] Ilias Maglogiannnisa, Euripidis Loukis,Elias Zafiropoulos,Antonis Stas, "Support Vectors Machinebased Identification of Heart Valve Disease using Heart Sounds", onlinr jaunary 2009.

[9] Todd,R.Reed,Nancy,E.Reed,PeterFrizson,"Heart Sound Analysis for Symptom Detection and Computer-Aided Diagnosis",2004.

[10] Resul Das, Ibrahim Turkoglu, Bdulkadir Sengur,"Diagnosis of Valvular Heart Disease Through Nerual Networks Ensembles", 2008.

[11] Harun Ug, Ahmet Arslam, Ridvan Sara, Ibrahim Turkogu,"Detection of Heart Valve Disease by Using Fuzzy Discrete Hidden Markov Model online", 2008.

[12] Durairaj M, Sivagowry S and Persia A,"An Empircial Study on Applying Data Mining Techniques for the Analysis and Prediction of Heart Disease",2013.

[13] Durairaj M, Revathi V,"Soft Computing Methodology To Measure Heart Sound-A Survey",2015.