

Intelligent and Effective Decision Support System Using Multilayer Perceptron

Sunila Godara*, Nirmal**

*(CSE Department , GJUS&T, Hisar)

** (Department of computer Science, Govt. College for women ,Hisar)

ABSTRACT

Medical science industry has huge amount of data, but most of this data is not mined to find out hidden information in data. Diagnosing of heart disease is one of important issue to develop medical decision support system which will help the physicians to take effective decision. In this paper an attempt is made to study the applicability supervised feed forward neural network. It uses backpropagation learning algorithm with momentum term and augmented learning rate .Missing data are handled using the substituting mean method. Performance of this technique is measured through sensitivity, specificity, accuracy and ROC. We have enhanced the performance of MLP using Dagging approach. In our study 10-fold cross validation method was used to measure the unbiased estimate of the model. Cleveland dataset was used for empirical comparisons.

Keywords - heart disease, artificial neural network, multilayer perceptron, supervised learning, dagging.

1. INTRODUCTION

According to survey, heart disease is leading cause of death in the world every year. Just in United States, almost 930,000 people die and its cost is about 393.5 billion dollars. The heart is the organ that pumps blood, with its life giving oxygen and nutrients, to all tissues of the body. If the pumping action of the heart becomes inefficient, vital organs like the brain and kidneys suffer and if the heart stops working altogether, death occurs within minutes. Life itself is completely dependent on the efficient operation of the heart. Cardiovascular disease is not contagious; you can't catch it like you can the flu or a cold. Instead, there are certain things that increase a

person's chances of getting cardiovascular disease. Cardiovascular disease (CVD) refers to any condition that affects the heart. A number of factors have been shown to increase the risk of developing CVD. Some of these are [1]: Low level of HDL (good) cholesterol

- Family history of cardiovascular disease
- High levels of LDL (bad) cholesterol
- Hypertension
- High fat diet
- Lack of regular exercise
- Obesity

With so many factors to analyze for a diagnosis of heart disease, physicians generally make a diagnosis by evaluating a patient's current test results. Previous diagnoses made on patients with the same results are also examined by physicians. These complex procedures are not easy. Therefore, a physician must be experienced and highly skilled to diagnose heart disease in a patient.

Data mining has been heavily used in the medical field, to include patient diagnosis records to help identify best practices. The difficulties posed by prediction problems have resulted in a variety of problem-solving techniques. For example, data mining methods comprise artificial neural networks and decision trees, and statistical techniques include linear regression and stepwise polynomial regression [2].

It is difficult, however, to compare the accuracy of the techniques and determine the best one because their performance is data-dependent. A few studies have compared data mining and statistical approaches to solve prediction problems. The comparison studies have mainly considered a specific data set or the distribution of the dependent variable.

This paper is organized as follows: A brief overview of related work will be given in section 2. ANN and Dagging are presented in section 3. Section 4 describes about data source. Section 5 and section 6 contains results and conclusion of our study.

2.BACKGROUND

Up to now, several studies have been reported that have focused on cardio vascular disease diagnosis. These studies have applied different approaches to the given problem and achieved high classification accuracies, of 77% or higher, using the dataset taken from the UCI machine learning repository. Here are some examples:

- Robert Detrano's experimental results showed correct classification accuracy of approximately 77% with logistic-regression derived discriminate function [3].
- Zheng Yao applied a new model called R-C4.5 which is based on C4.5 and improved the efficiency of attribution selection and partitioning models. An experiment showed that the rules created by R-C4.5s can give health care experts clear and useful explanations [4].
- Resul Das introduced a methodology that uses SAS base software 9.13 for diagnosing heart disease. A neural networks ensemble method is at the center of this system [5].
- Colombet et al. evaluated implementation and performance of CART and artificial neural networks comparatively with a LR model, in order to predict the risk of cardiovascular disease in a real database [6].
- Engin Avci and Ibrahim Turkoglu study an intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases [7].
- Imran Kurt , Mevlut Ture , A. Turhan Kurum compare performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease [8].
- The John Gennari's CLASSIT conceptual clustering system achieved a 78.9% accuracy on the Cleveland database [9].

3. ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANNs) are commonly known as biologically inspired, highly sophisticated analytical techniques, capable of modeling extremely complex non-linear functions. ANNs are analytic techniques modeled after the processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations from other observations (on the same or other variables) after executing a process of so-called learning from existing data. One of popular ANN architecture is called multi-layer perceptron (MLP)

with back-propagation (a supervised learning algorithm). The MLP is known to be a powerful function approximator for prediction and classification problems. It is arguably the most commonly used and well-studied ANN architecture. Given the right size and the structure, MLP is capable of learning arbitrarily complex nonlinear functions to arbitrary accuracy levels. The MLP is essentially the collection of nonlinear neurons (perceptrons) organized and connected to each other in a feedforward multi-layer structure. Fig 1 shows MLP feed forward Neural Network.

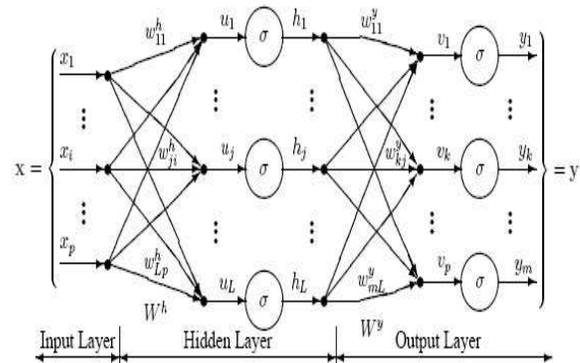


Fig 1: MLP

There is one neuron in the input layer for each predictor variable. In the case of categorical variables, $N-1$ neurons are used to represent the N categories of the variable.

Input Layer — A vector of predictor variable values ($x_1...x_p$) is presented to the input layer. The input layer standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j . The weighted sum (u_j) is fed into a transfer function, σ , which outputs a value h_j . The outputs from the hidden layer are distributed to the output layer.

Output Layer — Arriving at a neuron in the output layer, the value from each hidden layer neuron is

multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j . The weighted sum (v_j) is fed into a transfer function, σ , which outputs a value y_k . The y values are the outputs of the network[12].

$$Y_j = f(\sum W_{ji} X_i)$$

The back-propagation algorithm can be employed effectively to train neural networks; it is widely recognized for applications to layered feed-forward networks, or multi-layer perceptrons. The back-propagation algorithm consists of four steps:

1. Compute how fast the error changes as the activity of an output unit is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

2. Compute how fast the error changes as the total input received by an output unit is changed. This quantity (EI) is the answer from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.

3. Compute how fast the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the activity level of the unit from which the connection emanates.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{dy_j}{dx_j} = EA_j y_j (1 - y_j)$$

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = EI_j y_i$$

4. Compute how fast the error changes as the activity of a unit in the previous layer is changed. This crucial step allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activities of all the output units to which it is connected. So to compute the overall effect on the error, we add together all these separate effects on output units. But each effect is simple to calculate. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$EA_i = \frac{\partial E}{\partial x_i} = \sum_j \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial x_i} = \sum_j EI_j W_{ij}$$

By using steps 2 and 4, we can convert the EAs of one layer of units into EAs for the previous layer. This procedure can be repeated to get the EAs

for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to compute the EWs on its incoming connections

Dagging, This meta classifier creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via majority vote, since all the generated base classifiers are put into the Vote meta classifier. Useful for base classifiers that are quadratic or worse in time behavior, regarding number of instances in the training data.

4. DATA SOURCE

Cleveland cardiovascular disease dataset from UCI To compare these data mining classification techniques repository was used. The dataset has 14 attributes and 303 records. Table 1 below lists these attributes:

Table 1: Attributes of Cardiovascular dataset

| No. | Name | Description |
|-----|----------|---|
| 1 | Age | Age in years |
| 2 | Sex | 1 = male, 0 = female |
| 3 | Cp | Chest pain type (1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic) |
| 4 | Trestbps | Resting blood sugar (in mm Hg on admission to hospital) |
| 5 | Chol | Serum cholesterol in mg/dl |
| 6 | Fbs | Fasting blood sugar > 120 mg/dl (1 = true, 0 = false) |
| 7 | Restecg | Resting electrocardiographic results (0 = normal, 1 = having ST-T wave abnormality, 2 = left ventricular hypertrophy) |
| 8 | Thalach | Maximum heart rate |
| 9 | Exang | Exercise induced angina |
| 10 | Oldpeak | ST depression induced by exercise relative to rest |
| 11 | Slope | Slope of the peak exercise ST segment (1 = upsloping, 2 = flat, 3 = downsloping) |
| 12 | Ca | Number of major vessels colored by fluoroscopy |
| 13 | Thal | 3 = normal, 6 = fixed defect, 7 = reversible defect |
| 14 | Num | Class (0 = healthy, 1 = have heart disease) |

5. RESULTS

These data mining classification model were developed using data mining classification tool Weka version 3.6. Initially Cleveland dataset had 14 attributes and 303 records

A distinguished confusion matrix was obtained to calculate sensitivity, specificity and accuracy. Confusion matrix is a matrix representation of the classification results. Table 2 shows confusion matrix.

Table 2: Confusion Matrix

| | | |
|--------------------|-----------------------|---------------------------|
| | Classified as Healthy | Classified as not healthy |
| Actual Healthy | TP | FN |
| Actual not healthy | FP | TN |

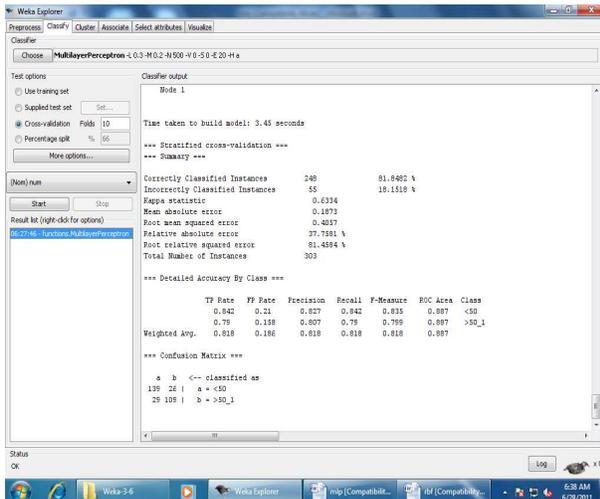
The upper left cell denotes the number of samples classified as true while they were true (i.e., TP), and the lower right cell denotes the number of samples classified as false while they were actually false (i.e., TN). The other two cells (lower left cell and upper right cell) denote the number of samples misclassified. Specifically, the upper right cell denoting the number of samples classified as false while they actually were true (i.e., FN), and the lower left cell denoting the number of samples classified as true while they actually were false (i.e., FP).

Below formulae were used to calculate sensitivity, specificity and accuracy:

Sensitivity = TP / (TP + FN)
 Specificity = TN / (TN + FP)
 Accuracy = (TP + TN) / (TP + FP + TN + FN)

Table 3a and 3b below show confusion matrix for MLP and MLP with Dagging approach.

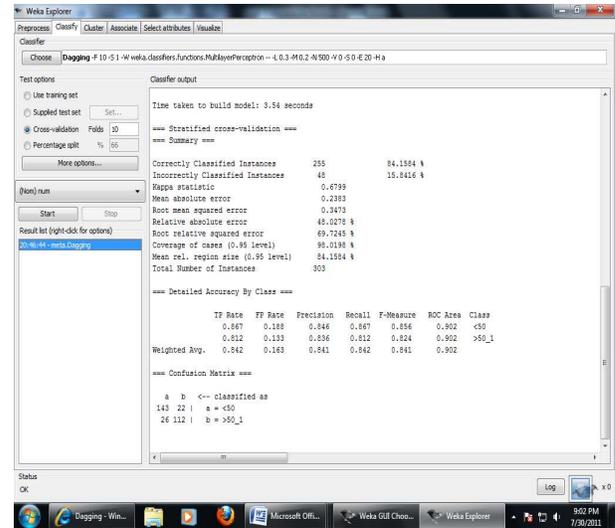
Table 3a: MLP Confusion Matrix



=== Confusion Matrix ===

```
a b <-- classified as
139 26 | a = <50
29 109 | b = >50_1
```

Table 3b: MLP with Dagging Confusion Matrix



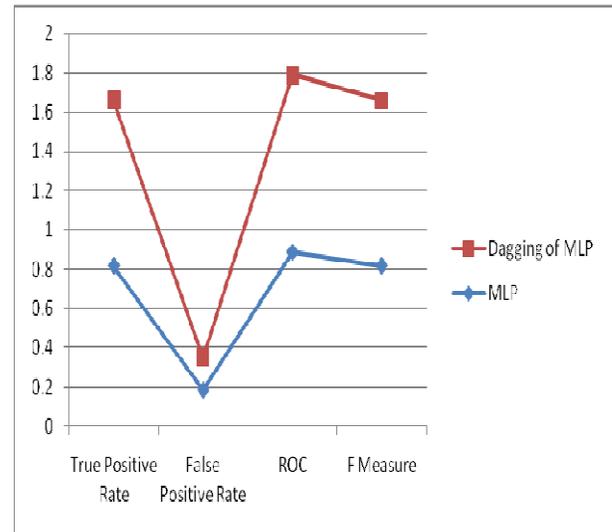
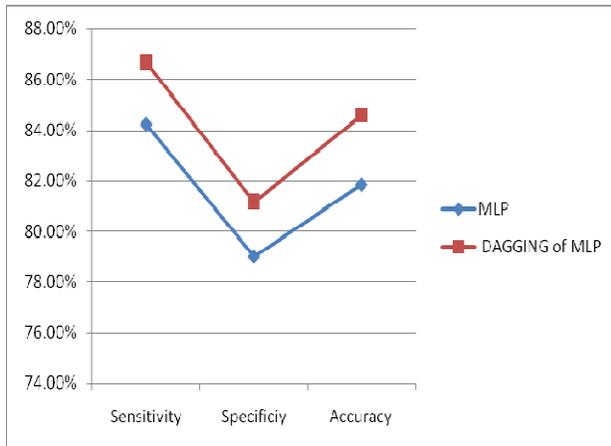
=== Confusion Matrix ===

```
a b <-- classified as
143 22 | a = <50
26 112 | b = >50_1
```

Table 4a shows sensitivity, specificity and accuracy for MLP and MLP with Dagging and Table 4b shows their graphical representation. Table 4a: Sensitivity, Specificity and Accuracy for MLP and MLP with Dagging approach

| | Sensitivity | Specificiy | Accuracy |
|----------------|-------------|------------|----------|
| MLP | 84.24% | 78.99% | 81.85% |
| DAGGING of MLP | 86.67% | 81.16% | 84.58% |

Table 4b: Graphical representation of Sensitivity, Specificity and Accuracy for MLP and MLP with Dagging approach



A Receiver Operating Characteristic (ROC) space is defined by False Positive Rate and True Positive Rate which shows relative trade-off between true positive and false positive.

$$\text{True Positive Rate} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{False Positive Rate} = \text{FP} / (\text{FP} + \text{TN})$$

model will have maximum ROC. ROC value 1.0 represents 100% True Positive Rate and no False Positive Rate which will be ideal case.

Table 5a shows TPR, FPR, ROC, F measure and Time taken to build model by MLP. Table 5b shows their graphical representation.

Table 5a: TPR, FPR, ROC , F measure and Time for MLP and MLP with Dagging approach

| | True Positive Rate | False Positive Rate | ROC | F Measure |
|----------------|--------------------|---------------------|-------|-----------|
| MLP | 0.818 | 0.186 | 0.887 | 0.818 |
| Dagging of MLP | 0.842 | 0.163 | 0.902 | 0.841 |

Table 5a: Graphical representation of TPR, FPR, ROC , F Measure for MLP and MLP with Dagging approach

Our result shows that in all parameters such as Sensitivity, specificity, accuracy and ROC value, Dagging of MLP performs better than MLP.

6.CONCLUSION

In this paper, we have presented a decision support system based on MLP neural network architecture for diagnosing heart disease. Back propagation algorithm was used to train the network. Result have shown that MLP can achieve an accuracy of 81.85% on Cleveland Heart Diseases dataset. Further we have enhanced Sensitivity, Specificity and Accuracy of MLP, using Dagging approach. MLP with Dagging approach had showed good results and attained accuracy of 84.58%. So more applications would bring out the efficiency of this model over other models when interactions or complex nonlinearities exists in the dataset. In future we will try to enhance the performance by changing their learning techniques and performing ensembles. Drawing clinical conclusions from data that are in part collected for decision making should be done with caution. Also TN classification was based on presumed status and not on well defined standards. Finally, the newly identified cases still need to be clinically validated to confirm their positivity

REFERENCES

- [1] Yanwei, X.; Wang, J.; Zhao, Z.; and Gao, Y. (2007). "Combination data mining models with new medical data to predict outcome of coronary heart disease".

Proceedings International Conference on Convergence Information Technology 2007, p 868 – 872.

- [2] Khemphila, A.; Boonjing, V. (2010). “Comparing performance of logistic regression, decision trees and neural networks for classifying heart disease patients”. *Proceedings of International Conference on Computer Information System and Industrial Management Applications 2010*, p 193 – 198.
- [3] Detrano, R.; Steinbrunn, W.; Pfisterer, M. (1987). “International application of a new probability algorithm for the diagnosis of coronary artery disease”. *American Journal of Cardiology*, Vol. 64, No. 3, 1987, p 304-310.
- [4] Yao, Z.; Lei, L.; Yin, J. (2005). “R-C4.5 Decision tree model and its applications to health care dataset”. *Proceedings of International Conference on Services Systems and Services Management 2005*, p 1099-1103.
- [5] Das, R.; Abdulkadir, S. (2008). “Effective diagnosis of heart disease through neural networks ensembles”. *Elsevier, 2008*.
- [6] Colombet, I.; Ruelland, A.; Chatellier, G.; Gueyffier, F. (2000). “Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression”. *Proceedings of AMIA Symp 2000*, p 156-160.
- [7] Avci, E.; Turkoglu, I. (2009). “An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases”. *Journal of Expert Systems with Application, Vol. 2, No. 1, 2009*, p 2873-2878.
- [8] Kurt, I.; Ture, M.; Turhan, A. (2008). “Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease”. *Journal of Expert Systems with Application, Vol. 3, 2008*, p 366-374.
- [9] Gennari, J. (1989). “Models of incremental concept formation”. *Journal of Artificial Intelligence*, Vol. 1, 1989, p 11-61.
- [10] Cohen, W. (1995). “Fast effective rule induction”. *Proceedings of International Conference on machine Learning 1995*, p 1-10.
- [11] Chau, M.; Shin, D. (2009). “A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms”. *Proceedings of IEEE*