Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DIAGNOSIS OF HEART DISEASE USING NEURAL NETWORKS-COMPARATIVE STUDY OF BAYESIAN REGULARIZATION WITH MULTIPLE REGRESSION MODEL

[1] K.SAI KRISHNASREE, [2] M.R.NARASINGA RAO

[1]MTech Student, Department of Computer Science and Engineering, KL University, Guntur, A.P, India

[2] Professor, Department of Computer Science and Engineering, KL University, Guntur, A.P, India

E-mail: [1]saikrishnasree36@gmail.com, [2]ramanarasingarao@kluniversity.in

ABSTRACT

Heart disease is one of the significant reasons of death and the progress of which is rampant all over the globe. Blood vessels carry blood with oxygen to all the cells in the body. It is a common reason that, Cholesterol and other substances can be deposited in blood vessels which block blood vessels and that no blood and oxygen can get to heart. This leads to heart disease. Several works have been made to predict the heart disease in different methods. The main aim of this paper is to predict heart disease using Multiple Regression and Bayesian Regularization methods and compare the results of these models. Multiple Regression is one of the strong model used for prediction and it shows the association between input variables and output variable. It predicts the output variable based on the relationship between one or more input variables and target variable. Bayesian regularization is a statistical model which process nonlinear dataset. It increases the generalization capability and decreases squared errors. Bayesian regularization works on with large inputs efficiently. The results are calculated using Multiple Regression and Bayesian Regularization methods and predicted the heart disease. The results of Multiple Regression and Bayesian Regularization are compared and it is observed that the results generated from Bayesian Regularization are more accurate than multiple regression model.

Keywords: *Bayesian Regularization, Multiple regression, Heart disease, Artificial Neural Network (ANN), Prediction.*

1. INTRODUCTION

Heart Disease generally refers to cardiovascular disease which causes death. In [1], by 2030 the expected death rate per year is more than 23.6 million. [2] Blood is transferred to all the cells in the body using blood vessels. High cholesterol creates a fat deposit in blood vessels, which reduces pumping of blood with oxygen to heart. Cardiovascular disease mainly causes when the flow of blood to heart through blood vessels gets blocked. [2], Blood pressure, cholesterol and diabetes are the main reason that causes cardiovascular disease. The common risk factors that cause heart disease are chest pain, blood pressure, cholesterol, blood sugar, heart rate, angina etc. According to National Diabetes Clearing house, people suffering from diabetes have more chances of attacking heart disease. There are many data mining techniques and neural network models used to predict heart disease. In the proposed paper, the main objective is to predict the heart disease using Bayesian regularization and multiple regression and comparing the results of these two models. [4], Neural network generates high-quality results for large data samples and multiple regression is good at small data samples.

Multiple Regression model is a strong tool in predicting the unknown value taking into account one or more input variables. Multiple regression allows any number of input variables. These multiple independent variables are also known as predictors, [3]. [5], If predictors have correlation with each other, then it leads to data redundancy. [3], Based on these multiple predictors, the unknown value is predicted. The unknown value is dependent on the multiple independent predictors. Hence multiple regression, models the association between multiple input variables and output variable. The output variable is dependent on input variables. To predict heart disease multiple predictor variables are required. Selecting multiple predictor variables is a difficult task.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Artificial neural networks are efficiently used for medical predictions. ANN is the solution for many problems that cannot be solved using data mining techniques. ANN is utilized for extensive nonlinear datasets. [6], The main aim of any neural network is to reduce the difference between actual output and predicted output. Training algorithm like Bayesian regularization is used to adjust the network to generate accurate output. Bayesian Regularization reduces squared errors and generalizes the network in an accurate manner.

2. RELATED WORK

Many Researches have been implemented on diagnosing heart disease. [8], Durairaj and Revathi V proposed a paper on "Prediction Of Heart Disease Using Back Propagation MLP algorithm". They have proposed that MLP in conjunction with a training algorithm is applied to predict heart disease. The results of different training algorithms are compared. Bayesian Regularization increased the accuracy in prediction. [9], Manisha Dinesh Bhosla and T.P.Singh proposed a paper on "Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle". Feed Forward algorithms (like Scaled conjugate gradient, Levenberg-Marquardt method, Broyden-Fletcher-Golfarb-Shanno quasi-Newton algorithm and Bayesian regularization) and multiple linear regression are used for milk yield prediction. To evaluate the best model, R square and RMSE values are calculated. BR is considered as best calculation for milk yield forecast. [10], Harwinder Kaur and Dalwinder singh Salaria has proposed a paper on "Bayesian Regularization Based Neural Network Tool for Software Effort Estimation". This paper deals to increase the effort estimation in software organizations using neural networks training algorithms like Levenberg-Marquardt method, Bayesian regularization and Back Propagation. It predicts cost estimations for successful control of time and schedule in project development. Similar efforts were made from Deepthi Gurram , M R Narasingarao using MLP Neural Network model and factor analysis for predicting the risk level of heart disease[11]. I.F. Barcelos Tronto, J.D. Simoes da Silva, N. SantAnna proposed a model on predicting time and budget required for the development of software using Artificial Neural Network and linear regression. The performance of Artifical Neural Network is better than linear regression in prediction [12].

Chaitrali S. Dangare, Dr. Sulabha S. Apte proposed a paper on heart disease prediction using data mining and neural networks. Multilayer perceptron neural network with back propagation is used to get better results to diagnose the disease than using data mining techniques [13].
Parsinejad Shahbaz, Bagheri Ahmad, Ebrahimi Atani Reza and Javadi Moghaddam Jalal proposed a paper on predicting stock market. There are several algorithms like regression, genetic algorithms, time series and fundamental analysis used to predict. To get more efficient results and profit, an appropriate non linear method like Artificial Neural Network is used to forecast stock market [14]. Sunila, Prabhat Panday, Nirmal Godara proposed a paper on diagnosing heart disease using improved multilayer perceptron and multiple layer perceptron. In improved multilayer perceptron dataset is divided into multiple datasets and the individual datasets are solved and results are combined to get accurate results in predicting heart disease. The effectiveness is measured using accuracy, specificity and sensitivity [15].

Aspy Palia proposed a paper on forecasting sales online using Multiple Regression. Based on the past behavior of sales the predictor variables are identified to forecast future sales [16]. Suchithra, Dr.P.Uma Maheswari proposed a paper on diagnosis of heart disease using J48 Decision tree algorithm. Information like symptoms, stage and treatment are gathered from an expert. Decision algorithm is applied to predict the disease and to give treatment for the disease based on symptoms [17]. There are several papers used to diagnose heart disease comparing with other papers, the aim of the present paper is to predict the heart disease using Multiple Regression and Bayesian Regularization. Bayesian regularization works accurately when there is noisy data than multiple regression. So hence for accurate prediction Bayesian regularization is used.

3. METHODOLOGY

In the proposed system, heart disease is predicted using multiple regression and Bayesian regularization. [4], Both the models are completely different from one another. The weights in multiple regression model depends on multiple input variables and in Bayesian regularization depends on the input layer, hidden layer and output layer. Multiple regression and Bayesian regularization predicts whether the person has a chance to get heart disease based on input parameters.

Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In [7] [4], Regularization network mainly contains 3 layers- input, hidden and output layers. Input layer includes independent input variables of the heart disease dataset. Hidden layer includes the non linear training data. Each input is connected to the hidden node. The output layer is a single unit which is connected with hidden layer. Finally, the output layer contains the predicted value of heart disease. To predict heart disease, consider different input parameters like age, gender, cholesterol, fasting blood pressure, resting blood pressure, angina etc. Too many parameters in the data, the network generates noisy results and it causes more squared errors this leads to poor prediction. This problem is referred as over fitting. So to avoid over fitting, Bayesian regularization has been used for the purpose. [10], BR model deals with large number of inputs; dealing with large input values produces best prediction results. BR model update weights in the network based on the input values. It helps to avoids squared errors and improves generalization.

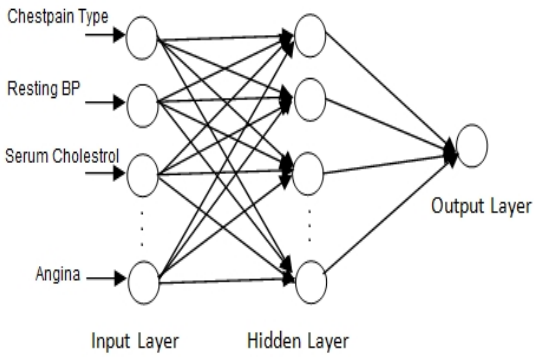

Figure1: Neural Network

3.1 Choosing Input Variables

Selecting input variables is a difficult task. Input variables have been considered from a domain expert. Select an independent input variable and perform multiple regression. The R^2 value (squared error) of the selected input variable should be high. Each time you add input variable, the R^2 value increases. Eliminate adding input variable by adding which R^2 value remains constant. R^2 value increases with increase in number of independent input variables. The heart disease dataset is collected from the UCI library from Internet. The main challenge is to select valid input variables to predict the heart disease effectively using multiple regression and Bayesian regularization. Input variables considered for predicting the heart disease are gender, age, chest pain type, BP, serum cholesterol, fasting blood sugar, resting ECG, Maximum heart rate, angina, old peak, ST depression, number of vessels and thal. The output variable predicts the heart disease. The output of 1 indicates presence of heart disease and 2 indicates non-presence of heart disease.

3.2 Using Multiple Regression

In Multiple Regression model, 100 records of patients are used to train the model. 13 input variables are used to train the model. [5], These input variables should not have any relationship with each other. If they correlate with each other, it increases standard errors and multicollinearity problems. The multiple regression equation for dependent variable X is as follows

$$X= a_0+a_1I_1+a_2I_2+a_3I_3+.....a_kI_k \quad (1)$$

In this equation, a_0 is the intercept, $a_1,a_2... a_k$ are the regression coefficients of the independent variables, $I_1,I_2...I_k$ are the input variables and X gives the predicted value of heart disease. Regression coefficients are calculated to reduce sum of squared errors using

$$\sum_{i=1}^{N} X^1 - X \quad (2)$$

In the above equation X^1 is the estimated value of regression equation. Multiple regression focuses on the relationship between dependent variable and multiple predictor variables. So thus heart disease prediction depends on all the 13 input variables. In the proposed system, Regression model has been applied to 100 samples of data. If the p-value of independent variables are <0.05 then the regression model is good fit. The above table shows the multiple regression statistics in prediction of heart disease.

Table 1: Statistics

Multiple R	0.7295
R Square	0.5321
Adjusted R Square	0.4614
Standard Error	0.3651
Observations	100

[3], R square lies between 0 and 1. The model is said to be good predictor when R square value is close to 1. Multiple R value shows the association of output variable and multiple input variables. In this case the correlation between the prediction of heart disease and input variables is 0.7295. Adjusted R square is used when there are multiple input variables. In this case there are 13 input variables ($I_1,I_2..I_{13}$).

Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.3 Using Bayesian Regularization

There is no specific procedure adopted using neural networks for predicting heart disease. MATLAB is used to work with Bayesian Regularization. In [7], it gives best prediction when there are large number of inputs and output. In the proposed system, to predict and examine performances, give test data as input to the network. Create work space for the heart disease prediction. 100 records of patients and 15 neurons are used to train the neural network. Separate the data as input, sample and target. Function trainbr is used to implement Bayesian regularization. Stop the training when number of repetitions (epochs) reached.

Steps used to implement Bayesian regularization:

1. Define the input and target data in the work space.
2. Determine the amount of neurons and function 'trainbr' used.
   network=fitnet(15,'trainbr');
3. Specify the ratio of data used for training, validation and testing.
   network.divideParam.trainRatio=.7;
   network.divideParam.valRatio=.15;
   network.divideParam.testRatio=.15;
4. Train the neural network with input and target data.
   [network,pr]=train(network,input',target');
5. Define the sample data.
6. If the output is not desired then again train the neural network with more input data and change the number of neurons used until we get desired output.

To get highest accuracy retrain the network. It gives different solutions for different input data and divisions of data for validating, training and testing. Bayesian regularization predicts heart disease with an accuracy of 91%.

4. RESULTS AND DISCUSSION

Both the models reduce squared difference between actual value and predicted value. Below table shows the % accuracy of multiple regression and Bayesian regularization models. The model is trained with 100 records of data. The output of sample data using both the models is compared. Multiple regression model predicts heart disease with 86% and the Bayesian regularization model predicts with 91%.

Table 2: Comparison table of heart disease diagnosis

No of Samples	% Accuracy of MR	% Accuracy of BR
1	89	95
2	98	100
3	74	74
4	69	68
5	92	97
6	88	93
7	89	100
8	93	100
9	82	96
10	94	88

The performance of the present work is measured using regression plots. Figures below are the regression plots of Bayesian regularization in heart disease prediction. The regression plot generally shows the association between the output and the target. Dashed line represents best results. The solid line represents regression line between targets and outputs.

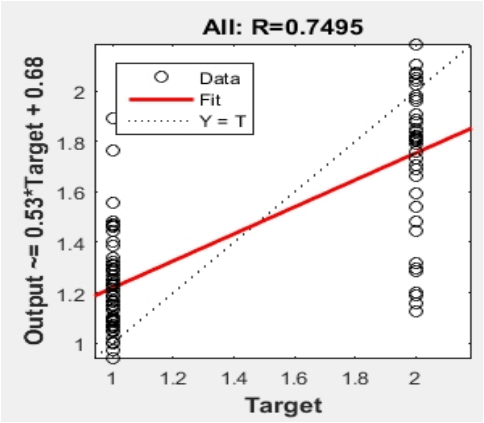

Figure 2: Regression Plot for heart dataset

Coefficient (R-Value) in the regression shows the correlation between output and target. If R-Value is 1 then there is good correlation between output and target. If R value is closer to zero, it means there is no correlation between output and target and gives inaccurate results. In this case R-Value for training is 0.7726, the value is closer to 1.To get more accurate results reinitialize once again and train the network. The limitation of this paper is it is done with limited dataset. If the result of the prediction is not accurate then train the system with more data and reinitialize the system multiple times.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

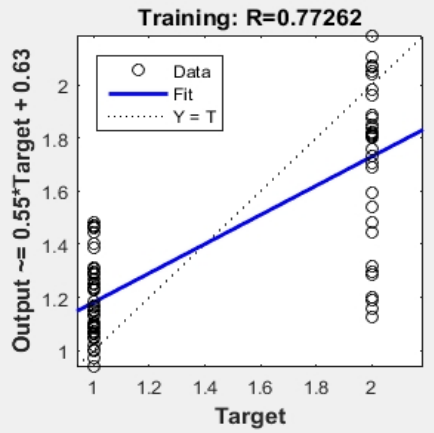

Figure 3: Regression Plot for Training data

5. CONCLUSION

The proposed system mainly compares the output of multiple regression and Bayesian regularization model while predicting heart disease. Hundred patient records have been used to train the neural network and multiple regression. 13 attributes are taken as input values. Multiple regression mainly shows the relationship between input variables and target variable. Using this correlation multiple regression predicts heart disease of a patient. Using Bayesian regularization train the network with 100 records of the patients. Trainbr is a function used by the network to predict heart disease. The results of both the models are compared. Multiple regression model predicted with an accuracy of 86% and Bayesian regularization with 91%. The performance of Bayesian regularization is better than multiple regression. This experiment successfully predicts heart disease using Bayesian regularization. Bayesian regularization works accurately with large non linear dataset but multiple regression model works less accurately with non linear dataset. So thus Bayesian regularization is a powerful tool in making prediction even with large noisy data.

REFRENCES:

[1] Heart Disease and Stroke Statistics- At-a-Glance [online], Availability: https://www.heart.org/idc/groups/ahamah-public/@wcm/@sop/@smd/documents/downloadable/ucm_470704.pdf.

[2] Causes and Risks Of Heart Disease [online], Availability: www.healthline.com/health/heart-disease/causes-risks#Overview1.

[3] Multiple Regression Analysis [online], Availability: https://explorable.com/multiple-regression-analysis.

[4] Susan L. King, "Neural Networks vs. Multiple Linear Regression for Estimating previous diameter", 12th central hardwood forest conference; 1999 February.

[5] Residual Analysis and Multiple Regression [online], Availability: http://www-personal.umich.edu/~gonzo/coursenotes/file7.pdf.

[6] K.K. Aggarwal, Yogesh Singh, Pravin Chandra and Manimala Puri , "Bayesian Regularization in a Neural Network Model to Estimate Lines of code using function points" , journal of Computer Science, Volume 1, Issue , Pages 505-509, 2005.

[7] Simon Haykin, "Neural Networks a Comprehensive Foundation", Second Edition, Pearson Education, 2006.

[8] Durairaj M, Revathi V, "Prediction of Heart Disease Using Back Propagation MLP Algorithm", International Journal Of Scientific & Technology Research Volume 4, Issue 08, August 2015.

[9] Manisha Dinesh Bhosale and T.P.singh, "Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle", Current Science, Vol. 108, No. 12, 25 June 2015.

[10] Harwinder Kaur, Dalwinder Singh Salaria, "Bayesian Regularization Based Neural Network Tool for Software Effort Estimation", Global Journal of Computer Science and Technology Neural & Artificial Intelligence, Volume 13, Issue 2, Year 2013.

[11] Deepthi Gurram, M.R.Narasinga Rao, "A Decision Support System for predicting Heart Disease using Multilayer Perceptron and Factor Analysis", International Review on Computers And Software, Vol:10,No:8,2015,PageNo:799-804.

[12] I.F. Barcelos Tronto, J.D. Simões da Silva, N. Sant'Anna, "Comparison of Artificial Neural Network and Regression Models in Software Effort Estimation", INPE ePrint, Vol.1, 2006.

[13] Miss. Chaitrali S. Dangare1 , Dr. Mrs. Sulabha S. Apte," A Data Mining Approach For Prediction Of Heart Disease Using Neural Networks", IJCET, Volume 3, Issue 3, October - December (2012).

[14] Parsinejad Shahbaz, Bagheri Ahmad, Ebrahimi Atani Reza and Javadi Moghaddam Jalal, "Stock Market Forecasting Using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Artificial Neural Networks", European Online Journal of Natural and Social Sciences 2013, Vol.2, No.3 Special Issue on Accounting and Management.

[15] Sunila, Prabhat Panday, Nirmal Godara, "Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron", International Journal of Computer Applications, Volume 45– No.8, May 2012.

[16] Aspy Palia," Online Sales Forecasting With The Multiple Regression Analysis Data Matrices Package", Developments In Business Simulation And Experiential Learning, Volume 31, 2004.

[17] Suchithra , Dr.P.Uma Maheswari, "Clinical Decision Support System for Diagnosing Heart Disease", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Special Issue 3, July 2014.