Logistic Regression Versus Neural Networks: The Best Accuracy in Prediction of Diabetes Disease

Musavir Hassan, Muheet Ahmad Butt and Majid Zaman Baba
PG Department of Computer Science, University of Kashmir, Srinagar , India

Abstract **- To derive actionable insights from vast amount of data in an intelligent fashion some techniques are used called machine learning techniques. These techniques support for predicting disease with correct case of training and testing. To classify the medical data logistic regression and artificial neural networks are the models to be selected. Today in world a major health problem is Diabetes Mellitus for which many classification algorithms have been applied for its diagnoses and treatment. To detect diabetes disease in early stage it needs greatest support of machine learning, since it cannot be cured and also brings great complication to our health system. In this paper, we establish a general framework for explaining the functioning of Artificial Neural Networks (ANNs) in binomial classification and implement and evaluate the variants of Back propagation algorithm (Standard Back Propagation, Resilient - Back propagation, Variable Learning Rate, Powell-Beale Conjugate Gradient, Levenberg Marquardt, Quasi-Newton Algorithm and Scaled Conjugate Gradient) using Pima Indians Diabetes Data set from UCI repository of machine learning databases. We also compare Artificial Neural Networks (ANNs) with one of the conventional techniques, namely logistic regression (LR) to predict diabetic disease decisions.**

Keywords: **Artificial Neural Networks (ANNs), logistic regression (LR), Quasi-Newton Algorithm, Resilient - Back propagation, Standard Back Propagation.**

I. INTRODUCTION

In recent years, electronic health records in modern hospitals and medical institutions become larger and larger to improve the quality of patient care and increase the productivity and efficiency of health care delivery. So methods for efficient computer based analysis are needed due to the inefficiency of traditional manual data analysis. Machine learning technique has been a tremendous support for making prediction of a particular system by training. In recent year's machine learning has been the evolving, reliable and supporting tool in medical domain. Due to recent advances in machine learning, medical analysis increases diagnostic accuracy, reduces cost and reduces human resource.

Diabetes is the chronic and progressive metabolic disorder that is caused due to increase in blood sugar, either because the pancreas does not produce enough insulin, or because cells don't respond to the insulin that is produced. The number of people suffering from diabetes mellitus (DM) [1,2,3] is increasing each year, which is predicted to reach 366 million by 2030 [4] , causing disabilities, economic hardship and even death. It is becoming more and more important to diagnose DM for its early stage also known as Impaired Glucose Regulation (IGR). Until now, the standard method to diagnose DM in many hospitals is fasting plasma glucose (FPG). FPG test is performed by analysing the patient's blood glucose level after the patient has gone at least 12 hours without taking any food. This method is accurate, but inconvenient. The blood required detecting method can be consider invasive and slightly painful, and even has a risk of infection (piercing process).

Diabetes can be classified into three categories diabetes 1, diabetes 2 and gestation diabetes. No insulin secretion in body or less amount of insulin can cause Diabetes 1 and is majorly found in young children, teens and young adults. They are mainly caused due to little insulin I in their body. This type of diabetes needs insulin to be injected in their body. These types of patients are called Insulin Diabetes Dependent patients (IDDM). Diabetes type2 is cause due to resistance less in insulin. Type2 diabetes is majorly found in adults but now found in younger people also. This type of patients has insulin in their body which is not sufficient. Major cause for type2 diabetes is high obesity rate, majorly when BMI is greater than 25 then there exist greater percentage of risk. These types of patients are Non –Insulin Dependent patients (NIDDM). Gestations Diabetes is cause during pregnancy period. This type of diabetes can be cured after birth of child. During this type of diabetes if proper treatment is not followed there is heavy chance to change into type2 diabetes.

II. DATA AND METHODS

A. DATA

Pima Indian Diabetes Database

PIMA Indian diabetes database is a data set that was obtained from the UCI Repository of Machine Learning Databases [5]. The data set was selected from a larger data set held by the National Institutes of Diabetes and Digestive and Kidney Diseases. The data in this database include database are Pima-Indian women at least 21 years old and living near Phoenix, Arizona, USA. The diagnostic, binary-

valued variable investigated is whether the patient shows signs of diabetes according to World Health Organization criteria. The binary response variable takes the values '0' or '1', where '1' means a positive test for diabetes and '0' is a negative test for diabetes. There are 268 (34.9%) cases in class '1' and 500 (65.1%) cases in class '0'. There are eight clinical findings:

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index
7. Diabetes pedigree function
8. Age (years).

A brief statistical analyse is given in Table 1.

TABLE I BRIEF STATISTICAL ANALYSE OF PID DATABASE

Attribute Number	Mean	Standard Deviation	Min/Max
1	3.8	3.4	0/17
2	120.9	32.0	0/199
3	69.1	19.4	0/122
4	20.5	16.0	0/99
5	79.8	115.2	0/846
6	32.0	7.9	0/67.1
7	0.5	0.3	0.078/2.42
8	33.2	11.8	21/81

As can be seen from Table 1, value range between the attributes is high. A normalisation process is performed on the data to overcome this problem and to get a better result. Normalised values are given in Table 2.

TABLE II NORMALISED STATISTICAL VALUES OF PID DATABASE

Attribute Number	Mean	Standard Deviation	Min/Max
1	3.8	3.4	0/17
2	12.09	3.2	0/19.9
3	6.91	1.94	0/12.2
4	2.05	1.6	0/9.9
5	0.798	1.152	0/8.46
6	3.20	0.79	0/6.71
7	5	3	0.078/24.2
8	3.32	1.18	2.1/8.1

B. Methods

Classification has emerged an important decision making tool. It has been used in variety of applications including credit scoring, prediction of events like credit card usage and as a tool in medical diagnosis. Unfortunately, while several classification procedures exist, many of the current methods fail to provide adequate results. Data classification can be done in a two-step process. The learning step in which a classification model is constructed and a classification step in which the model predicts class labels for given data. In the first step, a classifier is built describing a predetermined set of data classes or concepts. This step is also called training phase where classifier is constructed by learning from a training set that consists of database tuples and their associated class labels($X= x_1, x_2,\ldots,x_n$).

There are two different approaches to data classification: the first considers only a dichotomous distinction between the two classes, and assigns class labels 0 or 1 to an unknown data item. The second attempts to model P (y|x); this yields not only a class label for a data item, but also a probability of class membership. The most prominent representatives of the first class are support vector machines. Logistic regression, artificial neural networks, k-nearest neighbours, and decision trees are all members of the second class, although they vary considerably in building an approximation to P (y|x); from data. Some details on these models, including a comparison on their respective advantages and disadvantages, are given below. Currently, logistic regression and artificial neural networks are the most widely used models in biomedicine, as measured by the number of publications indexed in MEDLINE: 28,500 for logistic regression, 8500 for neural networks, 1300 for k-nearest neighbours, 1100 for decision trees, and 100 for support vector machines.

Decision tree

A decision tree comprises of a set of tree-structured decision tests functioning in a divide-and-conquer way. This algorithm recurrently splits the data set according to a standard that maximizes the separation of the data, resulting in a tree-like structure [6, 7]. Based on several input features decision tree predicts the value of a target label. The feature value pair with the largest information gain is selected for the split; this means that at each split, the decrease in entropy due to this split is maximized. Each interior node (decision node) specifies a test to be carried out on a single feature and corresponds to one of the input features. All the possible outcomes of the test of that input feature are represented by the edges to children and indicate the path to be followed. The value of the target label is indicated by leaf node, given the values of the input features represented by the path from the root to the leaf. A tree can be trained by splitting the source set into subsets based on testing a single feature. On each resulting subset this process is repeated in a recursive manner called recursive partitioning. The recursion is completed when the subset of the training patterns at a given node has the same label or when splitting no longer adds value to the predictions. DT is also simple to

understand, works well with hard data, and is easily combined. [8]. A major disadvantage of decision trees is given by the greedy construction process: at each step, the combination of single best variable and optimal split-point is selected; however, a multi-step look ahead that considers combinations of variables may obtain different (and better) results. A further drawback lies in the fact that continuous variables are implicitly discretized by the splitting process, losing information along the way. Compared with the other machine learning methods mentioned here, decision trees have the advantage that they are not black-box models, but can easily be expressed as rules. In many application domains, this advantage weighs more heavily than the drawbacks, so that these models are widely used in medicine.

Logistic regression

Logistic regression scrutinises the relationship between a binary outcome (dependent) variable such as presence or absence of disease and predictor (explanatory or independent) variables such as patient demographics or imaging findings [9]. For example, the presence or absence of diabetes within a specified time period might be predicted from knowledge of the patient's age (years), diastolic blood pressure (mm Hg), triceps skin fold thickness (mm), body mass index etc. The outcome variables can be both continuous and categorical. If X_1, $X_2 \ldots X_n$ denote n predictor variables (e.g., age (years), diastolic blood pressure (mm Hg), and so on), Y denotes the presence($Y = 1$) or absence ($Y = 0$) of disease, and p denotes the probability of disease presence (i.e., the probability that $Y = 1$), the following equation describes the relationship between the predictor variables and p:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n \quad (1)$$

Where β0 is a constant and β1, β2… βn are the regression coefficients of the predictor associated with each explanatory variable X_1, $X_2 \ldots X_n$. The dependent variable is the logarithm of the odds, which is the logarithm of the ratio of two probabilities: the probability that a disease outbreak will occur divided by the probability that it will not. The regression coefficients are estimated from the available data. The probability of disease presence p can be estimated with this equation.

The contribution of the corresponding predictor variable to the outcome is described by each regression coefficient. The odds ratio of the predictor variable determines the effect of the predictor variables on the outcome variable, which represents the factor by which the odds of an outcome change for a one-unit change in the predictor variable. The odds ratio is estimated by taking the exponential of the coefficient (e.g., exp [β1]). For example, if β1 is the coefficient of variable X_{dp} ("diastolic blood pressure"), and p represents the probability of diabetes, exp (β1) is the odds ratio corresponding to the diastolic blood pressure. The odds ratio in this case represents the factor by which the odds of having diabetes increase if the patient has a diastolic blood pressure and all other predictor variables remain unchanged. Logistic regression models generally include only the variables that are considered "important" in predicting an outcome. With use of P values, the importance of variables is defined in terms of the statistical significance of the coefficients for the variables. The significance criterion $P \leq .05$ is commonly used when testing for the statistical significance of variables; however, such criteria can vary depending on the amount of available data. For example, if the number of observations is very large, predictors with small effects on the outcome can also become significant. To avoid exaggerating the significance of these predictors, a more stringent criterion (e.g., $P \leq .001$) can be used. Significant variables can be selected with various methods. Although different techniques might give different regression models, they are often very similar. The maximum-likelihood method is used to estimate the coefficients,$\beta_1,\ldots,\beta_n$ in the logistic regression.

Artificial Neural Networks

ANNs are computer models patterned after the structure of biologic neural networks. They consist of highly interconnected nodes, and the connection between these neurons determines the overall ability to predict outcomes [10]. The processing elements (the units analogous to biological neurons) are organised into groups called layers. To produce the single output ANNs simulate neural processes by summing negative (inhibitory) and positive (excitatory) inputs. [11]. There are different ways in which the neurons are connected and inputs are processed, we will focus on "feed forward" networks, the most commonly used ANNs in medical research. Figure 1 illustrates the generic structure of an ANN. A typical ANN includes a sequence of nodes arranged in three layers (input, hidden, and output layers). The node in the input layer is called an input node that represents an input that is used as a predictor of the outcome. The single node in the output layer (output node) represents the predicted outcome (e.g., probability of diabetes). The inputs and the output of an ANN correspond to the predictor variables and the outcome variable Y, respectively, in a logistic regression model. The nodes in the hidden layer (hidden nodes) contain intermediate values calculated by the network that do not have any physical meaning. The hidden nodes allow the ANN to model complex relationships between the input variables and the outcome. The ANN in Figure 1 has N input nodes, K hidden nodes, and only one output node.

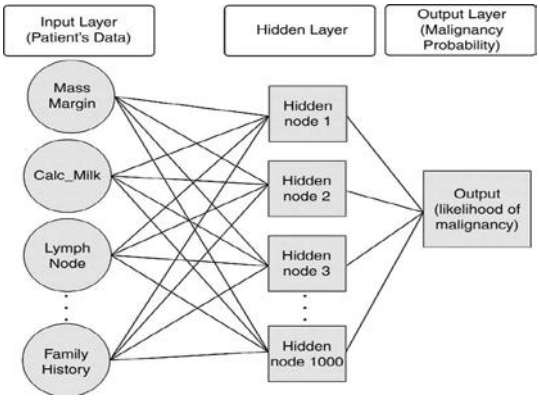


Fig. 1Chart illustrates the generic structure of an ANN

In above figure the connection weights represented by arcs connect the nodes in different layers. The connection weights contain the "knowledge" representing the relationships between variables and correspond to the coefficients in a logistic regression model. By strengthening (increasing) or weakening (decreasing) the values of these connection weights ANNs "learn" the relationships between input variables and the effects they have on outcome. The procedure of estimating the optimal weights that generate the most reliable outcomes is called learning or training [12]. Training an ANN is similar to reckoning parameters in a logistic regression model; however, an ANN is not an automated logistic regression model because the two models use different training algorithms for parameter estimation. There are several algorithms for training ANNs, the most popular of which is back propagation. The back propagation algorithm is based on the idea of adjusting connection weights to minimize the discrepancy between real and predicted outcomes by propagating the discrepancy in a backward direction (i.e., from the output node to the input nodes).The variants of back propagation algorithm are Standard Back Propagation, Resilient - Back propagation, Variable Learning Rate, Powell-Beale Conjugate Gradient, Levenberg-Marquardt, Quasi-Newton Algorithm and Scaled Conjugate Gradient

Standard Back Propagation

The most common method for training a neural network is Back propagation. It is a method in which the weights in an artificial neural network are used to calculate the gradient of the loss function. It is commonly used as a part of algorithms that optimize the performance of the network by adjusting the weights, for example in the gradient descent algorithm. It is also called backward propagation of errors.

The back propagation learning algorithm can be divided into two phases: propagation and weight update.

Phase 1: Propagation

Each propagation involves the following steps:

1. To generate the network's output values forward propagation of a training pattern's input through the neural network
2. To generate the deltas (the difference between the targeted and actual output values) of all output and hidden neurons backward propagation of the propagation's output activations through the neural network using the training pattern target.

Phase 2: Weight update

For each weight, the following steps must be followed:

1. The weight's output delta and input activation are multiplied to find the gradient of the weight.
2. A ratio (percentage) of the weight's gradient is subtracted from the weight.

Resilient - Back propagation

Resilient propagation, in short, RPROP is one of the fastest training algorithms available that can be used to train a neural network. This is similar to the more common (regular) back-propagation but it has two main advantages over back propagation: First, training with Rprop is often faster than training with back propagation. Second, there is no need of specifying any free parameter values, as in back propagation values for the learning rate (and usually an optional momentum term) are needed. The main disadvantage of Rprop is that it's a more complex algorithm to implement than back propagation.

Variable Learning Rate (traingda, traingdx)

Learning rate has a great impact on training of Back propagation neural network, the speed and accuracy of the learning process i.e. updating the weights, also depends the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. The algorithm will take too long to converge if the learning rate is too small. Some variations in the training procedure used by traingd are required in an adaptive learning rate. The initial network output and error are calculated first. New weights and biases at each epoch are calculated using the current learning rate. New outputs and errors are then calculated.

Levenberg-Marquardt algorithm

Levenberg-Marquardt algorithm was specially designed to minimize sum-of-square error functions and to approach second-order training speed without having to compute the Hessian matrix [13]. The Hessian matrix can be approximated as

$$H = J^t J \qquad (2)$$

When the performance function has the form of a sum of squares.

The gradient can be computed as

$$g = J^t e \qquad (3)$$

Where J is the Jacobian matrix computed through a standard back propagation and e is a vector of network errors. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix using the following update:

$$x_{k+1} = x_k - [J^t J + \mu I]^{-1} J^t e \qquad (4)$$

When the scalar μ is zero, this is just Newton's method which is faster and more accurate near an error minimum. So the aim is to shift toward Newton's method as quickly as possible. Performance function is always reduced at each iteration of the algorithm by decreasing μ after each successful step and is increased only when a tentative step would increase the performance function [14][15].

Conjugate Gradient Algorithms

Weights in the basic Back propagation algorithm are adjusted in the steepest descent direction (negative of the gradient). This is the direction in which the performance function is decreasing very fast. Although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In this algorithm search is accomplished along conjugate directions, which produces generally faster convergence than steepest descent directions. In the conjugate gradient algorithms the step size is adjusted at each iteration. To determine the step size search is made along the conjugate gradient direction which will minimize the performance function along that line. The search direction at each iteration is determined by updating the weight vector as:

$$w_{k+1} = w_k + \alpha p_k \qquad (5)$$

Where

$$p_k = -g_k + \beta_k p_{k-1}$$

And

$$\beta_k = \frac{\Delta g_k^t - 1 g k - 1}{g_k^t - 1 g k - 1}$$

$$\Delta g_{k-1} = g_k - g_{k-1}$$

Quasi-Newton Algorithms

Newton's method is used to find the maxima and minima of functions and for the fast optimization as an alternative to the conjugate gradient methods. The basic step of Newton's method is

$$w_{k+1} = w_k - A_k^{-1} g_k \qquad (6)$$

Where A_k is the Hessian matrix (second derivatives) of the performance index at the current values of the weights and biases. Newton's method converges fast but it is more complex to compute the Hessian matrix for feed forward neural networks. Based on Newton's method there is a class of algorithms, which doesn't require calculation of second derivatives. These are called quasi-Newton (or secant) methods. At each iteration these methods update an approximate Hessian that is computed as a function of the gradient. Although this method converges in fewer iterations but it requires more computation in each iteration and more storage than the conjugate gradient methods.

III. LITERATURE REVIEW

Over the years, outcome prediction studies have become the important in many areas of medical research, especially in diabetes. Traditionally, statistical classification procedures deal with these kinds of problems, however satisfactory models for outcome prediction have been challenging to develop. So far different models have been proposed for classification of disease. For classification of diseases one of the major ways is machine learning. Since diabetes shows various signs for the presence of blood glucose and various instances so more training is needed to predict diabetes with proper datasets. Recent study tells that 80% of complications can be prevented by identification of intelligent data analysis method like machine learning technique that are valuable in early detection [16].

Bourd`es *et.al*. [17] Compare multilayer perceptron neural networks (NNs) with standard logistic regression (LR) to identify key covariates impacting on mortality from cancer causes, disease free survival (DFS), and disease recurrence using Area Under Receiver-Operating characteristics (AUROC) in breast cancer patients.

Sa'di et.al. [18] Implemented various data mining techniques such as Naïve Bayes, J48 and Radial Basis Function Artificial Neural Networks for diagnosing diabetes type 2. They use a data set with 768 data samples, 230 of them selected for test phase. They found Naive Bayes algorithm with 76.95% accuracy outperformed J48 and RBF with 76.52% and 74.34% accuracies, respectively.

A. Mehdi *et.al* [19] perform a cross-sectional study with 12000 Iranian people in 2013 using stratified cluster sampling that include information on hypertension and diabetes and their risk factors. They determine the predictive precision of the bivariate Logistic Regression (LR) and Artificial Neutral Network (ANN) in concurrent diagnosis of diabetes and hypertension. They found that ANN is having high accuracy and is the best model for concurrent affliction to hypertension and diabetes than bivariate LR model.

Vijayarani *et al.*, [20] have used classification algorithms to check the presence of liver diseases in patients. Naive Bayes and support vector machine (SVM) are the two classification algorithms used in their paper to check for the occurrence of liver diseases. Accuracy and execution time are the two performance factors used in their paper to measure the effectiveness of algorithms. From the evaluated

results it is found that the SVM is a better classifier when compared to Naive Bayes for predicting the liver diseases.

J.Li *et al*., [21] propose a novel fusion method to jointly represent the tongue, face and sublingual information and discriminate between DM (or IGR) and healthy controls. Joint similar and specific learning (JSSL) approach is proposed to combine features of tongue, face and sublingual vein, which not only exploits the correlation but also extracts individual components among them. Experimental results on a dataset consisting of 192 Healthy, 198 DM and 114 IGR samples (all samples were obtained from Guangdong Provincial Hospital of Traditional Chinese Medicine) substantiate the effectiveness and superiority of our proposed method for the diagnosis of DM and IGR, achieving 86.07% and 76.68% in average accuracy and 0.8842 and 0.8278 in area under the ROC curves, respectively.

The objective of this study is to compare the performance of ANN and logistic regression models for prediction of diabetes based on initial clinical data and whether these models are reproducible. We used different variables even if they were interdependent.

IV. RESULT AND DISCUSSIONS

In our study, we reviewed logistic regression models and ANNs and illustrated an application of these algorithms in predicting the risk of diabetes. The comparison of prediction models in this study showed that accuracy in predicting in diabetes was significantly higher in the ANN model than in the LR model (p 0.001). Performance of selected classification algorithms were experimented on the Pima Indians diabetes dataset. This data set is open sourced gathered from UCI machine learning site. In each and every algorithm we have noticed the parameters sensitivity, specificity, false positive rate, positive predictor value, negative predictor value, false discovery rate, accuracy and error rate.
The parameters illustrated as follows:

Sensitivity: It is used to measure the performance of true positive rate

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity: It is used to measure the performance of true negative rate

$$\text{Specificity} = \frac{TN}{TN+FN}$$

False positive rate: It is the ratio of individual who incorrectly received a positive test result

$$\text{False Positive rate} = \frac{FP}{FP+TN}$$

Positive predictor value: If the test result is positive what is the probability that the patient actually has the disease

$$\text{Positive predictor value} = \frac{TP}{TP+FP}$$

Negative predictor value: If the test result is negative what is the probability that the patient does not have the disease

False discovery rate: It is a way of conceptualizing the *rate* of type I errors in null hypothesis testing when conducting multiple comparisons

$$\text{False discovery rate} = \frac{FP}{FP+TP}$$

Accuracy: It is defined as the ratio of correctly classified instances to total number of instances

$$\text{Accuracy} = \frac{TP+TN}{FP+TP+TN+FN}$$

Error rate: It is the number of bit errors per unit time. It can be calculated with the help of accuracy as shown below

$$\text{Error rate} = 1 - \text{Accuracy}$$

Precession: Precession is used to access the percentage of tuples labelled as diabetes that are actually diabetes tuples.

$$\text{Precession} = \frac{t-pos}{t-pos+f-pos}$$

Sensitivity: The sensitivity and specificity measures can be used respectively for the alternative to accuracy measure. Sensitivity is also referred to as the true positive (recognition) rate (that is, the proportion of positive tuples that are correctly identified).

$$\text{Sensitivity} = \frac{t-pos}{pos}$$

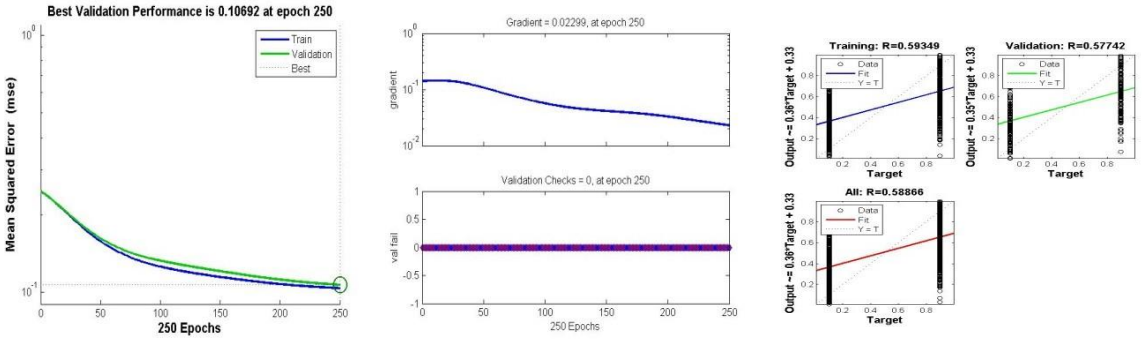
Specificity: Specificity is the true negative rate (that is, the proportion of negative tuples that are correctly identified).

$$\text{Specificity} = \frac{t-neg}{neg}$$

where t-pos is the no. of true positives (diabetes tuples that were correctly classified as such), pos is the no. of positive ("diabetes") tuples , t-neg is the no. of true negatives("not diabetes" tuples that were correctly classified as such), neg is the no. of negative ("not diabetes") tuples and f-pos is the no. of false positives ("not diabetes" tuples that were incorrectly labelled as "diabetes"). It can be shown that accuracy is a function of sensitivity and specificity. In this paper, the results are analysed with various classification algorithms on the given medical dataset. This work is implemented in Matlab. The performance measures of classification algorithms on Pima Indians diabetes dataset is show in table and figures below. An evaluated result shows the performance of ANN is better than Logistic regression

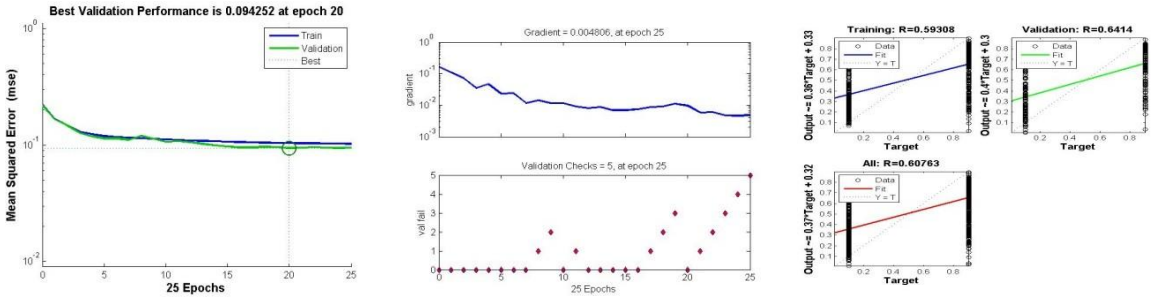
TABLE III PERFORMANCE MEASURES OF PIMA INDIANS DIABETES DATASET

Algorithm	accuracy	Sensitivity	specificity	False Positive rate	Negative Predictor Value	Positive Predictor Value	False Discovery rate	Error rate
Logistic Regression	0.768	0.602	0.863	0.13	0.7	0.79	0.28	0.231
Standard Back propagation	0.7860	0.6216	0.8645	0.1355	0.8272	0.6866	0.3134	0.2140
Resilient - Back propagation	0.8035	0.6627	0.8836	0.1164	0.8217	0.7639	0.2361	0.1965
Variable Learning Rate	0.7773	0.5745	0.9185	0.0815	0.7561	0.8308	0.1692	0.2227
Powell-Beale Conjugate Gradient	0.7904	0.5823	0.9000	0.1000	0.8036	0.7541	0.2459	0.2096
Levenberg-Marquardt	0.8035	0.6517	0.9000	0.1000	0.8025	0.8056	0.1944	0.1965
Quasi-Newton Algorithm	0.8035	0.6707	0.8776	0.1224	0.8269	0.7534	0.2466	0.1965
Scaled Conjugate Gradient	0.7948	0.5200	0.9286	0.0714	0.7989	0.7800	0.2200	0.2052



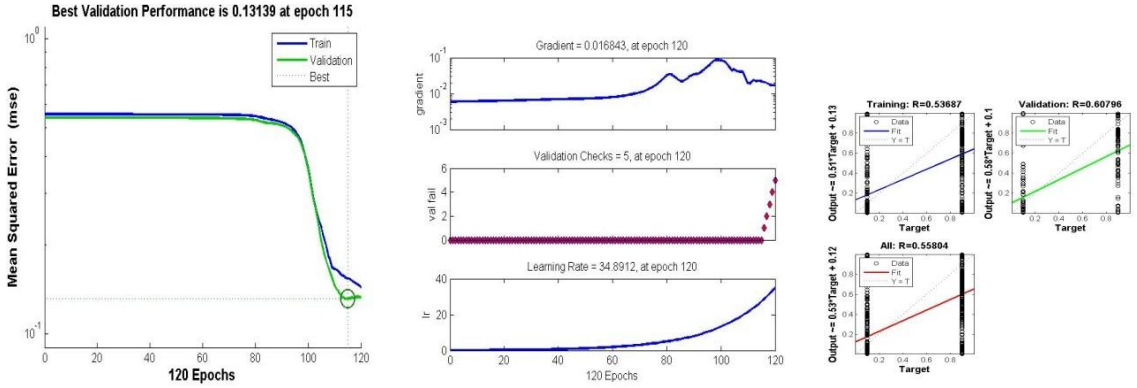
NN performance plot **NN Training State plot** **NN regression plot**

Fig.2 Performance of Standard back propagation algorithm



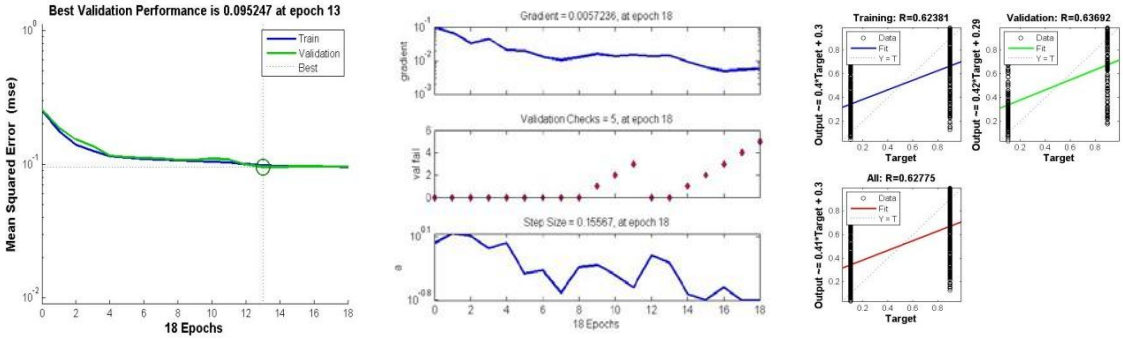
NN performance plot **NN Training State plot** **NN regression plot**

Fig.3 Performance of Resilient - Back propagation algorithm



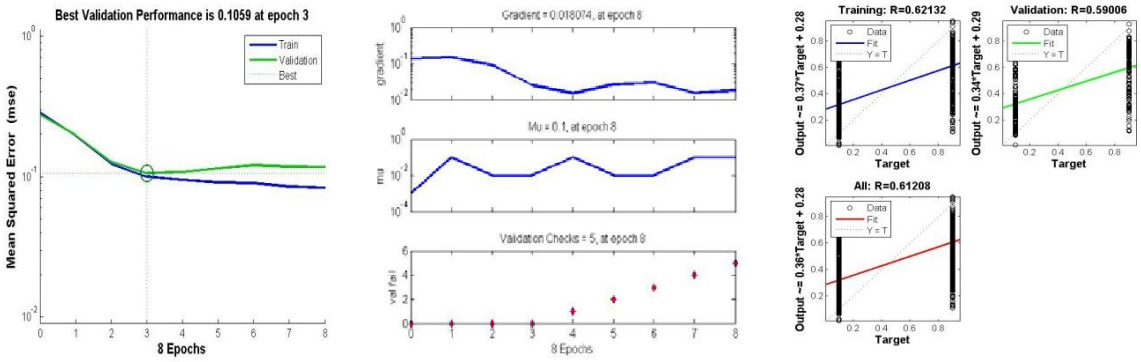
NN performance plot **NN Training State plot** **NN regression plot**

Fig.4 Performance of Variable Learning Rate back propagation algorithm



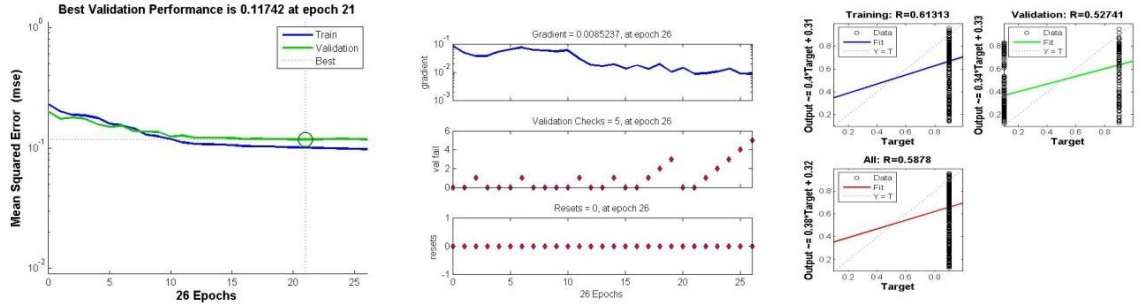
NN performance plot **NN Training State plot** **NN regression plot**

Fig5. Performance of Powell-Beale Conjugate Gradient back propagation algorithm



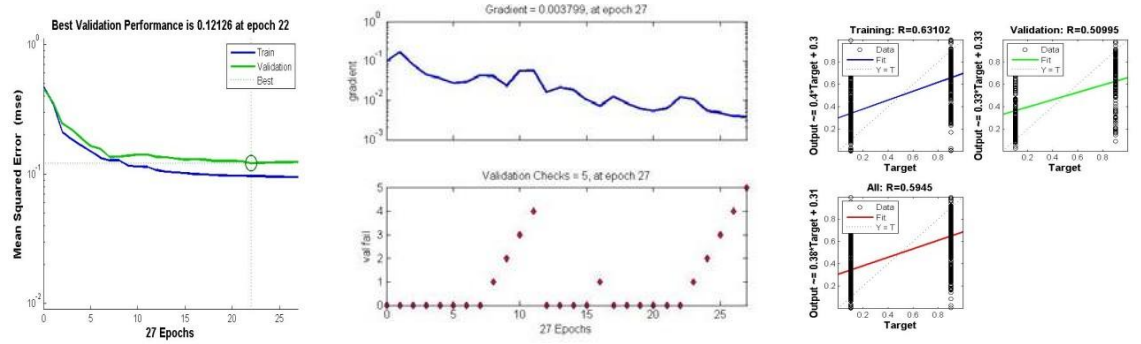
NN performance plot **NN Training State plot** **NN regression plot**

Fig.6 Performance of Levenberg-Marquardt back propagation algorithm



NN performance plot **NN Training State plot** **NN regression plot**

Fig.7 Performance of Quasi-Newton Algorithm back propagation algorithm



NN performance plot **NN Training State plot** **NN regression plot**

Fig.8 Performance of Scaled Conjugate Gradient back propagation algorithm

V. CONCLUSION

In this paper we have used Matlab tool for analysis and performed comparison of selected classification algorithms. After the comparative analysis we examined that neural network algorithms is more accurate and has less error rate. Our interface also provides the user the choice of selecting suitable prediction algorithm. We conclude that ANN has more precision than other models.

REFERENCES

[1] R. Deja, W. Froelich, G. Deja, and A. Wakulicz-Deja, "Hybrid approach to the generation of medical guidelines for insulin therapy for children," *Inf. Sci. (Ny).*, vol. 384, pp. 157–173, 2017.

[2] Y. Li, C. Bai, and C. K. Reddy, "A distributed ensemble approach for mining healthcare data under privacy constraints," *Inf. Sci. (Ny).*, vol. 330, pp. 245–259, 2016.

[3] J. A. Morente-Molinera, I. J. Pérez, M. R. Ureña, and E. Herrera-Viedma, "Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies," *Inf. Sci. (Ny).*, vol. 328, pp. 418–434, 2016.

[4] World Health Organization, "Prevention of blindness from diabetes mellitus," *Geneva WHO*, pp. 1–48, 2005.

[5] http://ftp.ics.uci.edu/pub/ml-repos/machine-learning databases/pima-indians-diabetes, 2003.

[6] Breiman L et al. Classification and regression trees. Belmont, CA: Wadsworth; 1984.

[7] S. Salzberg, "Book Review: C4.5: Programs for Machine Learning," *Mach. Learn.*, vol. 1, no. 16, pp. 235–240, 1994.

[8] Caruth, C.: "History as false witness". In: Ekman, U., Tygstrup, F. (eds.) Witness Memory, Representation and the Media in Question. Museum Tusculanum Press, University of Copenhagen (2008)

[9] D. W. Hosmer and S. Lemeshow, "Applied Logistic Regression," *Wiley Series in Probability and Sattistics*, no. 1. p. 373, 2000.

[10] M. T. Hagan, H. B. Demuth, and M. H. Beale, "Neural Network Design," *Bost. Massachusetts PWS*, vol. 2, p. 734, 1995.

[11] Guerriere MR, Detsky AS. "Neural networks: what are they? ", Ann Intern Med 1991;115:906–907.

[12] Haykin S. "Neural networks: a comprehensive foundation", Upper Saddle River, NJ: Prentice Hall, 1998.

[13] More ,J .J, in Numerical Analysis, edited by Watson, G.A, Lecture Notes in Mathematics 630, (Springer Verlag, Germany) 1997,105-116.

[14] Hagan M T & Menhaj. IEEE Trans Neural Networks 5(6) (1994) 989-993.

[15] S. A. Dudani. The Distance-Weighted k-NN Rule. IEEE Transactions on Systems, Man and Cybernetics, Volume 6, Number 4, pp. 325-327, 1976.

[16] Nahla H.Barakat, Andrew P. Bradley and Mohamed Nabil H. Barakat, "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus"," IEEE Transactions on Information Technology in biomedicine", vol. 14, no. 4, July 2010.

[17] Val´erie Bourd`es, St´ephane Bonnevay, Paolo Lisboa, R´emy Defrance, David P´erol, Sylvie Chabaud, Thomas Bachelot, Th´er`ese Gargi, and Sylvie N´egrier. "Comparison of Artificial Neural Network with Logistic Regression as Classification Models for Variable Selection for Prediction of Breast Cancer Patient Outcomes", "Advances in Artificial Neural Systems", Volume 2010, Article ID 309841, 10 pages

[18] S. Sa'di, A. Maleki, R. Hashemi, Z. Panbechi, K. Chalabi. Comparison of Data Mining Algorithms in the Diagnosis of Type Ii Diabetes. International Journal on Computational Science & Applications (IJCSA), Vol.5, No.5,(2015), pp. 1-12.

[19] M. Adavi, M. Salehi, and M. Roudbari, "Artificial neural networks versus bivariate logistic regression in prediction diagnosis of patients with hypertension and diabetes," pp. 2–6, 2016.

[20] S. Vijayarani, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms," vol. 4, no. 4, pp. 816–820, 2015.

[21] J. Li, D. Zhang, Y. Li, J. Wu, and B. Zhang, "Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection," *Inf. Sci. (Ny).*, vol. 384, pp. 191–204, 2017.