An Intelligent Expert Based System Neural Network For The Diagnosis Of Type2 Diabetes Patients.

V. Sujatha
School of Advanced Sciences
VIT University, India
sujathavishnu03@gmail.com

Abstract- This paper presents an intelligent expert based system neural network for the diagnosis of Type 2 diabetes patients, given the variable, NPG- Number of times pregnant, PGL-Plasma Glucose concentration a 2 hours in an oral glucose tolerance test, DIA-Diastolic blood pressure(mmHg),TSF-Triceps skin fold thickness (mm), INS-2-Hour serum insulin (mu U/ml), BMI-Body mass index (weight in kg/(height in m)^2), DPF-Diabetes pedigree function, AGE-Age (years) , the variable diabet was classified by ANN as either present or not present, Diabet- Class variable (0 or 1). In this study, general regression and neural network (GRNN) tools are applied to data sets they were investigated, the results are presented and discussed.

Key Words- GRNN, Regression, data mining, blood sugar, diabetics

I INTRODUCTION

Nowhere is the diabetes epidemic more pronounced than in India as the World Health Organization (WHO) reports show that 32 million people had diabetes in the year 2002. The International Diabetes Federation (IDF) estimates the total number of diabetic subjects to be around 40.9million in India and this is further set to rise to 69.9million by the year 2025. Diabetes, often referred to by doctors as diabetes mellitus, describes a group of metabolic diseases in which the person has high blood glucose (blood sugar), either because insulin production is inadequate, or because the body's cells do not respond properly to insulin, or both. In this study, general regression and neural network (GRNN) tools are applied to diagnose the Pima Indian diabetes data sets were investigated the results are presented and discussed. Research efforts on Neural Networks for predicting are considerable. The literature is vast and growing. Regression is a powerful Statistical technique to estimate the relationship between a dependent variable and two or more variables.

Regression is a statistical technique to estimate the relationship between a dependent variable and two or more independent variable. The general regression model does not require an iterative training procedure. Regression coefficients are estimated by minimizing the sum of residuals. The standard error of the regression is based on the sum of the residuals .Regression is the attempt to explain the variation in a dependent variable using the variation in independent variable.

Neural Network is an tool to predict the diabetes of a patient. Data mining also popularly known as knowledge discovery to find the inter relation pattern among the data. It provides an useful information from large set of data bases. Neural networks are used for prediction with various levels of success. The advantage includes the automatic learning of dependencies only from measured data without any need to add further information. The predictive accuracy on a neural network is more than the Regression technique of human experts. Neural networks can learn the dependency valid for a certain period. The knowledge stored in the form of neural network are strongly non-linear dependent and even there is significant noise in the training set. The knowledge is not a comprehensible.

II RELATED WORK

Smith et al. [2] used the PID data set to evaluate the perceptron- like Adaptive learning routine (ADAP).This study had 576 cases in the training set and 192 cases in the test set. Using 576 training instances, the sensitivity and specificity of their algorithm was 76% on the remaining 192 instances. Upadhyaya S, Farahmand K., Baker- Demaray F.[3]they Compare the Neural Network and Linear Regression. They classify in the context of screening Native American elders with diabetes. Kamer Kayaer, Tulay, Yildirim [1] 'Medical Diagnosis on Pima Indian Diabetes using general Regression Neural Networks'. They compare the linear General Regression Neural Network (GRNN) model with two different neural network structures, which are multilayer perceptron (MLP), radial basis function (RBF). They applied three tools to the Pima Indians Diabetes (PID), medical data. The best result achieved on the test data is the one using the GRNN structure (80.21%). This is very close to one with the highest true classification result that was achieved by using the more complex structured ARTMAP-IC network (81%) [1]. Therefore they show that, general regression neural network

(GRNN) can be a good and practical choice to classify a medical data. Similarly, Jankowskiand Kadirkamanathan developed a radial basis function network suite called Inc Net which used 100 neurons and trained for 5,000 iterations.

This approach yielded 77.6% accuracy. Au and Chan in attempted to improve the correct classification percentage on the PID dataset by using a fuzzy approach. Au and Chan first represented the revealed regularities and exceptions using linguistic terms, and then mined interesting rules for the classification based on membership degrees. Their approach yielded 77.6accuracy. Rutkowski and Cpalka in introduced a new neural-fuzzy structure. called a flexible neural fuzzy inference system (FLEXNFIS). Based on the input and out put data, they proposed the parameters of the membership functions and the type of the neuron systems (Mamdani or logical). However, their correct classification percentage on the PID dataset was 78.6%. Davis in [10] developed a fuzzy neural network by using the BK-Square products. This fuzzy neural network was then tested on the PID dataset. The result of his approach yielded 81.8the results obtained from the Stat Log project [11] when evaluating for many different classification algorithms on the PID dataset showed that their correct classification percentage was less than 78%.

The data set were downloaded from the UCI Repository of machine learning databases. All patients represented in this data set are females with at least 21 years old of Pima Indian heritage. The problem posed here is to predict whether a person would test positive given a number of physiological measurements and medical test results. Diabetes were diagnosed according to the World Health Organization criteria [5].

Table 1

Characteristics of data sets. Data set	No.of examples	Input attributes	Output classes	Total No. of attributes	Missing attributes status	Noisy attributes status
Pima	768	8	2	9	No	No

There are eight variables which are all numerically valued they are

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 (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)

The eight variables on diabetic set are considering as a risk factors among pima Indian diabetic patients. Out of 768 data sets 500 instances are non-diabetic and 268 are diabetic.
Pima Indian Diabetes Data (PIDD) set is downloaded from the machine learning database at UCI. The problem posed here is to predict whether a person would test positive given a number of physiological measurements and medical test results. This is a two-class problem with class value 1 being in-terpreted as “tested positive for diabetes”

III RESULTS AND DISCUSSION

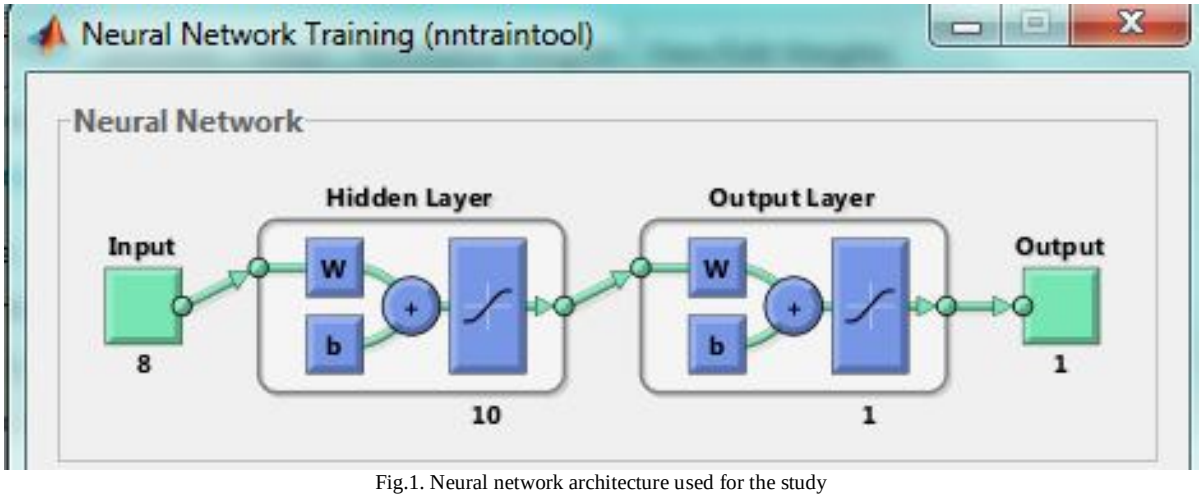

Fig.1. Neural network architecture used for the study

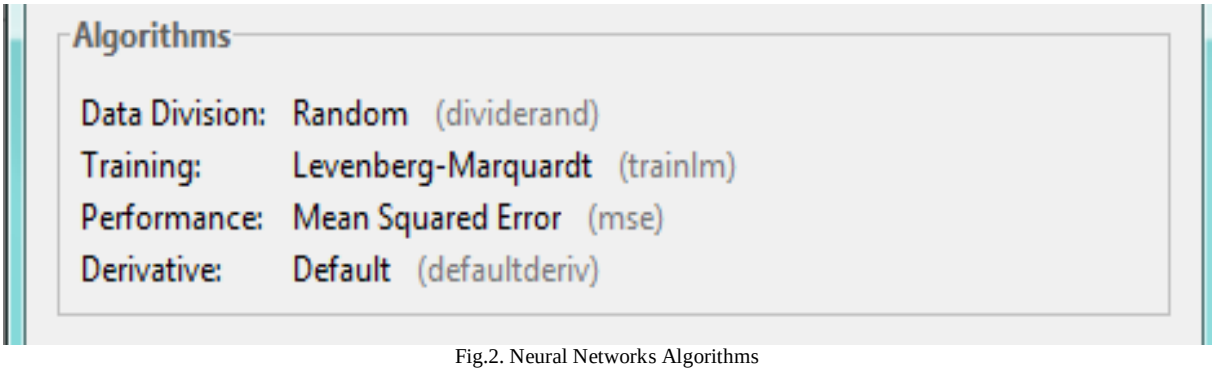

Fig.2. Neural Networks Algorithms

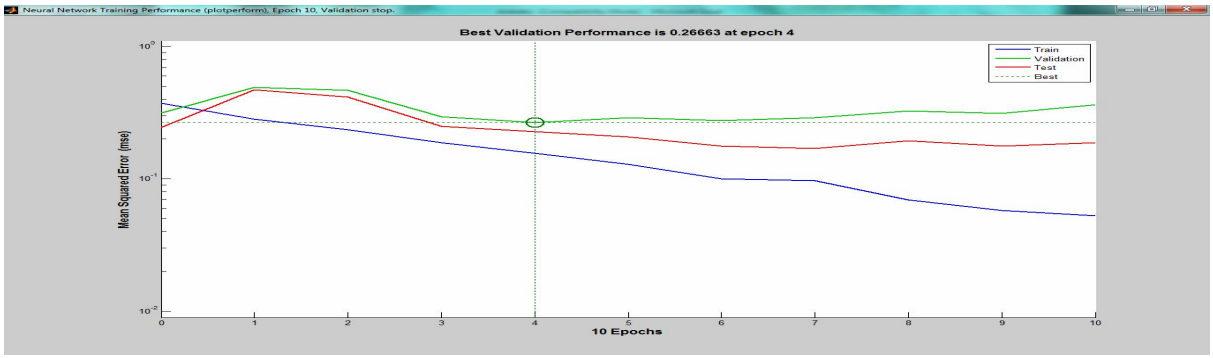

Fig.3. Result NN prediction shows only 0.266 as MSE

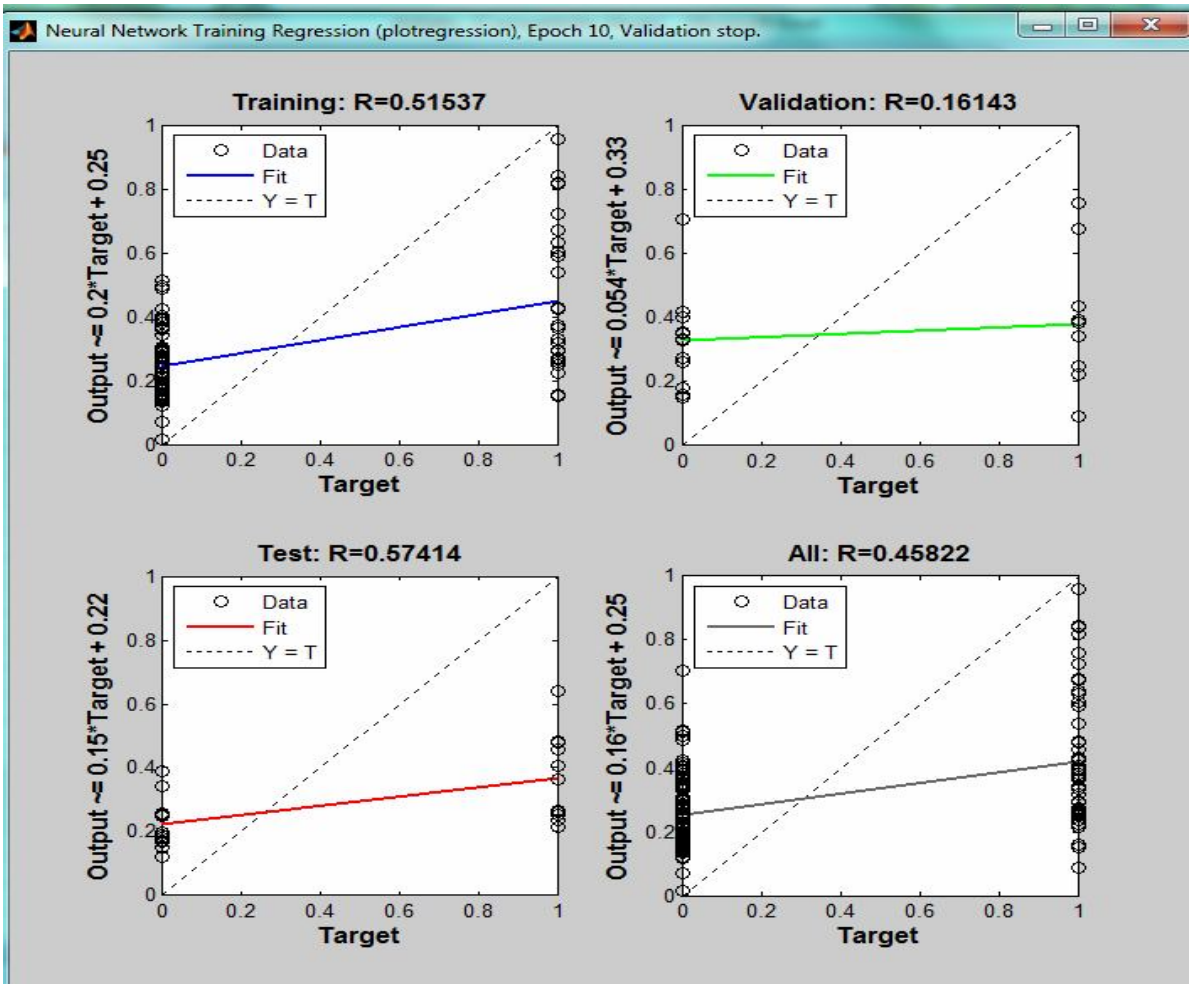

Fig.4. Results of regression testing

III CONCLUSIONS

From the regression results we can infer that the prediction accuracy is reported as 57%. From the ANN results, it is inferred that the prediction accuracy is 74%. Hence, ANN prediction is better than regression prediction. Given the input dataset, the class variable, diabetes dataset.

REFERENCES

[1] Kamer Kayer and Tulay yildirim,,*Medical diagnosis on PIMA Indian Diabetes using General Regression Neural Networks,*2003.
[2] Smith JN,Ever Hart JE, Dickson WC,et al, *Using the ADAP learning algorithm to forecast the onset of diabetes mellitus,* Proceedings of the symposium on computer Applications and medical care, IEEE Computer Society Press,1988,261-265.
[3] Jin Park,Dec W.Edington, *Asequential neural network model for diabetes prediction.*ACM Digital Library Volume 23 Issue 3,Nov-2001.
[4] Yuk Sal Oner, Taner Tunc,Erol Egrioglu,Yildiz Atasoy,*Comparisons of Logistic Regression and Artificial Neural Networks in Lung Cancer Data, American Journal of Intelligence systems.*p-Issn 2165-2978.-2013.
[5] WHO- World Health Organization criteria for diabetes
[6] UCI –Machine Learning Data Repository for diabetes data set.
[7] Rutkowski.L, Cpalka.K, *Discussion of a Hybrid interior point Assisted Differential evaluation Algorithms for Economic Dispends,* IEEE Transactions.Vol.14.Issue 4-2003.
[8] Huy Nguyen Anh Pham and Evangelos Triantaphyllou, *Prediction of Diabetes by Employing a New Data Mining Approach, Which Balances Fitting and Generalization* ,Computational Intelligence Volume 131, pp 11-26, 2008.
[9] Zahra Farzanyar, Mohammadreza Kangavari, *Efficient Mining Of Fuzzy Association Rules from the Pre-Processed Dataset,* Computing and Informatics, Vol. 31, 2012, 331–347.
[10] Warren L. Davis, IV and Ladislav J. Kohout, *Enhancing Performance of Relational Fuzzy Neural Networks with Square BK- Products,* DigiNole Commons Electronic Theses, Treatises and Dissertations 2006.
[11] Stat Log Project On Diabetes, Jeroen Eggermont and Joost N. Kok and Walter A. Kosters. *Genetic Programming for data classification: partitioning the search space.* SAC. 2004.