ISSN:0975-9646

Prediction of Type 2 Diabetes using Optimized ANFIS with Genetic Algorithm and Particle Swarm Optimization

Alby S[1], Dr. B L Shivakumar[2]

[1]*Research Scholar, Research and Development Centre*
Bharathiar University, Coimbatore – 44

[2]*Principal, Sri Ramakrishna Polytechnic College*
Coimbatore – 22

***Abstract*-This paper proposes two different approaches for the prediction of type2 diabetes. Many different techniques have been used for the prediction of chronic diseases by different researchers. Among them Adaptive Neuro Fuzzy Inference system (ANFIS) is very popular and already used for the prediction of type 2 diabetes. In this paper, the proposed system is optimization of ANFIS using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) which reduces the complexity of ANFIS and increases the accuracy of prediction.**

I. INTRODUCTION

In today's world health and fitness remains a crucial subject as the number of health issues continues to increase worldwide. Usually many chronic diseases emerge in middle aged people because of unhealthy lifestyle which includes lack of physical activity, tobacco use, consumption of fast food etc. This lifestyle results in high risk factors which lead to chronic disease like diabetes, hypertension, cancer etc. Here arises the importance of prediction of diseases which can be prevented before they occur. Research is progressing in this field by many researchers. Since large amount of data has to be analyzed, data mining techniques are the most suitable method for this purpose.

A. Data Mining

Data Mining (DM) can be viewed as a result of the natural evolution of information technology [1]. In recent years the information industry deals with a huge amount of data and needs turning such data into useful information and knowledge. The information and knowledge gained can be used for a wide range of applications. Data mining refers to extracting knowledge from large amounts of data. It is also treated as knowledge discovery in database (KDD). Knowledge discovery as a process is depicted in Fig. 1 and consists of an iterative sequence of the following steps:

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures)
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)

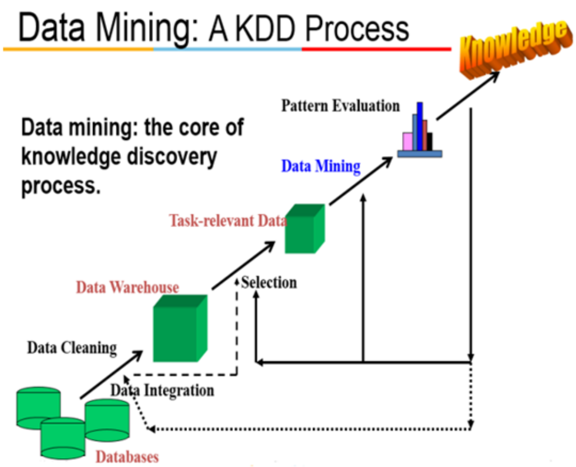

Fig.1: KDD process

There are many different technologies available in data mining including regression, neural networks, and decision trees. Neural networks are widely used by data mining practitioners in the field of disease detection and prediction. This paper presents two different approaches of Adaptive Neuro Fuzzy Interface System (ANFIS) for the prediction of diabetes.

B. Diabetes

Diabetes mellitus (or diabetes) is a chronic, lifelong condition that affects the body's ability to use the energy found in food. There are three major types of diabetes: type 1 diabetes, type 2 diabetes, and gestational diabetes [2].
According to the different causes of occurrence, diabetes are classified mainly into 3 types: type1, type 2 and

gestational diabetes. Type 1 diabetes is insulin –dependent and type 2 diabetes is non insulin – dependent. Gestational diabetes occurs temporarily during pregnancy. Diabetes is serious condition and life threatening. Proper exercise and diet control reduces the severity of diabetes. About 90% of people with diabetes have type 2 diabetes [3].So if it can be predicted the chance of occurrence can be reduced.

Many researches have been done in this field. Different data mining techniques have been used for the prediction of type 2 diabetes by researchers. This paper presents two approaches for the prediction of type 2 diebetes using ANFIS optimized by Genetical Algorithm(GA) and ANFIS optimized by Particle Swarm Optimization (PSO).

II. Literature Review

Authors in [6] have constructed association rules for classification of type -2 diabetic patients. They generated 10 association rules to identify whether the patient goes on to develop diabetes or not. Several of machine learning algorithms have been proposed in the context and have been successfully used in some parts. In their first stage of work the missing values were handled and applied equal interval binning with approximate values and lastly Pima Indian dataset were applied by the Apriori algorithm to generate the rules. In this study they have included only pregnant women below 21 years who are type-2 diabetic. Many other factors which influence diabetes may be considered for improving the generalization of rule.

In paper [7], two classified techniques with principle component analysis (PCA) are implemented for the forecasting of diabetes and concluded with the best forecasting. The techniques are Neural Network and ANFIS, the dataset which they have utilized is the same Pima Indian dataset with 8 features. Using NN 72.9% and ANFIS 70.56 % with PCA 89.2% and 90.4% of accuracy respectively were being obtained. They have also designed one GUI in MATLAB using GUIDE to represent the work in simplest manner, so that any doctor who is not familiar with MATLAB can use this and it is inferred that the classification ability of all the classifiers is better for non-diabetic samples than that for diabetic ones.

M.W. Aslam et. al in [8] uses genetic Programming (GP) and a variation of genetic programming called GP with comparative partner selection (CPS) for diabetes detection. The system produces a single individual as output from training data, that converts the available features to a single feature such that it has different values for healthy and patient(diabetes) data and in the next stage use test data for testing of that individual. The proposed system was able to achieve only 78.5% accuracy.

In 2003 K. Kayaer et. al applied three different neural network structures namely MLP, RBF and GRNN to PID medical data [9]. They came up with an accuracy of 80.21% to classify a medical data. Even though in our proposed work we are also using GRNN for prediction of diabetes comparing to this work we have used a different approach for the prediction. In [10] Xiao Fang tried to make use of different DM concepts to gain knowledge about diabetes. Using several challenging applications of DM, an application for identifying diabetic patients in a small US town was presented. Even though different methods produced different accuracy levels, on an average 75% of accuracy was obtained. The limitations of this work are the population is not clustered into different risk related populations and the different variables that found critical for identifying diabetic patients are not subcategorized.The authors in [11] propose a system which can improve the strategy to a better level where artificial metaplasticity on perceptrons is implemented on neural network. The nodes of neural network are parameters containing features like thirst increase, hungry increase, nausea, fatigue, vomiting etc. This proposed system can increase the efficiency of the system which is in existence.Madhavi Pradhan et. al in [12] uses a neural network implementation of the fuzzy k-nearest neighbor algorithm for designing of classifier for detection of diabetes. This method of detection of diabetes proposes a system that will be implemented in client – server architecture. The training dataset will be kept on the server, which will be used to train the neural network classifier on the mobile device. The mobile device is a feature add-on for convenience of the doctor. The accuracy of this method is calculated as 72.82%.

Kalliopi.V.Dalakleidi et. al present a hybrid approach based on the combined use of a genetic algorithm (GA) and a nearest neighbours classifier for the selection of the critical clinical features which are strongly related with the incidence of fatal and non-fatal Cardiovascular Disease (CVD)in patients with Type 2 Diabetes Mellitus (T2DM)[13].In order to overcome the problem of unbalanced data the dual weighted k-nearest neighbor classifier was used.

The aim of this study [14] was to examine whether waist circumference (WC) or WHR improve diabetes prediction beyond body mass index in older men and women, and to define optimal cut-off points. Both overall and central adiposity indices are strong predictors of type 2 diabetes in older adults. BMI is as strong as WC in predicting type 2 diabetes in men. In women, however, WC was a significantly better measure for the identification of diabetes risk.

Studies aiming at preventing or delaying diabetes are critically dependent on the ability to accurately predict Type1 and Type 2 diabetes [15]. After a median follow up time of 8 years, Islet cell auto antibodies did not predict diabetes. BMI measured at base line was as effective as 2-h plasma glucose and fasting plasma glucose to predict diabetes in the adult population.

In [16] the objective of Meredith F Mackay et. al was to compare different anthropometric measures in terms of their ability to predict type 2 diabetes and to determine whether predictive ability was modified by ethnicity. Waist – height ratio was the most predictive measure, followed by BMI. Measures of central and overall adiposity predicted type 2 diabetes to a similar degree.

C. Kalaiselvi and Dr. G .M. Nasira in [17] uses the same method ANFIS for diagnosis of diabetes and prediction of cancer.C.P.Ronald Reagan and Dr.S.Prasanna Devi presented in [18] an android app for intelligent dosage planning in type2 diabetes using ANFISGA.

III. METHODOLOGIES

A. Dataset

In this research work, data of females aged at least 21 years from Pima Indian Dataset were analysed.768 instances have been considered. The following attributes have been considered:

1. *Number of times pregnant*
2. *Plasma glucose concentration a 2 hours in an oral glucose tolerance test*
3. *Diastolic blood pressure*
4. *Triceps skin fold thickness*
5. *2-hour serum insulin*
6. *Body mass index*
7. *Diabetes pedigree function*
8. *Age*

B. Adaptive network based fuzzy inference system (ANFIS)

Concept and Structure: ANFIS [4] combines the advantages of two intelligent approaches neural network and fuzzy logic to allow good reasoning in quantity and quality. A network obtained has an excellent ability of training by means of neural networks and linguistic interpretation of variables via fuzzy logic. The both of them encode the information in parallel and distribute architecture in a numerical framework.

Rule: if x is A1 and y is B1 then f (x) = px + qy + r

Where x and y are the inputs, A and B are the fuzzy sets, f are the output, p, q and r are the design parameters that determined during the training process. ANFIS is composed of two parts is the first part is the antecedent and the second part is the conclusion, which are connected to each other by rules in network form. Five layers are used to construct this network. Each layer contains several node sits structure shows in figure 2.

layer1: executes a fuzzification process which denotes membership functions (MFs) to each input. In this paper we choose Gaussian functions as membership functions:

$$O_i^1 = \mu_{Ai} = \exp\left(\frac{-(x-c)^2}{\sigma^2}\right)$$

layer2: executes the fuzzy AND of antecedents part of the fuzzy rules

$$O_i^2 = w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2), i = 1,2,3,4$$

layer3: normalizes the MFs

$$O_i^3 = \overline{w_i} = \frac{w_i}{\sum_{j=1}^{4} w_j}, i = 1,2,3,4$$

layer4: executes the conclusion part of fuzzy rules

$$O_i^4 = \overline{w_i} y_i = \overline{w_i}\left(\alpha_1^i x_i + \alpha_2^i x_2 + \alpha_3^i\right), i = 1,2,3,4.$$

layer5: computes the output of fuzzy system by summing up the outputs of the fourth layer which is the defuzzification process.

$$O_i^5 = overll\_output = \sum_{i=1}^{4} \overline{w_i} y_i = \frac{\sum_{i=1} w_i y_i}{\sum_{i=1}^{4} w_i}$$

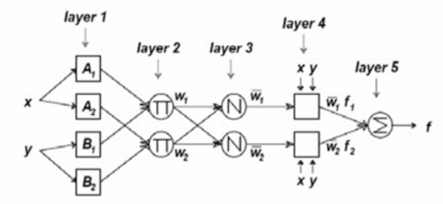

Fig.2: ANFIS architecture

Circles in ANFIS represent fixed nodes that predefined operators to their inputs and no other parameters but the input participate in their calculations. While square are the representative for adaptive nodes that affected by internal parameters.

C. Genetic Algorithm

Genetic Algorithm was developed by J. Holland in 1975[5]. Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Although randomized, GAs are by no means random, instead they exploit historical information to direct the search into the region of better performance within the search space. The basic techniques of the GAs are designed to simulate processes in natural systems necessary for evolution.

D. Adaptive Neuro- Fuzzy Inference System Genetic Algorithm (ANFISGA).

This method is based on Neuro-fuzzy inference system and genetic algorithm.

In ANFISGA there are eight inputs and one output.

In ANFISGA we have five layers.

Layer 1 is the input layer. Neurons in this layer simply pass external crisp signal to Layer 2. Layer 2 is the fuzzification layer. Neurons in this layer perform fuzzification.

Layer 3 is the rule layer. Each neuron in this layer corresponds to signal Sugeno-type fuzzy rule.

Layer 4 is the normalization layer. Each neuron in this layer receives inputs from all neurons in the rule layer, and calculates the normalized firing strength of given rule.

Layer 5 is the de-fuzification layer. Each neuron in this layer is connected to the respective normalization neuron.

The figure 1 shows the mean square error in training the data using ANFISGA. After train, the data is tested with the same algorithm. In testing the mean square error rate is reduced from 1136.9437 to 1049.6364. The same data set is then trained and tested with ANFISPSO. The rate of mean square error shows that using PSO is better than using GA.While using ANFISPSO the error is minimum.

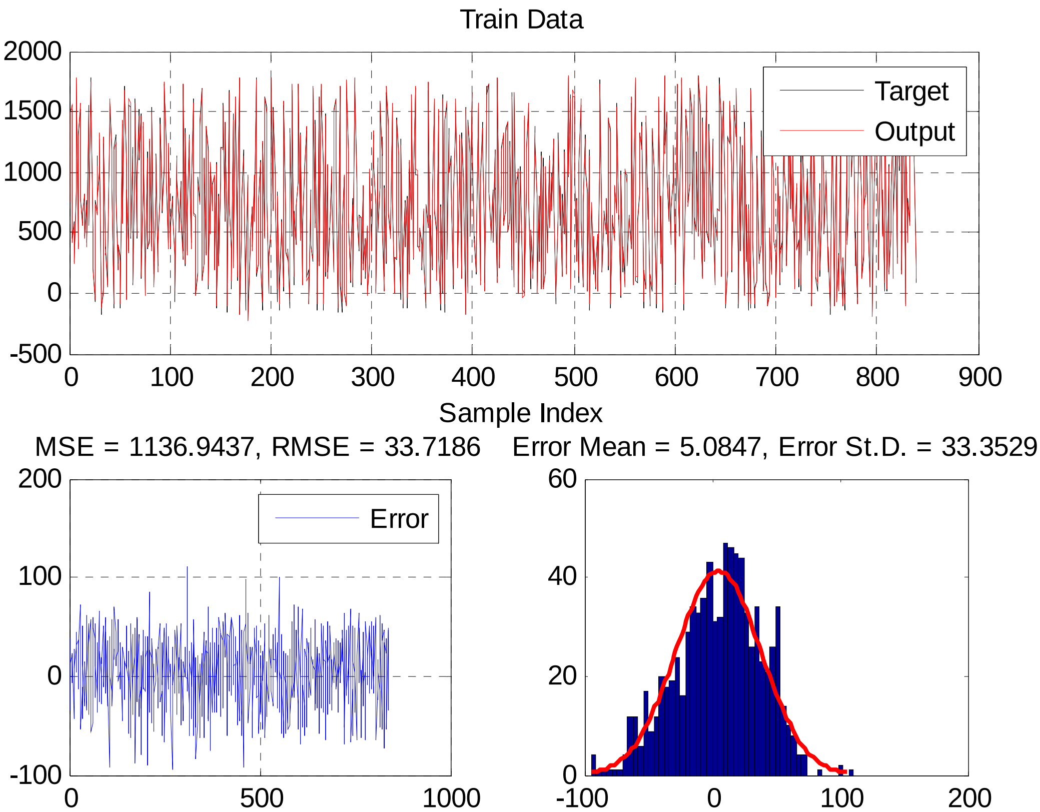

Fig.1: ANFISGA

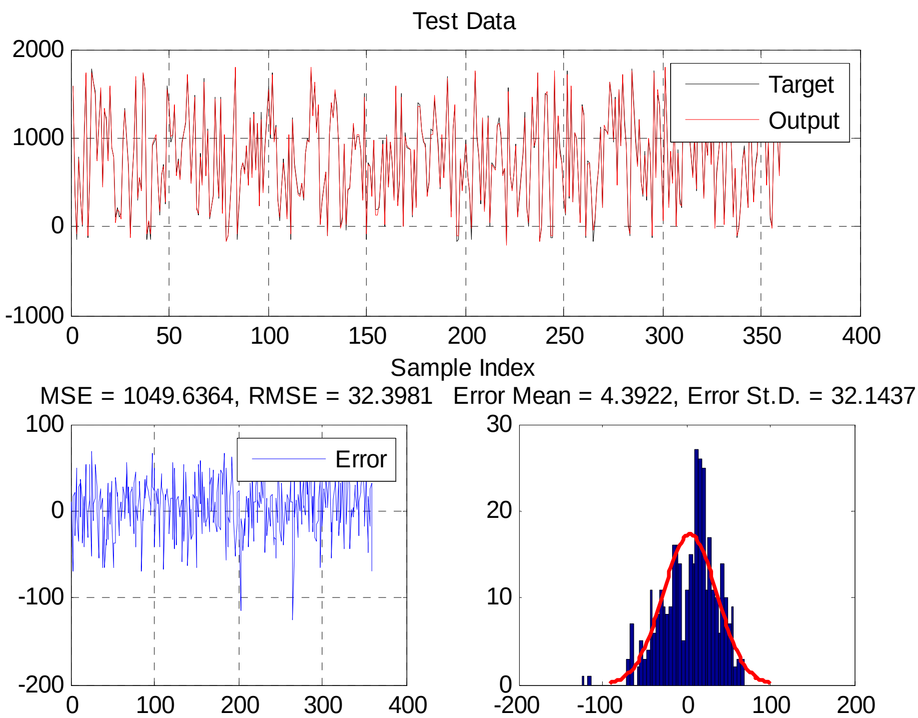

Fig.2: ANFISGA

E. Particle swarm optimization (PSO)

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995[19]. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. The detailed information will be given in following sections.

Compared to GA, the advantages of PSO are that PSO is easy to implement and there are few parameters to adjust. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other areas where GA can be applied.

F. Adaptive Neuro- Fuzzy Inference System Particle Swarm Optimization (ANFISPSO)

The PSO is used to improve the performance of ANFIS, tuning the membership functions required to achieve a lower error[20].

Initialize PSO with random Parameters
While(not PSO_stopping_criteria_met) foreach PSO_particle
train FIS via ANFIS with starting parameters from
PSO_particle
get FIS score (this is your particle score)
save the best FIS, along with its score, etc.
modify PSO particles (remember to use the original PSO particles)
output the best FIS mode

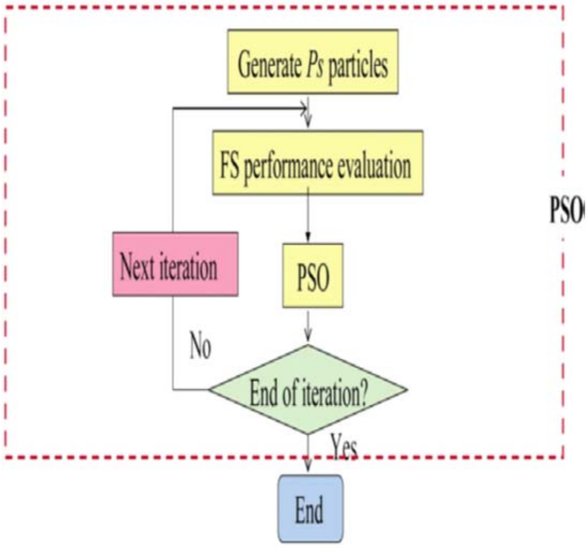

Fig.3: Flow chart of sequential combination of ANFIS and PSO [21].

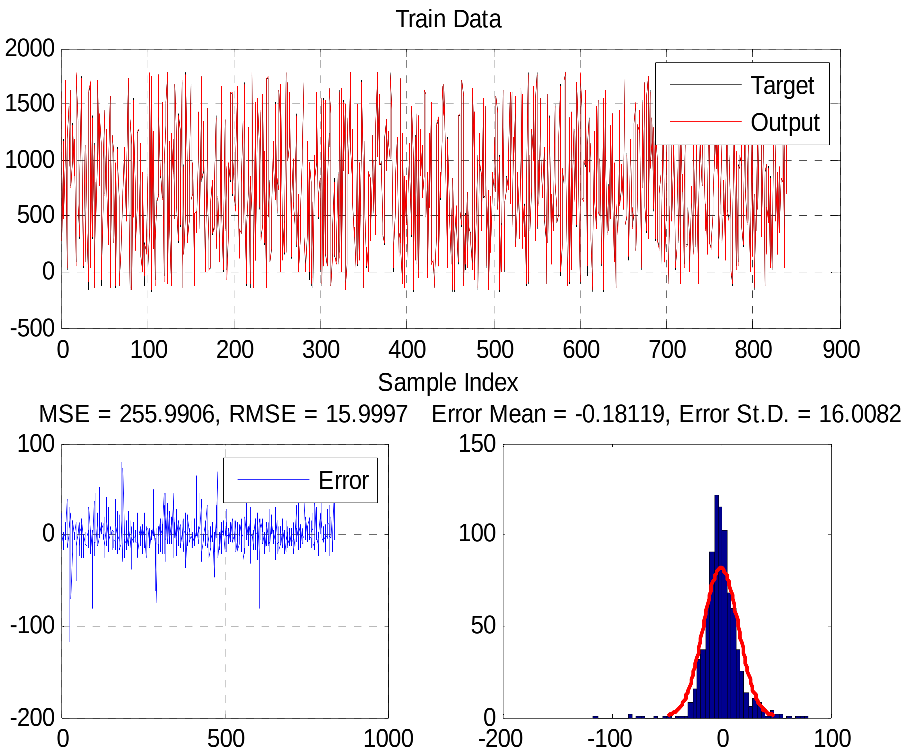

Fig 4: ANFISPSO

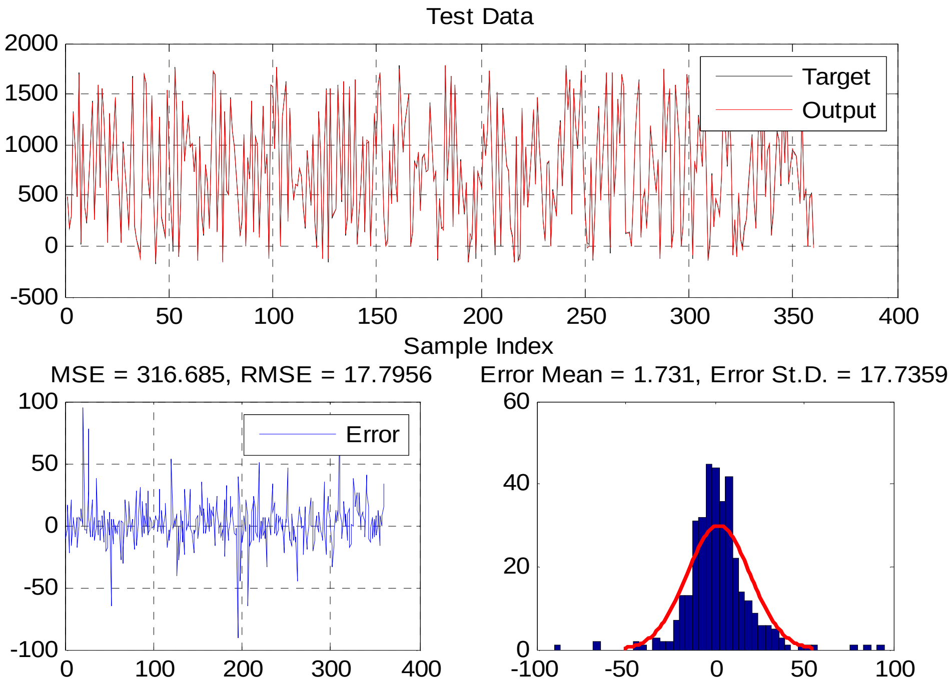

Fig.5 :ANFISPSO

IV. CONCLUSION

This paper presented two approaches for the presiction of type 2 diabetes using ANFIS optimized by Genetic Algorithm and ANFIS optimized by Particle Swarn Optimization algorithm . The result shows the better result is with ANFISPSO. For further work a new optimization algorithm Mine Blast Algorithm(MBA) can be used as an optimizer .This optimizer will efficiently reduce the complexity of ANFIS architecture.

REFERENCES

[1] Jiawei Han and Micheline Kamber, "Data Mining : Concepts and Techniques", 2nd edition.

[2] http://www.webmd.com/diabetes/guide /types-of-diabetes-mellitus.

[3] http://www.diabetes.ca/about-diabetes/ types-of-diabetes.

[4] J.S.R. Jang, "ANFIS: Adaptive Network Based Fuzzy Inference Systems", *IEEE Trans, Syst.ManCybernet*. Vol 23, No 3, pp. 665-685, 1993.

[5] J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.

[6] Patil B.M., Joshi R.C., Toshniwal D., "Association Rule for Classification of Type-2 Diabetic Patients," *Second International Conference on Machine Learning and Computing*(lCMLC), pp.330-334, Feb 2010.

[7] Rakesh Motka, Viral Parmar, Balbindra Kumar, A. R. Verma, "Diabetes Mellitus Forecast Using Different Data Mining Techniques", *4th International Conference on Computer and Communication Technology (ICCCT)*, 978-1-4799-1572-9/13/$31.00 ©2013 IEEE,2013.

[8] M.W. Aslam , A.K.Nandi , " Detection of Diabetes using Genetic Programming",*18th European Signal Processing Conference*, 2010.

[9] K. Kayaer and T. Yildrim, "Medical Diagnosis on Pima Indian Diabetes using General Regression Neural Networks",*International Conference on artificial neural networks and neural information processing*, pp 181-184, 2003.

[10] Xiao Fang, "Are you Becoming a Diabetic ? a Data Mining Approach", *Sixth International Conference on Fuzzy systems and Knowledge Discovery, 2009.*

[11] Poonguzhali E, SabarmathiKabilan, Sandia Kannan, Sivgami P, " Diagnosis of Diabetes Mellitus Type -2 using Neural Network", *International Journal of Emerging Technology and Advanced Engineering, Volume 4,Issue 2, February 2014.*

[12] Madhavi Pradhan,Ketki Kohale, ParagNaikade, AjinkayaPachore, EknathPaiwe, " Design of Classifier for Detection of Diabetes using Neural Network and Fuzzy k-Nearest Neighbour Algorithm", *International Journal of Computational Engineering Research, Vol. 2, Issue 5, pp 1384-1387, September 2012.*

[13] Kalliopi V Dalakleidi, Konstantia Zarkogianni, Vassilios G Karamanos, Anastasia C Thanopoulou, Konstantina S Nikita, "A Hybrid Genetic Algorithm for the Selection of the Critical Features for Risk Prediction of Cardiovascular Complications in Type 2 Diabetes Patients*", IEEE 13th International Conference on Bioinformatics and Bioengineering(BIBE)*, DOI: 10.1109/BIBE.2013.6701620 , pp 1 – 4, 2013.

[14] S. G. Wannamethee, O. Papacosta, P. H. Whincup, C. Carson ,M. C. Thomas ,D. A. Lawlor, S. Ebrahim, N. Sattar, "Assessing prediction of diabetes in older adultsusing different adiposity measures: a 7 year prospectivestudy in 6,923 older men and women",*Springer*, Diabetologia (2010) 53:890–898 ,DOI 10.1007/s00125-010-1670-7.

[15] O. Rolandsson, E. Hagg, M. Nilsson, G. Hallmans , L. Mincheva-nilsson, A . Lernmark, "Prediction of diabetes with body mass index, oral glucosetolerance test and islet cell autoantibodies in a regionalpopulation", *Journal of Internal Medicine* 2001; 249: 279±288.

[16] Meredith F Mackay, Ralph B Agostino, " Prediction of type-2 diabetes using Alternate Anthropometric Measures in a Multi-Ethnic Cohort", *Diabetes Care*, Volume 32, Number 5,May 2009.

[17] C. Kalaiselvi , Dr. G .M. Nasira "A New Approach for Diagnosis of Diabetes and Prediction of Cancer using ANFIS", *World Congress on Computing and Communication Technologies*,2014.

[18] C.P.Ronald Reagan, Dr.S.Prasanna Devi, "An android app for intelligent dosage planning in type2 diabetes using ANFISGA", *International Conference on Recent Trends in Information Technology*,2014.

[19] http://www.swarmintelligence.org/tutorials. php

[20] H.M.I. Pousinho, V.M.F. Mendes , J.P.S. Catalao, "A hybrid PSO-ANFIS approach for short-term wind power prediction in Portuga l".

[21] C.-F. Juang, Combination of particle swarm and ant colony optimization algo-rithms for fuzzy systems design, computer and information science, in: A.T.Azar (Ed.), Numerical Analysis and Scientific Computing. Fuzzy Systems, 2010,ISBN 978-953-7619-92-3, http://dx.doi.org/10.5772/7226.