Data Mining Application in Diabetes Diagnosis using Biomedical Records of Pathological Attribute

Naila[1], Anuradha Sharma[2]

[1]P.G. Student, Department of Computer Science & Engineering, Amity University, Luck now Campus, India

[2]Assistant Professor, Department of Computer Science & Engineering, Amity University, Luck now Campus, India

Abstract: *Data mining aims at discovering knowledge out of data and presenting it in a form that is easily compressible to humans. It is a process that is developed to examine large amounts of data routinely collected. Artificial intelligence technique like fuzzy, ANN etc are currently used for solving a wide range of problems in different application domain for decision based model designing. These systems allows us to introduce the learning and adaptation capabilities hence such type of framework has been used in several different process of diagnosis of disease. It helps in creating computational paradigm that provides a mathematical tool for dealing with the uncertainty and the imprecision typical of human reasoning. Relational between symptoms and risks factors for Diabetic based on the expert's medical knowledge is taken and also related complications or due to some common metabolic disorder it may lead to vision loss, heart failure, stroke, foot ulcer, nerves. In this work review is provided on various methods which are is considered in analysis where symptoms observed in the patient and relation representing the medical knowledge that relates the symptoms in set S to the diseases in set D to diagnose the set B of the possible diseases of the patients can be inferred by means of the compositional rule of inference. It has been observed that Neural Networks are efficiently used for learning membership functions, fuzzy inference rules and other context dependent patterns; fuzzification of neural networks extends their capabilities in applicability [17]. The classifier is the heart of the automatic diagnosis system. The reliable classifier should diagnose the disease at as high accuracy as possible even though there are many uncontrolled variations. In literature, different classifiers have been proposed for automatic diagnosis of PD. The NNs and adaptive neuro fuzzy classifier with linguistic hedges (ANFIS-LH) are investigated for automatic diagnosis of PD. The performance of probabilistic neural network (PNN) for automatic diagnosis of PD is evaluated. SVM classifier is also investigated for the same goal. NNs have some drawbacks such as need of long training time and uncertainties in activation function to be used in hidden layer, number of cells in hidden layer, and the number of hidden layer. In case of SVM, type of kernel function and penalty constant and so forth affects the classification performance. If these parameters are not appropriately selected, the classification performance of SVM degrades. Similarly, the performance of ANFIS depends on type and parameters of membership function and output linear parameters.*

Keywords: Biomedical Data, A. I. Techniques, Data Prediction, ANN, Fuzzy

1. Introduction

Diabetes mellitus is a group of metabolic diseases characterized by hyperglycemias resulting from defects in insulin secretion, insulin action, or both. The chronic hyperglycemias of diabetes are associated with long-term damage, dysfunction, and failure of various organs, especially the eyes, kidneys, nerves, heart, and blood vessels. Several pathogenic processes are involved in the development of diabetes. These range from autoimmune destruction of the -cells of the pancreas with consequent insulin deficiency to abnormalities that result in resistance to insulin action. The basis of the abnormalities in carbohydrate, fat, and protein metabolism in diabetes is deficient action of insulin on target tissues. Deficient insulin action results from inadequate insulin secretion and/or diminished tissue responses to insulin at one or more points in the complex pathways of hormone action. Impairment of insulin secretion and defects in insulin action frequently coexist in the same patient, and it is often unclear which abnormality, if either alone, is the primary cause of the hyperglycemias. Symptoms of marked hyperglycemias include polyuria, polydipsia, weight loss, sometimes with polyphagia, and blurred vision. Impairment of growth and susceptibility to certain infections may also accompany chronic hyperglycemias. Acute, life-threatening consequences of uncontrolled diabetes are hyperglycemias with ketoacidosis or the nonketotic hyperosmolar syndrome.

Long-term complications of diabetes include retinopathy with potential loss of vision; nephropathy leading to renal failure; peripheral neuropathy with risk of foot ulcers, amputations, and Charcot joints; and autonomic neuropathy causing gastrointestinal, genitourinary, and cardiovascular symptoms and sexual dysfunction. Patients with diabetes have an increased incidence of atherosclerotic cardiovascular, peripheral arterial and cerebrovascular disease. Hypertension and abnormalities of lipoprotein metabolism are often found in people with diabetes. The vast majority of cases of diabetes fall into two broad etiopathogenetic categories. In one category, type 1 diabetes, the cause is an absolute deficiency of insulin secretion. Individuals at increased risk of developing this type of diabetes can often be identified by serological evidence of an autoimmune pathologic process occurring in the pancreatic islets and by genetic markers. In the other, much more prevalent category, type 2 diabetes, the cause is a combination of resistance to insulin action and an inadequate compensatory insulin secretory response.

Figure 1: Disorders of glycerine: etiologic types and stages. Even after presenting in ketoacidosis, these patients can briefly return to normoglycemia without requiring continuous therapy (i.e., "honeymoon" remission); in rare instances, patients in these categories (e.g., Vac or toxicity, type 1 diabetes presenting in pregnancy) may require insulin for survival.

In the latter category, a degree of hyperglycemias sufficient to cause pathologic and functional changes in various target tissues, but without clinical symptoms, may be present for a long period of time before diabetes is detected. During this asymptomatic period, it is possible to demonstrate an abnormality in carbohydrate metabolism by measurement of plasma glucose in the fasting state or after a challenge with an oral glucose load. The classifier is the heart of the automatic diagnosis system. The reliable classifier should diagnose the disease at as high accuracy as possible even though there are many uncontrolled variations. In literature, different classifiers have been proposed for automatic diagnosis of PD. The NNs and adaptive neuro fuzzy classifier with linguistic hedges (ANFIS-LH) are investigated for automatic diagnosis of PD. The performance of probabilistic neural network (PNN) for automatic diagnosis of PD is evaluated. SVM classifier is also investigated for the same goal. NNs have some drawbacks such as need of long training time and uncertainties in activation function to be used in hidden layer, number of cells in hidden layer, and the number of hidden layer. In case of SVM, type of kernel function and penalty constant and so forth affects the classification performance. If these parameters are not appropriately selected, the classification performance of SVM degrades. Similarly, the performance of ANFIS depends on type and parameters of membership function and output linear parameters.

2. Related Work

The numbers of medical decision support systems are employed using different approaches. George et al. have anticipated decision support system to classify and detect agitation transition. In this system support vector machines is used for detection. This system is for Dementia patients. This system presents a decision confidence measure and two new SVM architectures, which were useful to agitation detection and agitation transition detection. An accuracy of 91.4% was achieved; in assessment with 90.9% for the conventional SVM [1]. Haitham and Alan have projected automated recognition of obstructive sleep apnoea syndrome using SVM classifier. In this study, they evaluated features from the magnitude and phase of the thoracic and abdominal respiratory effort signals for OSA detection. This is supports on the physiological fact that all through normal breathing the abdominal and thoracic efforts happen simultaneously. The aim of this study is to appraise classification of whole night regular and apnoeic epochs using mined features from the phase and magnitude of the respiratory efforts signals, compared and combined with some other features from HRV and oxygen saturation signals [2] [3]. Support Vector machines have also been utilized in decision support systems such as [4].An intelligent system based support vector machine along with a radial basis function network is accessible for the diagnosis. The support vector machine with sequential minimal optimization algorithm is applied to India based patients' data set. Then, the Radial Basis Function (RBF) network structure qualified by Orthogonal Least Square (OLS) algorithm is functional to same data set for predictions [5]. Tsai and Watanabe proposed a genetic algorithm(GA) based method and implemented for influential the set of fuzzy membership functions that can provide an optimal classification of myocardial heart disease from ultrasonic images. In this method an average classification rate of 96% is achieved [6]. In a different advance genetic algorithm is used to determine the attributes which donate more towards the diagnosis of heart ailments which ultimately reduces the number of tests which are desirable to be taken by a patient [7]. **Yang and Honavar** have projected a feature subset algorithm using genetic algorithm. A genetic algorithm to decide on optimal feature subset for use with back propagation artificial neural networks has been described [8]. A genetic algorithm for feature selection as well as for optimization of Support Vector Machine (SVM) parameter has been proposed by Haung. The projected method performs feature selection and parameters setting in an evolutionary way [9]. Very recently, a real coded Genetic algorithm for critical feature analysis for heart disease diagnosis has been portrayed [10]. Diagnosis of Heart Disease using Data mining Algorithm proposed by Rajkumar and Sophia. In their approach the preliminary diagnosis of a heart attack is made by a combination of clinical symptoms and characteristic electrocardiogram (ECG) changes. The accuracy with this technique is 52.33% [11]. Palaniappan and Awang proposed Intelligent Heart Disease Prediction System Using Data Mining Techniques. This investigation has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Nave Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals [12].

3. Methodology

An adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system. In early 1990's this technique was developed. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions. Adaptive Neuro-Fuzzy Inference System (ANFIS) is considered to be a universal estimator.

Adaptive neuro-fuzzy inference system is a fuzzy inference system implemented in the framework of an adaptive neural network. By using a hybrid learning procedure, ANFIS can construct an input/output mapping based on both human-knowledge as fuzzy if-then rules and approximate membership functions from the stipulated input-output data pairs for neural network training. This procedure of developing a FIS using the framework of adaptive neural networks is called an adaptive neuro fuzzy inference system (ANFIS). There are 2 methods that ANFIS learning employs for change membership function parameters: 1) back propagation for all parameters (a steepest descent method), and 2) a hybrid technique consisting of back propagation for the parameters related to the input membership and method of least squares estimation for the parameters related to the output membership functions. As a result, the training error decreases a minimum of locally, throughout the learning method. It applies the least-squares technique to identify the resultant parameters that define the coefficients of every output equation within the Sugeno-type fuzzy rule base. The training process continues until the required number of training steps (epochs) or the required root mean square error (RMSE) between the required and therefore the generated output is achieved. This study uses a hybrid learning algorithm, to identify premise and consequent parameters of first order Takagi-Sugeno type fuzzy system for predicting surface roughness in ball end milling.

Adaptive Neuro Fuzzy Inference System (ANFIS) was first proposed by Jang [4, 5, and 6]. ANFIS is suitable for tackling any input output relationship and thus it can be applied to many Practical applications. Their applications in different areas are worth mentioning. ANFIS is the first of its kind of NFS when applied to data classification. It is a hybrid model using both the ANN and FIS into one capsule. Thus once developed, there are no differences to mark out into ANN and FIS [7]. Finally, the predicted diagnosis is evaluated by a clinical specialist. The major steps can be summarized as
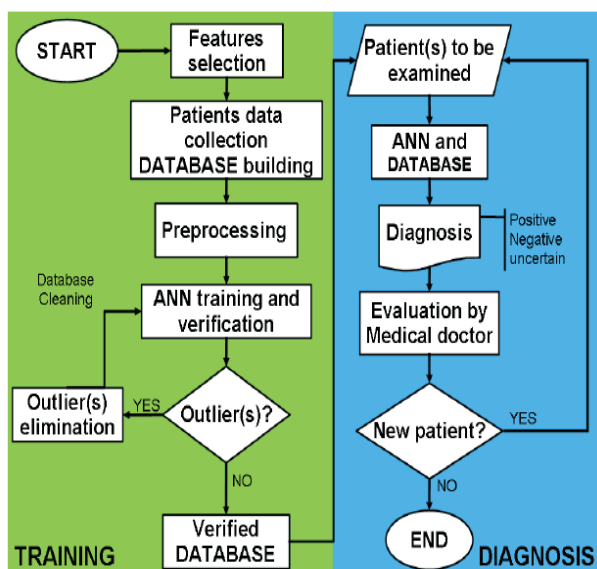
Figure 2: Diagram of fundamental steps in ANNs-based medical diagnosis. Building of the database and "learning" represents the left half (green) and its application for the diagnosis is the right part (blue).

3.1 Fuzzy Logic

It is a conceptual framework of classical logic. It was framed by Lotfi Zadeh [8] within the mid-1960s to model those issues during which inaccurate data should be used or in which the rules of inference are developed during a} very general means making use of diffused categories [8]. In fuzzy logic that is additionally generally known as diffuse logic there aren't simply 2 alternatives however a full continuation of truth values for logical propositions. A proposition A will have the truth value zero.4 and its complement A^c the reality worth zero. In line with the kind of negation operator that's used, the 2 truth values should not be essentially add up to one. Fuzzy logic features a weak association to probability theory. Fuzzy logic doesn't have to be compelled to be even employing a probability technique. The general system is to simplify the result of multi-valued logic in such a way on preserve a part of the algebraically structure [8]. Fuzzy set theory corresponds to fuzzy logic and therefore the semantic of fuzzy operators is understood employing a geometric model. The geometric idea of fuzzy logic can provide us a touch on the potential reference to neural networks. Fuzzy logic can be used for interpreting the properties of neural network model, moreover as for giving a additional precise description of their performance. Fuzzy logic may be used to specify networks directly while not having to apply a learning algorithmic rule. An expert in a certain field will generally produce control rules for a system which is dynamic, with less effort than those pertaining to training a neural network. An example given by Zadeh to the people of neural network is to develop a system to park a car. It's simple to formulate a collection of fuzzy rules for this task; however it's not immediately obvious the way to neither build a network to try and do the same nor how to train it. Fuzzy logic is currently being used in several products of business and consumer electronics for which a decent system is ample and wherever the question of optimal control doesn't essentially arise.

The process of fuzzy logic is explained in algorithmic rule. Firstly, converting a crisp input data into fuzzy set using fuzzy linguistic variable, fuzzy linguistic terms and membership functions. This step is understood as fuzzification. Afterwards, an inference is created which supported a collection of rules. Lastly, the fuzzy output is mapped to a crisp output using the membership functions, within the defuzzification step.

3.1.1 Fuzzy logic algorithm
1) Define the linguistic variables and terms (initialization)
2) Construct the membership functions (initialization)
3) Construct the rule base (initialization)
4) Convert crisp input data to fuzzy values using the membership functions (fuzzification)
5) Evaluate the rules in the rule base (inference)
6) Combine the results of each rule (inference)
7) Convert the output data to non-fuzzy values (defuzzification)

3.2 Proposed Work

The forecast of Internet traffic issue has received little attention from the computer networks field. By improving

this task, efficient traffic engineering and anomaly detection tools can be created, resulting in economic gains from better resource management. Thus the present study is about using soft computing technique, ANFIS for the prediction of network traffic using time series data analysis. For this freely downloadable online network traffic data is used for carrying out the analysis on MATLAB environment. The performance evaluation of the prediction model will be done using RMSE, Regression analysis, etc.

3.2.1 Algorithm

The algorithm for the work is given below
1. Start MATLAB.
2. Load data.
3. Divide the data into training and testing data sets using Matlab commands.
4. Start ANFIS Editor using commands.
5. Load training data into ANFIS editor.
6. Generate Fuzzy Inference System (FIS) using Subtractive clustering algorithm.
   a. Input Selection: Number and type of input / output membership functions.
7. ANFIS Training
   a. Optimization method selection: Error tolerance, no. of epochs.
8. ANFIS Testing
   a. Plot ANFIS output against Observed training and testing data.
9. Record the plot of training and testing datasets.
10. ANFIS Evaluation based on RMSE i.e. RMSE < RMSE ref?
11. No .got to step 5 and repeat step 6 to 10 ,else
12. END

4. Result and Discussion

This chapter covers the simulation results of the algorithm developed using MATLAB for detection of diabetes using ANFIS approach from minimum number of data inputs selection using regression analysis. The database that is used is shown below.
Data name =
Nop   : no. of time pregnant
Gconc   : glucose concentrations
Bp   : blood pressure
Skin thick: skin thickness
Insulin: insulin level

Bmi   : bmi
Pedigree: pedigree value
Age
Output   : diabetes present or not
Thus the data consist of 8 types of input values and the one output. There are 768 input samples for all 8 inputs some of the data and respective output data are shown in table 1.

**Table 1:** Data record for diabetes symptoms and respective input.

S. No.	I/P 1 Nop	I/P 2 gconc	I/P 3 bp	I/P 4 Skinthick	I/P 5 insulin	I/P 6 bmi	I/P 7 pedigree	I/P 8 age	O/P output
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	0	0	25.6	0.201	30	0
7	3	78	50	32	88	31	0.248	26	1
8	10	115	0	0	0	35.3	0.134	29	0
9	2	197	70	45	543	30.5	0.158	53	1
10	8	125	96	0	0	0	0.232	54	1
11	4	110	92	0	0	37.6	0.191	30	0
12	10	168	74	0	0	38	0.537	34	1
13	10	139	80	0	0	27.1	1.441	57	0
14	1	189	60	23	846	30.1	0.398	59	1
15	5	166	72	19	175	25.8	0.587	51	1
16	7	100	0	0	0	30	0.484	32	1
17	0	118	84	47	230	45.8	0.551	31	1
18	7	107	74	0	0	29.6	0.254	31	1
19	1	103	30	38	83	43.3	0.183	33	0
20	1	115	70	30	96	34.6	0.529	32	1
21	3	126	88	41	235	39.3	0.704	27	0
22	8	99	84	0	0	35.4	0.388	50	0
23	7	196	90	0	0	39.8	0.451	41	1
24	9	119	80	35	0	29	0.263	29	1
25	11	143	94	33	146	36.6	0.254	51	1
26	10	125	70	26	115	31.1	0.205	41	1
27	7	147	76	0	0	39.4	0.257	43	1
28	1	97	66	15	140	23.2	0.487	22	0
29	13	145	82	19	110	22.2	0.245	57	0
30	5	117	92	0	0	34.1	0.337	38	0
31	5	109	75	26	0	36	0.546	60	0
32	3	158	76	36	245	31.6	0.851	28	1
33	3	88	58	11	54	24.8	0.267	22	0
34	6	92	92	0	0	19.9	0.188	28	0
35	10	122	78	31	0	27.6	0.512	45	0

All the input values are normalized by diving it with respective maximum value. By normalization all the input shifts to a similar range of 0 to 1 as shown in table 2.

Table 2: Data record for diabetes symptoms and respective input after normalization.

S.No.	I/P 1 Nop	I/P 2 gconc	I/P 3 bp	I/P 4 skinthick	I/P 5 insulin	I/P 6 bmi	I/P 7 pedigree	I/P 8 age	O/P output
1	0.353	0.744	0.59	0.354	0	0.5	0.259	0.62	1
2	0.059	0.427	0.541	0.293	0	0.4	0.145	0.38	0
3	0.471	0.92	0.525	0	0	0.35	0.278	0.4	1
4	0.059	0.447	0.541	0.232	0.11	0.42	0.069	0.26	0
5	0	0.688	0.328	0.354	0.2	0.64	0.945	0.41	1
6	0.294	0.583	0.607	0	0	0.38	0.083	0.37	0
7	0.176	0.392	0.41	0.323	0.1	0.46	0.102	0.32	1
8	0.588	0.578	0	0	0	0.53	0.055	0.36	0
9	0.118	0.99	0.574	0.455	0.64	0.45	0.065	0.65	1

10	0.471	0.628	0.787	0	0	0	0.096	0.67	1
11	0.235	0.553	0.754	0	0	0.56	0.079	0.37	0
12	0.588	0.844	0.607	0	0	0.57	0.222	0.42	1
13	0.588	0.698	0.656	0	0	0.4	0.595	0.7	0
14	0.059	0.95	0.492	0.232	1	0.45	0.164	0.73	1
15	0.294	0.834	0.59	0.192	0.21	0.38	0.243	0.63	1
16	0.412	0.503	0	0	0	0.45	0.2	0.4	1
17	0	0.593	0.689	0.475	0.27	0.68	0.228	0.38	1
18	0.412	0.538	0.607	0	0	0.44	0.105	0.38	1
19	0.059	0.518	0.246	0.384	0.1	0.65	0.076	0.41	0
20	0.059	0.578	0.574	0.303	0.11	0.52	0.219	0.4	1
21	0.176	0.633	0.721	0.414	0.28	0.59	0.291	0.33	0
22	0.471	0.497	0.689	0	0	0.53	0.16	0.62	0
23	0.412	0.985	0.738	0	0	0.59	0.186	0.51	1
24	0.529	0.598	0.656	0.354	0	0.43	0.109	0.36	1
25	0.647	0.719	0.77	0.333	0.17	0.55	0.105	0.63	1
26	0.588	0.628	0.574	0.263	0.14	0.46	0.085	0.51	1
27	0.412	0.739	0.623	0	0	0.59	0.106	0.53	1
28	0.059	0.487	0.541	0.152	0.17	0.35	0.201	0.27	0
29	0.765	0.729	0.672	0.192	0.13	0.33	0.101	0.7	0
30	0.294	0.588	0.754	0	0	0.51	0.139	0.47	0
31	0.294	0.548	0.615	0.263	0	0.54	0.226	0.74	0
32	0.176	0.794	0.623	0.364	0.29	0.47	0.352	0.35	1
33	0.176	0.442	0.475	0.111	0.06	0.37	0.11	0.27	0
34	0.353	0.462	0.754	0	0	0.3	0.078	0.35	0
35	0.588	0.613	0.639	0.313	0	0.41	0.212	0.56	0
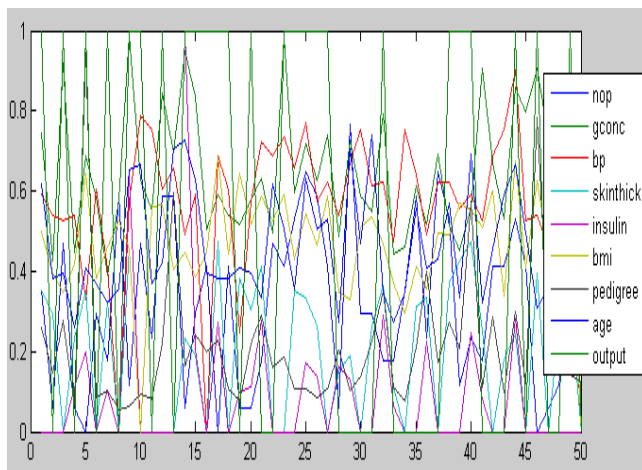
**Figure 3:** Plot showing 50 samples of all the normalized input data along with output.

Figure 3 shows the plot of all the 8 input signals after normalization. We can observe that data input are ranging from 0 to 1.Hence all of them are in similar range now we can check the dependency of output using the regression analysis.

Now it shows that for the determination of diabetes we can work out with pedigree (7), no. of time pregnancy (1); blood pressure (3); bmi (6) and glucose concentration in descending order of significance. Thus the input data set will consist of 5 inputs with 768 samples this data will be further divided into training and testing data for modelling the ANFIS predictor for diabetes.
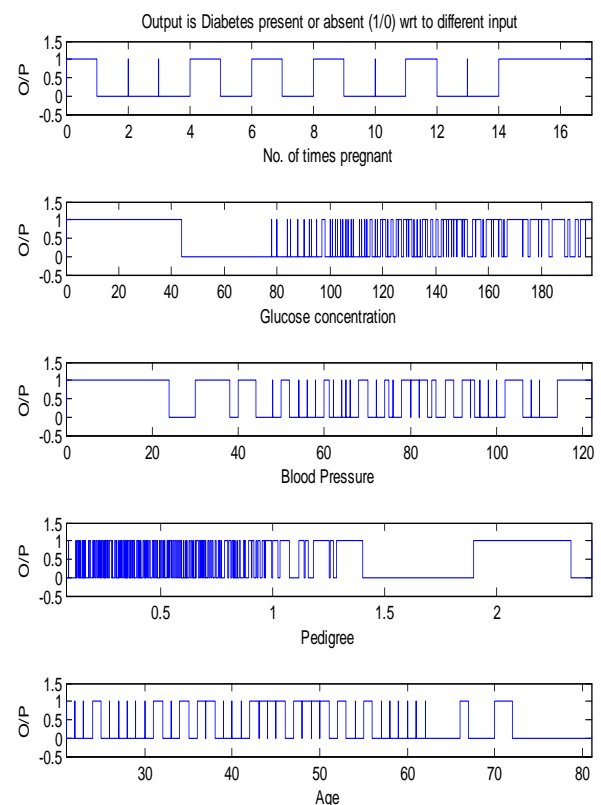

Figure 4: Plot of diabetes presence wrt five significant inputs in individually.

After applying subtractive clustering the the generated FIS system is updated by the ANFIS training. The training is performed for 100 epochs and the plot of prediction error is shown in figure 4.
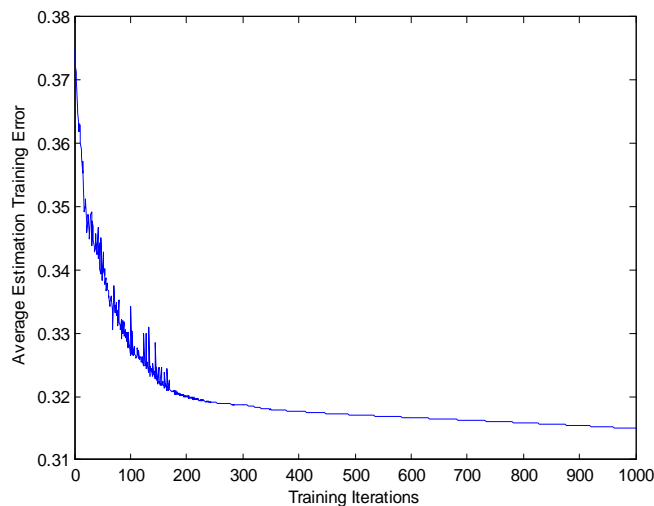
**Figure 5:** Training Error plot at different iterations.

From figure 5 we can observe that the error in prediction of diabetes using the 5 significant input is further reduced from 0.38 to 0.31.now after getting the ANFIS model testing is performed to check the accuracy on remaining 30 % data. The testing accuracy is found to be 93%.
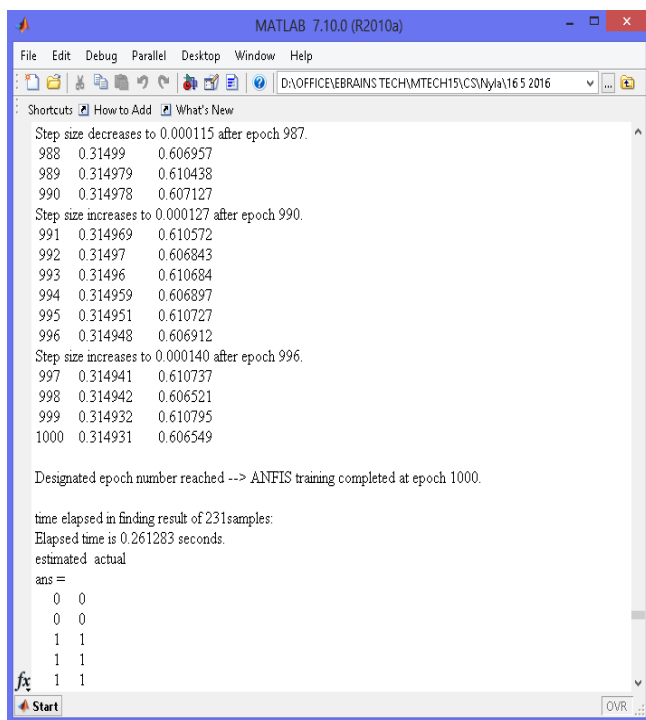
Figure 6: Snapshot at the end of ANFIS training

The details of developed ANFIS model are given below:
ANFIS info:
Number of nodes: 140
Number of linear parameters: 66
Number of nonlinear parameters: 110
Total number of parameters: 176
Number of training data pairs: 538
Number of checking data pairs: 231
Number of fuzzy rules: 11
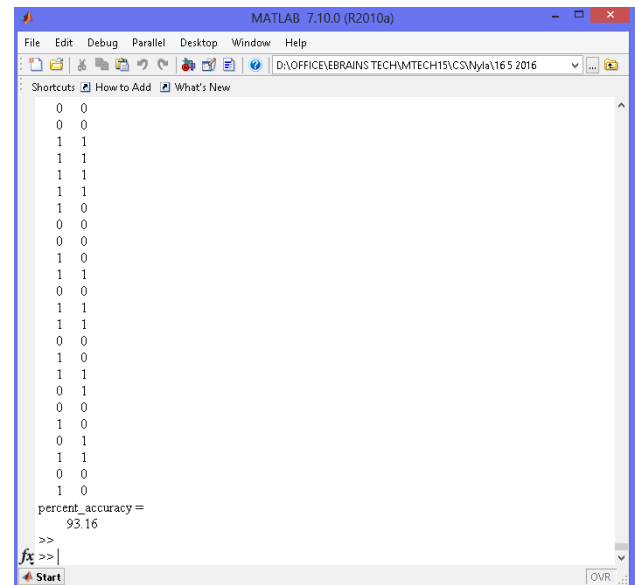
Figure 7: Snap after testing to display the testing accuracy.

5. Conclusion

Efficient prediction schemes are very important in order to achieve required accuracy of biomedical diagnosis. The goal of this thesis was to predict the presence of diabetes for an efficient allotment of minimum input variables biomedical signal using ANFIS. Here it was demonstrated that models developed using ANFIS technique could be used for solving this critical issue. It is clearly demonstrated that ANFIS can be trained to accurately predict in prediction of presence of diabetes using variable fuzzy process parameters. The interpretation results in chapter 4 demonstrate one of the real strengths of ANFIS is that they perform well even when the training data contains noise and measurement errors. That is, during learning, ANFIS are able to filter out noise and measurement error and effectively generalize the system behavior. From the analysis of the results, given under heading "Results and Discussions", Chapter-4, it is seen that the prediction model developed using ANFIS technique has been able to perform well. The MAE value obtained from ANFIS model is 0.31 for training datasets. Accuracy is of 93% is obtained for testing data.

References

[1] G. E. Sakr, I. H. Elhajj and H. A. Huijer, "Support vector machines to define and detect agitation transition," IEEE Transactions On Affective Computing, vol. 1, pp. 98-108, December 2010.

[2] M. Haitham, A. Angari and A. V. Sahakian, "Automated recognition of obstructive sleep apnea syndromeusing support vector machine classiifier," IEEE Transactions On Information Technology In Biomedicine,vol. 16, pp. 463-468, May 2012.

[3] F. Azuaje, W. Dubitzky, P. Lopes, N. Black and K. Adamsom, "Predicting coronary disease risk based on shortterm rr interval measurements: A neural network approach," Artificial Intelligence in Medicine, vol.15, pp. 275-297, March 1999.

[4] E. Comak, A. Arslan and T. Ibrahim, "A decision support system based on support vector machines

fordiagnosis of the heart valve diseases," Computers in biology and Medicine, vol. 37, pp. 21-27, January 2007.

[5] S. Ghumbre, C. Patil, and A. Ghatol, "Heart disease diagnosis using support vector machine,"International Conference on Computer Science and Information Technology (ICCSIT), pp. 84-88,December 2011.

[6] D. Y. Tsai and S. Watanabe, "Method for optimization of fuzzy reasoning by genetic algorithms and is application to discrimination of myocardial heart disease," IEEE Nuclear Science Symposium and Medico1 Imaging Conference, pp. 2239-2246, December 1966.

[7] E. A. M. Anbarasi and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," International Journal of Engineering Science and Technology, vol. 2, pp. 5370-5376, November 2010.

[8] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," IEEE Intelligent Systems,pp. 44-49, March 1998.

[9] C. L. Huang and C. J.Wang, "A ga-based feature selection and parameters optimization for support vector machines," Expert Systems with applications, vol. 31, pp. 231-240, October 2006.

[10] J. Z. H. Yan and C. Xiao, "Selecting critical clinical features for heart diseases diagnosis with a realcoded genetic algorithm," Applied Soft Computing, vol. 8, pp. 1105-1111, March 2008.

[11] A. Rajkumar and G. S. Reena, "Diagnosis of heart disease using datamining algorithm," Global Journal of Computer Science and Technology, vol. 10, pp. 38-43, December 2010.

[12] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques,"International Journal of Computer Science and Network Security, pp. 343-350, January 2008.

[13] W. G. Baxt, "Application of artificial neural networks to clinical medicine," Lancet, vol.346, pp. 1135-1138, October 1995

[14] Sarvestan Soltani A. , Safavi A. A., Parandeh M N. and Salehi M., "Predicting Breast Cancer Survivability using data mining techniques," Software Technology and Engineering (ICSTE), 2nd International Conference, 2010, vol.2, pp.227-231.

[15] Anunciacao Orlando, Gomes C. Bruno, Vinga Susana, Gaspar Jorge, Oliveira L. Arlindo and Rueff Jose, "A Data Mining approach for detection of highrisk Breast Cancer groups," Advances in Soft Computing, vol. 74, pp. 43-51, 2010.

[16] Naila, Anuradha Sharma, "Survey on Feature based Prediction in Biomedical application by Advanced Data Mining Technique", International Journal of Research and Development in Applied Science and Engineering (IJRDASE) , Vol.9, Issue 2, April 2016, available at www.ijrdase.com.

interests are Data Mining Techniques Using ANFIS (Adaptive Neuro Fuzzy Inference System) Algorithm etc.

Author Profile

NAILA is a Research Scholar in the Computer Science and Engineering Department, College of Amity School Of Engineering, Amity University, Luck now Campus, Uttar Pradesh. I received Master of Technology (M.Tech) degree in 2016-06-10 from Amity University, Luck now Campus, Uttar Pradesh. In My research