

Automated Detection of Diabetes by Regression Analysis and ANFIS approach

Chhavi Bajpai

cbajpai7@gmail.com

Department of Computer Science and Engineering

BNCET, Lucknow UP

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. 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. In this paper, we have tried to develop a method for the prediction of type 2 diabetes using adaptive neuro-fuzzy inference system (ANFIS) with regression analysis.

Keywords: ANN, ANFIS, Fuzzification, Biomedical.

1. Introduction:

Diabetes mellitus is a collection of metabolic illnesses characterized through hyperglycemia because of defects in insulin secretion, insulin motion, or each. The chronic hyperglycemia of diabetes is related to lengthy-term damage, disorder, and failure of several organs, in particular the eyes, kidneys, nerves, coronary heart, and blood vessels. Several pathogenic tactics are involved within the improvement of diabetes. This varies from autoimmune destruction of the cells of the pancreas with consequent insulin deficiency to abnormalities that bring about resistance to insulin motion. The basis of the abnormalities in carbohydrate, fat, and protein metabolism in diabetes is terrible action of insulin on the proper music tissues. Deficient insulin motion consequences from insufficient insulin secretion and/or faded tissue responses to insulin at one or extra elements in the complex pathways of hormone action. Impairment of insulin secretion and defects in insulin movement regularly coexist within the equal patient, and it's far often dubious which abnormality, if either by myself, is the number one reason of the hyperglycemia. Symptoms of marked hyperglycemia consist of polyuria, polydipsia, weight reduction, on occasion with polyphagia, and blur vision. Impairment of boom and

susceptibility to tremendous infections may additionally accompany persistent hyperglycemia. Acute, existence-threatening results of out of manipulate diabetes are hyperglycemia with ketoacidosis or the nonketotic hyperosmolar syndrome.

Long-time period complications of diabetes consist of retinopathy with capacity loss of vision; nephropathy main to renal failure; peripheral neuropathy with risk of foot ulcers, amputations, and Charcot joints; and autonomic neuropathy inflicting gastrointestinal, genitourinary, and cardiovascular symptoms and signs and sexual ailment. Patients with diabetes have an extended incidence of atherosclerotic cardiovascular, peripheral arterial and cerebrovascular illness. Hypertension and abnormalities of lipoprotein metabolism are often found in humans with diabetes. The enormous majority of times of diabetes fall into two sizeable etiopathogenetic lessons. In magnificence, kind 1 diabetes, the reason is an absolute deficiency of insulin secretion. Individuals at extended hazard of developing this form of diabetes can regularly be recognized with the resource of serological proof of an autoimmune pathologic gadget occurring in the pancreatic islets and with the resource of genetic markers. In the alternative, lots extra ordinary category, the kind 2 diabetes, the motive is a combination of resistance to insulin movement and an inadequate compensatory insulin secretory response. In the latter elegance, a degree of hyperglycemia enough to purpose pathologic and realistic adjustments in numerous target tissues, but without clinical symptoms, may be gift for a long time length before diabetes is detected. During this asymptomatic length, it's far feasible to illustrate an abnormality in carbohydrate metabolism with the aid of measurement of plasma glucose within the fasting United States or after a venture with an oral glucose load.

2. Related Work:

The numbers of medical choice aid systems are employed the use of one-of-a-kind approaches. George et al. Have predicted decision help machine to classify and locate agitation transition. In this gadget assist vector machines is used for detection. This device is for Dementia patients. This system offers a choice self assurance degree and new SVM architectures, which were useful to agitation detection and agitation transition detection. An accuracy of 91.4% become accomplished; in evaluation with 90.9% for the conventional SVM [1]. Haitham and Alan have projected automated reputation of obstructive sleep apnea syndrome the use of SVM classifier. In this have a look at, they evaluated

capabilities from the magnitude and phase of the thoracic and abdominal respiratory effort alerts for OSA detection. This supports at the physiological reality that all through everyday breathing the abdominal and thoracic efforts occur simultaneously. The aim of this take a look at is to appraise class of entire night ordinary and apneic epochs the usage of mined features from the section and value of the breathing efforts alerts, compared and combined with a few different capabilities from HRV and oxygen saturation alerts [2] [3]. Support Vector machines have additionally been utilized in decision guide structures along with [4]. An wise system based totally support vector system along with a radial basis function community is accessible for the prognosis. The guide vector machine with sequential minimum optimization set of rules is applied to India primarily based patients' data set. Then, the Radial Basis Function (RBF) community shape qualified by means of Orthogonal Least Square (OLS) algorithm is practical to same records set for predictions [5]. Tsai and Watanabe proposed a genetic algorithm (GA) based totally method and carried out for influential the set of fuzzy membership capabilities that may offer an most excellent type of myocardial heart disorder from ultrasonic snap shots. In this approach a mean class charge of ninety six% is done [6]. In a one of a kind strengthen genetic set of rules is used to decide the attributes which donate greater in the direction of the prognosis of heart ailments which in the long run reduces the quantity of exams that are ideal to be taken by using a affected person [7]. Yang and Honavar have projected a characteristic subset algorithm the use of genetic set of rules. A genetic algorithm to determine on most useful function subset to be used with again propagation artificial neural networks has been described [8]. A genetic set of rules for characteristic selection as well as for optimization of Support Vector Machine (SVM) parameter has been proposed by Haung. The projected technique plays function selection and parameters placing in an evolutionary way [9]. Very these days, a actual coded Genetic algorithm for critical characteristic evaluation for coronary heart disorder diagnosis has been portrayed [10]. Diagnosis of Heart Disease using Data mining Algorithm proposed by Rajkumar and Sophia. In their technique the initial prognosis of a coronary heart assault is made through a mixture of scientific signs and symptoms and feature electrocardiogram (ECG) modifications. The accuracy with this approach is fifty two.33% [11]. Palaniappan and Awang proposed Intelligent Heart Disease Prediction System Using Data Mining Techniques. This research has evolved a prototype Intelligent Heart Disease Prediction System (IHDP) the use of information mining techniques, namely, Decision Trees, Nave Bayes and Neural Network. Results display that each approach has its specific power in realizing the objectives of the defined mining dreams [12]. The features of the artificial neural community (ANN), excessive accuracy and learning charge, make it worth trying as an algorithm to the prediction of coronary heart disorder [13].

3. Methodology:

An adaptive neuro-fuzzy inference machine or adaptive network-based totally fuzzy inference gadget (ANFIS) is a sort of synthetic neural network that is based on Takagi-Sugeno fuzzy inference device. In early 1990's this technique modified into evolved. Since it integrates both neural networks and fuzzy logic standards, it has capability to seize the advantages of each in a single framework. Its inference tool corresponds to a set of fuzzy IF-THEN guidelines which have learning functionality to approximate nonlinear capabilities. Adaptive Neuro-Fuzzy Inference System (ANFIS) is taken into consideration to be a commonplace estimator.

Adaptive neuro-fuzzy inference device is a fuzzy inference device applied in the framework of an adaptive neural community. By using a hybrid reading method, ANFIS can assemble an input output mapping based totally mostly on both human-information as fuzzy if-then regulations and approximate club capabilities from the stipulated input-output data pairs for neural network schooling. This method of growing a FIS the usage of the framework of adaptive neural networks is called an adaptive neuro fuzzy inference system (ANFIS). There are 2 techniques that ANFIS mastering employs for exchange membership function parameters: 1) decrease lower back propagation for all parameters (a steepest descent technique), and 2) a hybrid approach collectively with once more propagation for the parameters associated with the input membership and technique of least squares estimation for the parameters related to the output membership features. As a cease end result, the education errors decreases, at the least regionally, during the studying method. It applies the least-squares technique to pick out the following parameters that define the coefficients of every output equation in the Sugeno-type fuzzy rule base. The schooling process keeps till the desired variety of schooling steps (epochs) or the desired root imply rectangular errors (RMSE) a number of the required and therefore the generated output is completed. This study makes use of a hybrid studying set of rules, to discover premise and consequent parameters of first order Takagi-Sugeno type fuzzy gadget for predicting floor roughness in ball prevent milling.

Adaptive Neuro Fuzzy Inference System (ANFIS) turned into first proposed with the aid of Jang [4, 5, 6]. ANFIS is appropriate for tackling any enter output courting and as a end result it is able to be applied to many realistic packages. Its applications in extraordinary regions are worth mentioning. ANFIS is the primary of its sort of NFS at the same time as applied to information class. It is a hybrid version the use of both the ANN and FIS into one pill. Thus once superior, there aren't any variations to mark out into ANN and FIS [7].

Proposed Work:

The present study is about using soft computing technique, ANFIS to predict the presence of diabetes for an efficient allotment of minimum input variables biomedical signal. 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.

The algorithm for the work is given below

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

4. Result and Discussion:

This section covers the simulation results of the algorithm developed using MATLAB for detection of diabetes using ANFIS approach from minimum number of data input 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
- skinthick: skin thickness
- insulin: insulin level
- bmi: body mass index
- pedigree: pedigree value age
- output: diabetes present or not

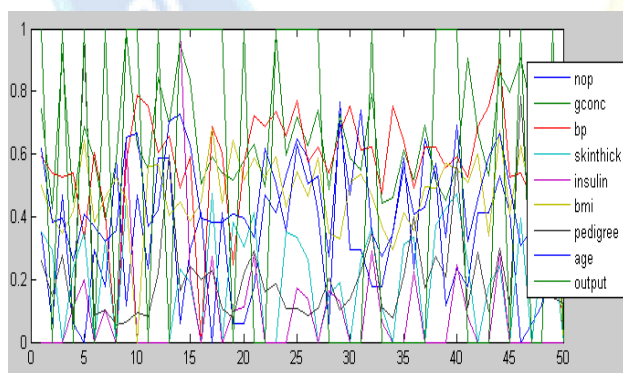


Fig 1: Plot showing 50 samples of all the normalized input data along with output.

The above figure 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 modeling the ANFIS predictor for diabetes.

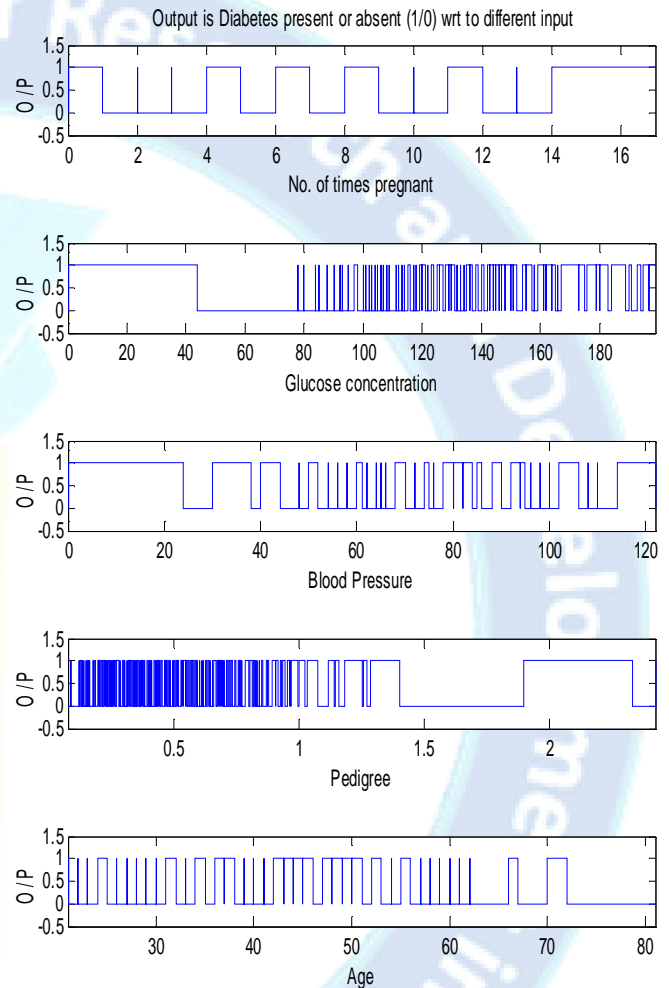


Fig 2: Plot of diabetes presence with respect to five significant inputs in individually

For developing ANFIS model the significant input data matrix is shuffled by randomly distributing them and thereafter 70 % data is taken as training data out of 768 samples and remaining is kept for testing the performance accuracy of developed prediction model. Initial Fuzzy inference system is generated by subtractive clustering method for 3 membership functions partitions of all 5 input.

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 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 behaviour. The MAE value obtained from ANFIS model is 0.31 for training datasets. Accuracy is of 93% is obtained for testing data.

In future a fixed centre-based approach to membership function selection may be added for the fuzzy system design to make an effective and concrete decision of the number of fuzzy output values, each with the same function width. The connectionist approach can also be hybrid to reduce human efforts, such as using GAs, to select membership functions is based on training a neural network. The improved gaussian membership function may also offer an alternative approach to let the data classification nature of the membership functions. In future scope such work can be shown that can lead to satisfactory performance for fuzzy and neural networks. Further research is planned towards on-line adaptation of MFs in an optimization structure.

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 syndrome using support vector machine classifier," 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 for diagnosis 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 its application to discrimination of myocardial heart disease," IEEE Nuclear Science Symposium and Medical 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.