Application of Adaptive Neuro-Fuzzy Inference System for Diabetes Classification and Prediction

Oana Geman[1], Iuliana Chiuchisan[2], Roxana Toderean (Aldea)[1]
[1]Health and Human Development Department, Stefan cel Mare University of Suceava, Suceava, Romania, oana.geman@usm.ro
[2]Computers, Electronics and Automation Department, Stefan cel Mare University of Suceava, Suceava, Romania, iuliana.chiuchisan@usm.ro

***Abstract*— Diabetes is one of the most common metabolic diseases and the statistics show that one in eleven adults has diabetes, but one in two adults with diabetes is undiagnosed, and in 2040 one in 10 adults will have diabetes. In this paper is proposed a hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) model for classifying patients with diabetes based on data sets with diabetic patients (Pima Indians Diabetes Dataset). The Pima Indians Diabetes Dataset contains 768 samples. In order to set the features vector of this system is used Diabetes Pedigree Function to define the fuzzy rule base with multiple premises. The Neuro-Fuzzy ANFIS modeling was implemented using ANFIS Fuzzy Logic Toolbox and MATLAB Toolbox. The performances of the algorithm were analyzed in terms of specificity, precision and sensitivity. The proposed neural network was trained and tested on Pima Indians Diabetes Database, proving an accuracy of 85.35% for training data and 84.27% for testing data.**

Keywords*— *Diabetes; Classification; Neural network; Adaptive Neuro-Fuzzy Inference System; Diabetes Pedigree Function.

I. Introduction

According to the World Health Organization (WHO) report, over the last 10 years, diabetes prevalence has been rising rapidly in middle-income and low-income countries [1]. Diabetes type 1 occurs when the pancreas does not produce enough insulin to regulate blood sugar levels. In diabetes type 2, the body does not properly use the insulin it produces. In both cases, the disease can cause a number of complications such as myocardial infarction, stroke, renal or vascular problems. Globally, the number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014 [1], according to the WHO.

Worldwide, only in 2016, a number of 1.5 million people died from diabetes, to be added to the 2.2 millions of deaths from diseases related to diabetes, leading to a total of 3.7 million deaths [1]. WHO states that by 2030 diabetes will be the seventh leading cause of death worldwide and the direct and indirect costs for the 2011 - 2030 periods are estimated at 1,700 billion dollars [1]. Also, the statistics show that one in eleven adults has diabetes, but one in two adults with diabetes is undiagnosed, and in 2040 one in 10 adults will have diabetes [1].

II. II. Diabetes Database and Diabetes Pedigree Function

In the study presented in this paper is used the Pima Indian Diabetes Dataset, available free on UCI Machine Learning Database from Department of Information and Computer Science, University of California [2]. The dataset was selected from a larger database held by National Institute of Diabetes and Digestive and Kidney Diseases [2]. In this database the diagnostics are binary-valued variable and all patients have assigned the 0 or 1 values, where 0 is negative test for diabetes (500 number of instances) and 1 is a positive test (268 number of instances).

The features vector contains the following clinical attributes [2]:

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;
9. Class variable (0 or 1).

In reference [3], authors developed a function named Diabetes Pedigree Function (DPF) that uses the genetic relationship of the subject's relatives in order to provide a synthesis of the Diabetes Mellitus (DM) history for the patient [3]. Diabetes Pedigree Function uses the information from relatives (parents, grandparents, siblings, first cousins, aunts and uncles). This function provides "a measure of the expected genetic influence of affected and unaffected relatives on the subject's eventual diabetes risk" [3]. The DPF [3] has the constants 88 and 14 which represent "the maximum and minimum ages at which relatives of the subjects developed diabetes mellitus (DM)" [3]. Also, the authors of the paper [3]

978-1-5386-0358-1/17/$31.00 ©2017 IEEE

have set the constants 20 and 50, according with DPF values for the relatives of the subject.

III. Adaptive Neural Fuzzy Inference System (ANFIS)

A survey of Artificial Neural Network (ANN) applications in various medical systems are presented in references [4] – [10]. Also, a series of Neuro-Fuzzy applications are presented in papers [11] – [14]. During the years, the authors of this paper have conducted research on the application of the nonlinear dynamics, artificial neural networks and neuro-fuzzy expert systems, especially in neurological diseases [15] – [19].

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid adaptive neural network [12], [13] that is functional equivalent with Takagi-Sugeno-type system [20]. Unlike the fuzzy systems, ANFIS has the ability to adjust during a learning process. Applying an optimization method, the membership functions and the consequent parameters can be adjusted.

The minimized criterion function may be the mean square error between the current neuro-fuzzy system output and its proposed output. In order to explain the structure of the ANFIS it is considered a Takagi-Sugeno-type system of the first order [20] that has two input quantities, a and b, and the output c. The basis of the rules of the fuzzy system is considered to be formed from two rules:

Rule 1: If a is A_1 and b is B_1 then c is:

$$f_1 = p_1 \cdot a + q_1 \cdot b + r_1 \tag{1}$$

Rule 2: If a is A_2 and b is B_2 then c is:

$$f_2 = p_2 \cdot a + q_2 \cdot b + r_2 \tag{2}$$

The Takagi-Sugeno-type system of the first order is illustrated in Figure 1 [20] and the structure of ANFIS corresponding to the Takagi-Sugeno-type considered is presented in Figure 2 [20].

The equivalent ANFIS structure consists of 5 layers [5]. In this system is noted with O1,i the output node of the first layer of ANFIS structure.

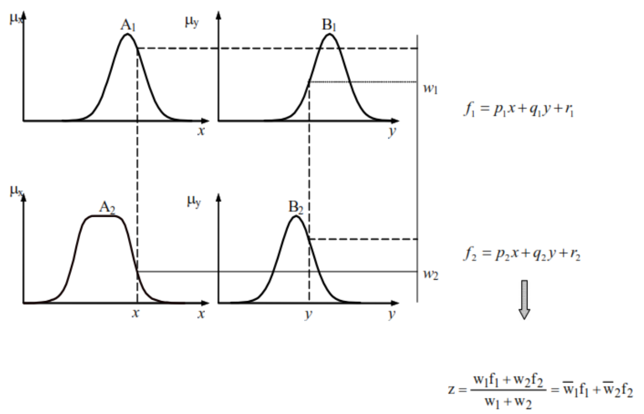

Fig. 1. Takagi-Sugeno-type system of the first order.

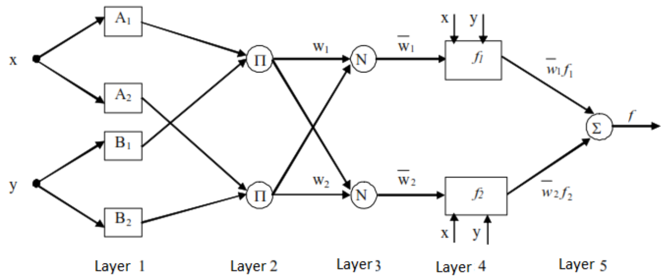

Fig. 2. Structure of the ANFIS neuro-fuzzy system corresponding to the Takagi-Sugeno-type system.

The process of learning in an ANFIS system is based on a hybrid structure. The five-layer structure of ANFIS [21] has the following functions presented below:

Layer 1: Every node is an adaptive node with a node function.

Layer 2: Every node is a fixed whose output is the product of all the incoming signals and represents the firing strength of a rule.

Layer 3: Every node is a fixed node. The ith node calculates the ratio of the ith rule's firing strength to the sum of all rules'firing stregths [21].

Layer 4: Every node is an adaptive node with a node function.

Layer 5: The single node in this layer is a fixed node, which computes the overall output as the summation of all incoming signals.

IV. Results

The generated structure was based on fuzzy clustering. We set the features vector, the number of rules and the way they interact, building the integrated priori knowledge for the proposed system.

It was chosen different data set: the training set - 80% of data, and the test set/validation test - 20% of data. In previous step of the training are computed the neurons output until the layer 4 and the linear function coefficients are calculated by least squares method RMS. In the next step, the error signals are back-propagated and the antecedent parameters of the membership function are updated using descending gradient method.

The base rules must be known a priori because the ANFIS structure can only adjust the membership functions of the adaptive and consequent parameters. The learning algorithm requires large computing resources and the modeling approach has difficulties in designing large models that involve a large number of partitions, rules and consequent parameters.

The proposed ANFIS structure is illustrated in Figure 3. The iterative algorithm is executed until a minimum amount of training error was reached. On the training data is performed a training procedure, which yields the rule base.

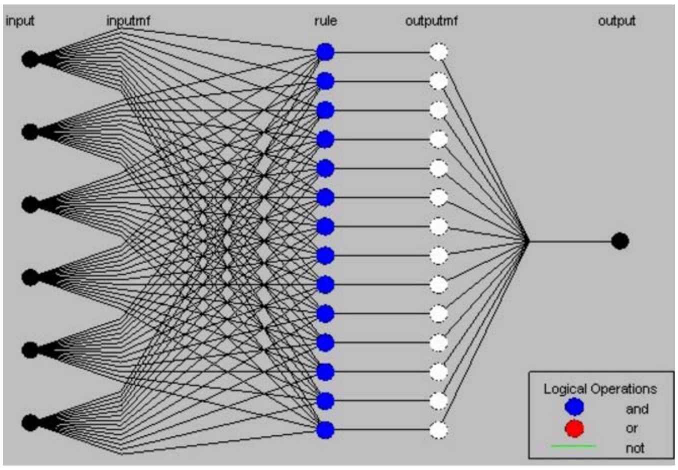

Fig. 3. Structure of the ANFIS neuro-fuzzy system.

The database of rules obtained using the described methods must be checked in terms of specifications that describe the quality: completeness, accuracy, consistency, complexity, and reproducibility.

The data set selected for the system testing includes six variables, commonly associate with the risk for diabetes:

- Plasma glucose concentration a 2 hours in an oral glucose tolerance test;
- Diastolic blood pressure (mm Hg);
- Triceps skin fold thickness (mm);
- 2-Hour serum insulin (mu U/ml);
- Body mass index (weight in kg/(height in m)^2);
- Diabetes pedigree function.

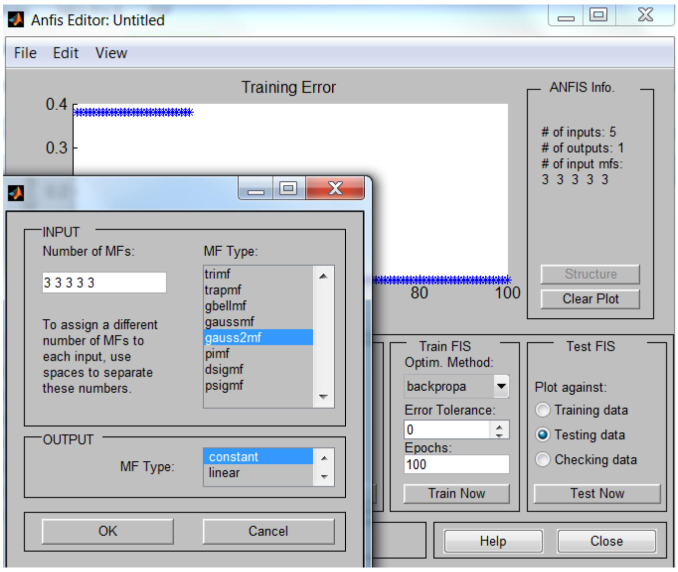

Fig. 4. Training Error and MF for ANFIS.

The neuro-fuzzy model testing on different set of data is necessary in order to validate the model obtained, in particular by the estimation error. The data set is used to evaluate the performance of the model. The ANFIS structure and algorithm training are implemented in MATLAB Toolbox.

Training		
RMSE	r	Correct
0.388246	1.146079	99.15%
0.321521	1.202728	99.18%
0.070078	0.515705	100.00%

Cross Validation		
RMSE	r	Correct
0.435473	1.219596	100.00%
0.319371	1.282056	100.00%
0.290909	1.971689	100.00%

Testing		
RMSE	r	Correct
0.574224	1.479728	100.00%
0.449625	1.425563	100.00%
0.304415	1.873252	100.00%

Fig. 5. Performance metrics I.

The quality of a classifier [22] in terms of correct identification of a class is measured using information from confusion matrix that contains:

- The number of data correctly classified as belonging to the class interests: *True positive cases* (TP);
- The number of data correctly classified as not belonging to the class of interest: *True negative cases* (TN);
- The number of data misclassified as belonging to the class of interest: *False positive cases* (FP);
- The number of data misclassified as not belonging to the class of interest: *False negative cases* (FN).

Based on these values the following measures are calculated: *Sensitivity, Specificity* and *Precision*.

	Training	Cross Val.	Testing
# of Rows	7520	1612	1612
RMSE	0.388246	0.435473	0.574224
Correlation (r)	1.146079	1.219596	1.479728
# Correct	7456	1612	1612
# Incorrect	64	0	0
% Correct	99.15%	100.00%	100.00%

Fig. 6. Performance metrics II.

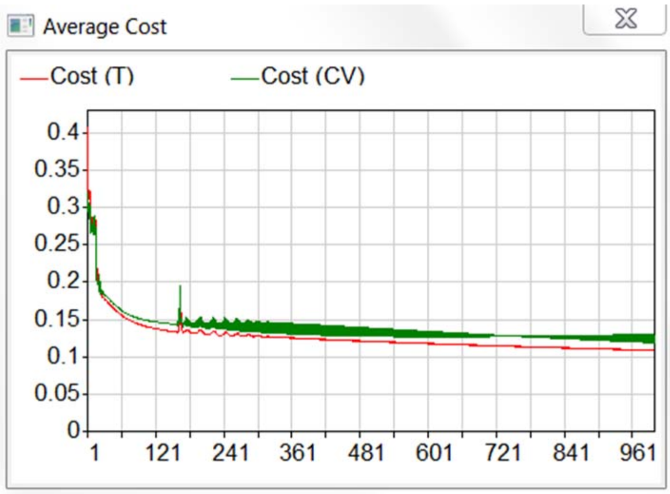

Fig. 7. Average Cost Function.

The backpropagation algorithm is a supervised learning algorithm based on minimizing the difference between the desired output and actual output by descending gradient method. The descending gradient method presents how the function varies in different directions.

The error is given by the difference between the desired output and the actual output of the network. In our case, in all three stages of ANFIS testing it was obtained Root Mean Square Error (RMSE) values below 0.5, with 85.35% for training data and 84.27% for testing data.

V. Conclusions

Worldwide, diabetes is one of the most common metabolic diseases. Diabetes management is a difficult process, based on the experience of the physician. One of the most important advantages of using the proposed system is the reduction of fuzzy rules, the impact on memory and the implementation of the proposed structure.

The ANFIS network performance consists precisely in the reduction of the number of initial rules, by reducing the feature vector and by introducing the DPF in features vector, as a criterion for initial screening. The results show that the proposed system, trained on a combination of algorithms has a good accuracy of the results classification and a minimum error.

References

[1] World Health Organization Reports and Facts (www.who.int), 2017.

[2] UCI Machine Learning Database, Department of Information and Computer Science, University of California, (www.archive.ics.uci.edu), 2017.

[3] J.W. Smith, J.E. Everhart, W.C. Dickson, W.C. Knowler, R.S. Johannes, “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, pp. 261-265.

[4] F. Amato, A. Lopez, E.M. Pena-Mendez, P. Vanhara, A. Hampl, J. Havel, “Artificial neural networks in medical diagnosis”, Journal Appl. Biomed., **11**, 2013, pp. 47–58.

[5] Q.K. Al-Shayea, “Artificial neural networks in medical diagnosis”, IJCSI International Journal of Computer Science, **8**(2), 2011, pp. 150-154.

[6] M. Catalogna, E. Cohen, S. Fishman, Z. Halpern, U. Nevo, E. Ben-Jacob, “Artificial neural networks based controller for glucose monitoring during clamp test”, PloS One, **7**, 2012.

[7] J. Fernandez de Canete, S. Gonzalez-Perez, J.C. Ramos-Diaz, “Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes”, Comput Meth Progr Biomed., **106**, 2012, pp. 55–66.

[8] M. Saghiri, K. Asgar, K. Boukani, M. Lotfi, H. Aghili, A. Delvarani, K. Karamifar, A. Saghiri, P. Mehrvarzfar, F. Garcia-Godoy, “A new approach for locating the minor apical foramen using an artificial neural network”, Int Endod J., **45**, 2012, pp. 257–265.

[9] G. Zhang, P. Yan, H. Zhao, X. Zhang, “A computer aided diagnosis system in mammography using artificial neural networks”, International Conference on BioMedical Engineering and Informatics, **2**, 2008, pp. 823-826.

[10] P. Dey, A. Lamb, S. Kumari, N. Marwaha, “Application of an artificial neural network in the prognosis of chronic myeloid leukemia”, Anal Quant Cytol Histol., **33**, 2012, pp. 335–339.

[11] A. Abraham, “Neuro fuzzy systems: state-of-the-art modelling techniques”, International Work-Conference on Artificial and Natural Neural Networks: Connectionist Models of Neurons, Learning Processes and Artificial Intelligence, 2001, pp. 269-276.

[12] Sikchi Smita, Sikchi Sushil, M.S. Ali, “Fuzzy expert systems (FES) for medical diagnosis”, International Journal of Computer Applications, **63**(11), 2013.

[13] S. Moein, S.A. Monadjemi, P. Moallem, “A novel fuzzy-neural based medical diagnosis system”, International Journal of Biological & Medical Sciences, **4**(3), 2009, pp. 146-150.

[14] S.A. Monadjemi, P. Moallem, “Automatic diagnosis of particular diseases using a fuzzy-neural approach”, International Review on Computers & Software, **3**(4), 2008, pp. 406-411.

[15] H.N. Teodorescu, M. Zbancioc, O. Voroneanu (Geman), Knowledge-based system applications, Performantica Publisher, Iasi, 2004, pp.10-293.

[16] O. Geman, “Nonlinear dynamics, artificial neural networks and neuro-fuzzy classifier for automatic assessing of tremor severity”, Proceedings of the E-Health and Bioengineering Conference, 2013, pp. 112-116.

[17] O. Geman, C.O. Turcu, A. Graur, “Parkinson's disease screening tools using a fuzzy expert system”, Advances in Electrical and Computer Engineering, **13**(1), 2013, pp. 41- 46.

[18] O. Geman, H.N. Costin, “Automatic assessing of tremor severity using nonlinear dynamics, artificial neural networks and neuro-fuzzy classifier”, Advances in Electrical and Computer Engineering, 14(1), 2014, pp. 133-138.

[19] O. Geman, “A fuzzy expert systems design for diagnosis of Parkinson's Disease”, Proceedings of the E-Health and Bioengineering Conference, 2011, pp. 122-126.

[20] T. Takagi, M. Sugeno, “Fuzzy Identification of Systems and Its Application to Modeling and Control”, IEEE Transactions on Systems Man Cybernetics, **15**, 1985, pp. 16-132.

[21] J.S.R. Jang, “ANFIS; Adaptive network-based fuzzy inference systems”, IEEE transactions on systems, man and cybernetics, **23**(3), 1993, pp. 665-685.

[22] O. Schipor, O. Geman, I. Chiuchisan, M. Covasa, From Fuzzy Expert System to Artificial Neural Network: Application to Assisted Speech Therapy, Artificial Neural Networks-Models and Applications, InTech Publisher, ISBN 978-953-51-2705-5, 2016, pp. 165-193.