

# ANFIS Based Classification Model for Heart Disease Prediction

Negar Ziasabounchi\*, Iman Askerzade

Department of Computer Engineering, Ankara University, Ankara, Turkey

Email: n.z.sabounchi@gmail.com, imasker@eng.ankara.edu.tr

**Abstract--** Heart disease diagnosis procedure is very vital and critical issue for the patient's health. Furthermore, it will help to decrease disease to a more specific level. The role of using machine learning techniques and data mining algorithms in diagnosis of heart disease is very considerable. The aim of this study was to develop a method of classifying for heart disease degree of patient based characteristic data using adaptive neuro fuzzy inference system. The data were obtained from the University of California at Irvine (UCI) machine learning repository. Seven variables are used as input of prediction model. To test the ability of the trained anfis models to recognize heart disease diagnosis, we used k-fold cross validation method. The experimental results demonstrate that the model successfully forecasts the patient's heart disease degree with an accuracy rate of 92.30%.

**Index Term--** ANFIS; Adaptive Neuro; Fuzzy Interface System; Classification; Heart Disease.

## 1. INTRODUCTION

As stated in [1], heart and blood vessel diseases called cardiovascular diseases contain several problems, many of which are related to atherosclerosis process which is a condition develops when a substance called plaque builds up in the walls the arteries. This process shrinks these arteries and cause problems for blood flowing through the vessels. Therefore, as a bad result, the potential risk of having heart diseases starts to increase. The World Health Organization report that world people die each year from heart diseases than from any other causes and it is estimated that more than 80% death from cardiovascular disease take place in low middle income countries and it is on the rise [2]. Hence, the main question comes that whether or not we can predict it before it comes true. As stated in [1, 3], with the advent of computer technology, the use of intelligent methods and algorithms (e.g. neural network, fuzzy logic and genetic algorithm) has started to play crucial role in complex and uncertain medical tasks such as diagnosis of diseases.

In last decade, the literature about the use of intelligent methods in medicine domain has seen enormous number of related studies [3, 4, 5, 6, 7]. As reported in [6], computer assisted applications and tools for diagnosis and treatment of patients seems to be more recent area of interest. Furthermore, the medical practitioners are also employing computerized technologies to assist in diagnosis and opinions as medical diagnosis is full of uncertainty [6]. On the other hand, fuzzy logic and neural networks stand as good methodologies dealing with these uncertainties [1, 6]. As noted in [1], in case of vague data or prior knowledge is involved, both of them serve certain advantages over classical methods. Nonetheless, as reported in [1], individual usage of these two methods can

cause some weaknesses. At this point, neuro-fuzzy integration presents a hybrid intelligent system that combines the power of human-like reasoning style of fuzzy logic with the connectionist structure of neural networks [8]. Moreover, Adaptive Neuro-Fuzzy Inference System (ANFIS) is one of the hybrid neuro-fuzzy inference expert systems and it works in Takagi-Sugeno-type fuzzy inference system, which was developed by Jyh-Shing and Roger Jang in 1993 [9]. As reported in [1], the ANFIS method provides a method for the fuzzy modeling procedure to learn information about a data set, in order to calculate membership function parameters which allow the associated fuzzy inference system to track the given input/output data.

When the literature is investigated, it can be easily seen that there exist diverse types of studies based on fuzzy and ANFIS methodologies [2, 10, 11, 12, 13, 14]. In [15], Malhotra and Malhotra introduced an ANFIS based smart predictive model for screening potential defaulters on consumer loans. Furthermore, in [16] Soyguder and Alli employed ANFIS method to build an expert system for the humidity and temperature control in HVAC systems. When the medical related ones are encountered, some of them can be listed as follow: In [11], Sungging et al. designed and developed an ANFIS based artificial intelligence system for lung cancer diagnosis. On the other hand, in [17] a fuzzy rule based expert system was implemented for asthma diagnosing. Likewise, Ucar et al. [18] carried out a study about tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets.

In this study we applied ANFIS method to Cleveland Clinic Foundation heart disease dataset which has been obtained from the well-known UCI machine learning data repository. The dataset consists of 303 subjects. The followed analysis methodology and the experiment results are given in the respective section. According to results, ANFIS method stands as a good and flexible predicting mechanism for heart disease prediction.

Remaining of this paper is organized as follows: The theoretical background of ANFIS is demonstrated in section 2. Section 3 presents the properties and structural information of dataset. The experiments and results of followed methodology is given in section 4 and whole study is summarized in section 5.

## 2. OVERVIEW OF ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

ANFIS is a network structure which represents Takagi-Sugeno-type fuzzy inference system. ANFIS computed initial membership functions by training itself with training data,

afterwards adjusting membership functions using either a back propagation algorithm or a hybrid-learning algorithm (a combination of back propagation and the least squares method) to minimize error measure. Actually, ANFIS benefits Artificial Neural Network's learning ability and Fuzzy-Logic decision making capability together. In [9], the learning rules of ANFIS have been described in detail.

### 2.1 ANFIS Architecture

In the following lines we express the five layers of ANFIS system.

#### Layer 1:

The first layer is fuzzification layer. Every node  $i$  in this layer is a square node. Which are given by:

$$O_i^1 = \mu_{A_i}(x), \quad \text{for } i=1,2, \quad (1)$$

$$O_i^1 = \mu_{B_{i-2}}(y), \quad \text{for } i=3,4, \quad (2)$$

where  $x$  and  $y$  are the inputs to node  $i$  and outputs are fuzzy membership grade of inputs [9]. In order to calculate the degree of membership of the input, every node uses Gaussian membership function.

$$O_i^1 = \mu_{A_i}(x) = e^{\frac{-1}{2} \left( \frac{x-c_i}{\sigma_i} \right)^2}, \quad (3)$$

where  $\{c, \sigma\}$  is a parameter set.  $C$  represents the membership function's center and  $\sigma$  determines the membership function's width. These parameters called premise parameters.

#### Layer 2:

Rule layer is the second layer. In this layer the input values are the membership functions and each node multiplies inputs and gives output which represents the firing strength of rule. The output of this layer given in equation

$$O_i^2 = w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2) \times \mu_{C_i}(x_3) \times \mu_{D_i}(x_4) \times \mu_{E_i}(x_5) \times \mu_{F_i}(x_6) \times \mu_{G_i}(x_7) \quad i=1,2,\dots,7, \quad (4)$$

#### Layer 3:

Here the  $i$ -th node is calculated by the ratio of the  $i$ -th rules firing strength to the sum of the rule's firing strengths.

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2,\dots,7, \quad (5)$$

#### Layer 4:

In this layer the nodes are adaptive nodes.

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i x_3 + a_i x_4 + b_i x_5 + c_i x_6 + d_i x_7 + t_i) \quad i=1,2,\dots,7, \quad (6)$$

where,  $(w)$  is the output of layer 3, and  $\{p_i, q_i, r_i, \dots, t_i\}$  is the parameter set which referred to as consequent parameters.

#### Layer 5:

This layer there is single fixed node. In this layer computes the overall output as the summation of the all incoming signals, given in equation

$$O_i^5 = y = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{w_i f_i} \quad i=1,2,\dots,7, \quad (7)$$

Fig. 1. shows the ANFIS architecture:

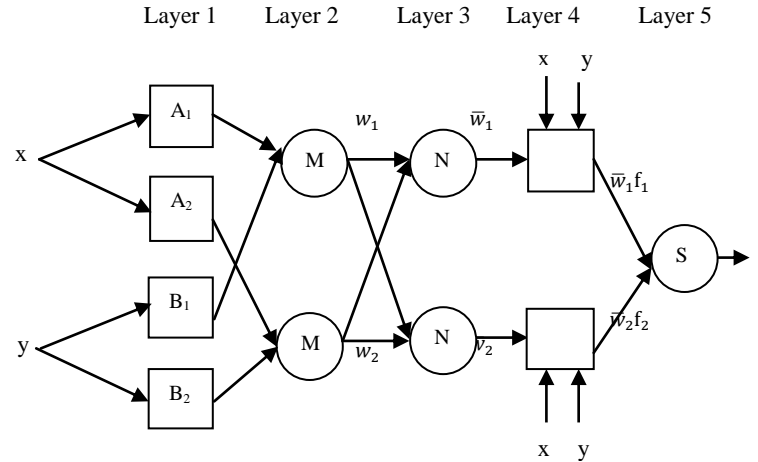


Fig. 1. ANFIS Architecture

### 2.1 Hybrid-learning Algorithm

There are two passes for hybrid algorithm, forward pass and backward pass. In the forward pass of the hybrid algorithm when the values of premise algorithm are fixed, the overall output can be expressed as a linear combination of the consequent parameters. Output  $y$  can be written as

$$\begin{aligned} f &= \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \\ &= \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2) \\ &= (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2 \end{aligned} \quad (8)$$

$$y = f(i, S), \quad (9)$$

where  $i$  is the vector of input variables,  $S$  is the set of parameters. If there is a function  $H$  such that the composite function  $H \circ f$  is linear in some of the elements of  $S$ , then these elements can be identified by the least-squares method [15]. If the parameter set  $S$  can be divided into two sets  $S_1$  and  $S_2$ , which  $S_1$  is set of premise parameters and  $S_2$  is set of consequent parameters we have

$$S = S_1 \oplus S_2, \quad (10)$$

(where  $\oplus$  represents the direct sum) such that  $H \circ f$  is linear in the elements of  $S_2$ , after applying  $H$  to Eq. (9), we have

$$H \circ O = H \circ f(i, S), \quad (11)$$

which is linear in the elements of  $S_2$ . After substituting  $p$  training data in (11) a matrix equation is obtained

$$A\theta = Y, \quad (12)$$

where  $\theta$  is an unknown parameter vector whose elements are parameters in  $S_2$ . Let  $S_2 = M$ , and dimensions of  $t$  is  $M \times I$  parameter vector,  $A$  is  $p \times M$  matrix and  $Y$  is  $p \times I$  output vector.

This is a standard linear least-squares problem, and the best solution for  $\theta$ , which minimizes  $\|A\theta - Y\|^2$  is the least-squares estimator (LSE)  $\theta^*$  [15].

$$\theta^* = (A^T A)^{-1} A^T y, \quad (13)$$

In the back propagation learning rule, consequent parameters are constant and a gradient model is applied to update premise parameters with output of network distribute through back. The following equation is used for updating premise parameters

$$\Delta x = -\eta \frac{\partial E}{\partial x}, \quad (14)$$

where  $x$  is premise parameters,  $\eta$  training rate,  $E$  is error value at out of the network.

### 3. DATASET

Table I  
Information about the input variables

| Variable Name | Min | Max | Number of MF | Description   |
|---------------|-----|-----|--------------|---|
| age           | 29  | 77  | 3            | Age in years  |
| cp            | 1   | 4   | 4            | Chest pain type(4 type)                                       |
| trestbps      | 94  | 200 | 3            | resting blood pressure  |
| chol          | 126 | 564 | 3            | cholesterol   |
| fbs           | 0   | 1   | 2            | resting blood sugar (0=false, 1=true) it is true when fbs>120 |
| restecg       | 0   | 2   | 3            | resting electrocardiographic(ECG)                             |
| talach        | 71  | 202 | 2            | maximum heart rate  |

## 4. EXPERIMENTS AND RESULTS

### 4.1 Data preparation

At the first stage, our crisp data which was collected from Cleveland heart disease dataset was converted in order to be used in MATLAB environment and preprocessed by normalization method. After that seven input variables are introduced to input of ANFIS with their membership functions. Gaussian membership function is used for fuzzy set description.

### 4.2 ANFIS Model Design

ANFIS or adaptive neuro fuzzy inference system is a class of adaptive networks that are functionally equivalent to fuzzy inference systems [9]. Our network has premise parameters and consequent parameters for the network structure illustrated in section 2. In this study, hybrid learning algorithm is used for our ANFIS model to identify parameters. There are two passes for hybrid algorithm, forward pass and backward pass. After presentation of premise parameters, in the forward pass a node outputs move ahead until layer 4 and the consequent parameters are calculated with least square estimate (LSE) then error measure is calculated for each node. In the backward pass, the error signals distribute backward to update premise parameters with gradient descent. In this

In this paper we used Cleveland heart disease dataset which is obtained from the UCI machine learning repository. This dataset contains 303 cases, 13 input fields and one output field which refers presence of heart disease in the patient.

We used 7 attributes for input and one attribute for output. Input fields are *age*, *chest pain*, *resting blood pressure*, *cholesterol*, *fasting blood sugar*, *resting electrocardiographic* and *maximum heart rate achieved* and output field refers to *presence of heart disease* which has 5 classes (*healthy and disease degrees {1, 2, 3, 4}*); increasing value show increasing heart disease risk.

In this dataset, fields are divided in some sections and each section has a value [3]. The settings of membership functions are kept as default values of MATLAB. The detailed information about variables is presented in Table I.

section, we used MATLAB software to develop ANFIS model. In the first stage, it is necessary to generate a fuzzy inference system for the heart diseases. To create an initial set of membership functions we used Grid Partition method. At the beginning of training, this method divides the data space in to rectangular sub-spaces using axis-paralleled partition based on predefined number of membership functions and their types in each dimension [19]. Seven variables were used with their membership functions. We used Gaussian membership function for each of the input variables. The number of these membership functions is shown in Table1. In the training process, consequent parameters are learnt during forward pass, when least squared error estimate approach is employed and premise parameters are learnt, when gradient descent method is applied during backward pass. After repeating forward and backward passes, premise and consequent parameters are determined. For our ANFIS model, training error tolerance set to 0.01. Three Gaussian membership function were used for blood pressure and cholesterol, The graph of these membership functions before training are shown in Fig. 2. and modified membership functions after designed ANFIS controller assume the form shown in Fig. 3.





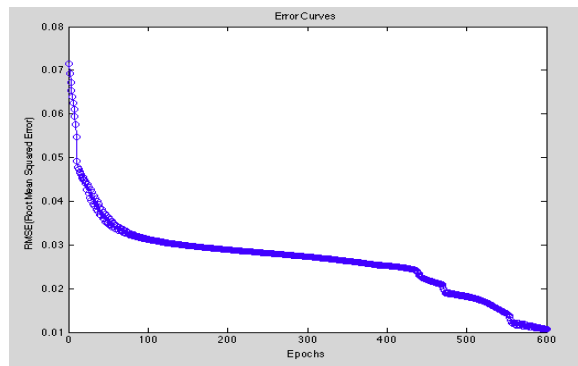


Fig.5.The diagram of error rate

Table II  
Error values for ANFIS

| ANFIS       |      |
|-------------|------|
| Train Error | 0.01 |
| Test Error  | 0.15 |

Surface viewer is a three dimensional curve that demonstrate the mapping with two input parameters to one output for obtaining heart disease degree. Fig. 5. and Fig. 6. show result surface after training.

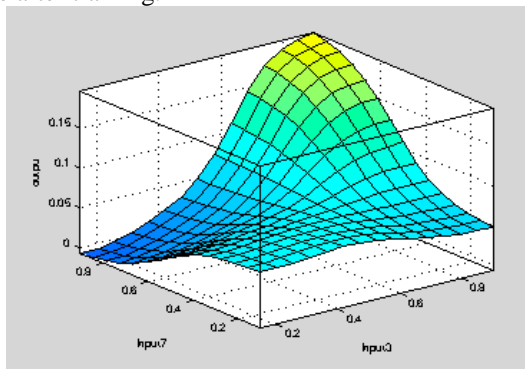


Fig.5. Surface view of max heart rate versus blood pressure

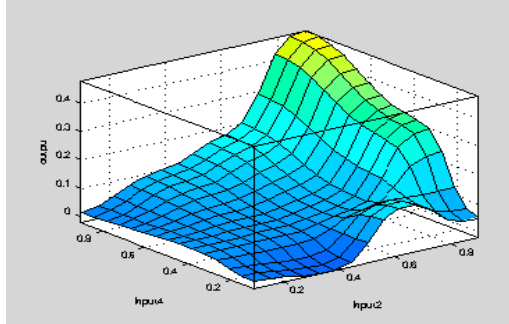


Fig.6. Surface view of cholesterol versus chest pain

To determine accurate diagnosis, the accuracy was obtained using this model and the result was compared with classify results of previous methods which used Cleveland heart disease dataset. The reported classification accuracy on this database [22] varies between 46.2% and 90% [20]. In this database researchers used varies data mining techniques in the diagnosis of heart disease to identify which data mining technique can provide more reliable accuracy. Table III illustrates a sample of these data mining techniques.

N.Jankowski used incremental net method in the detection of heart disease showing accuracy of 90% [22]. Kahramanli, H & Allahverdi, N used hybrid system to classify the data of heart disease to detect absence or presence of disease [20]. The proposed method showed the accuracy of 86.8% in the detection of heart disease. Comparison of these data mining techniques in the diagnosis of heart disease on the Cleveland heart disease dataset shows different accuracies. Hybridized data mining techniques are enhancing the accuracy of heart disease diagnosis [23]. In this research, neuro-fuzzy integration presents a hybrid intelligent system with the best accuracy in this database. As it is shown in Table III the proposed method in this study is the best classifier method in Cleveland heart disease dataset.

Table III  
Comparison of proposed method accuracy with other classifiers for the Cleveland heart disease

| Method                           | Accuracy % | Reference                            |
|----------------------------------|------------|--------------------------------------|
| Suggested approach in this study | 92.3       | This study                           |
| Inc Net                          | 90         | Nobert Jankowski                     |
| Hybrid system                    | 86.8       | Humar Kahramanli & Novruz Allahverdi |
| 28-NN, stand, Euclid, 7 features | 85.1± 0.5  | WD KG                                |
| LDA                              | 84.5       | Ster&Dobnikar                        |
| Fisher discriminate analysis     | 84.2       | Ster&Dobnikar                        |
| K=7, Euclid, std                 | 84.2± 6.6  | WD Gostminer                         |
| 16-NN, stand, Euclid             | 84± 0.6    | WD KG                                |
| FSM, 82.4-84% on test only       | 84.0       | Rafal Adamczak                       |
| K=1:10, Manhattan, std           | 83.8±5.3   | WD, GostMiner                        |
| Native Bayes                     | 82.5–83.4  | Rafal; ster, Dobnikar                |
| SNB                              | 83.1       | Ster&Dobnikar                        |

## 5. CONCLUSION

This paper suggests a technique for classification of heart diseases for helping patients to early predict and reliable diagnosis. Adaptive Neuro-Fuzzy Inference System has been used for classification which has both the advantages of neural network and fuzzy logic. In proposed model, training and average testing errors are 0.01 and 0.15 which are very satisfying. The experimental results show that the proposed

technique has high accuracy especially when compared with other studies that use the same database of heart disease. therefore this model is an appropriate model to classification of heart diseases. In the future an attempt will be made to develop this simulation with more input parameters for different databases.

#### REFERENCES

- [1] Opeyemi, O., Justice, E. O., Development of Neuro-fuzzy System for Early Prediction of Heart Attack, *Information Technology and Computer Science*, Vol:9, pp. 22-28, 2012.
- [2] World Health Organization, <http://www.who.org>, 9.2.2014.
- [3] Adeli, A. & Neshat, M. A Fuzzy Expert System for Heart Disease Diagnosis, *Proceeding of the International Multi Conference of Engineers and Computers Scientists*, 2010.
- [4] Allahverdi, N., Torun, S., Saritaş, İ. Design of a Fuzzy Expert System for Determination of Coronary Heart Disease Risk, *International Conference on Computer Systems and Technologies, CompSysTech'07*, 2007.
- [5] Nazmy, T.M., El-Messiry, H., Al-Bokhity, B., Classification of Cardiac Arrhythmia based on Hybrid System, *International Journal of Computer Applications*, 2(4), 2010.
- [6] Sikchi, S.S, Sikchi, S., Ali, M.S., Design of Fuzzy Expert System for Diagnosis of Cardiac Diseases, *International Journal of Medical Science and Public Health*, 2(1), 2012.
- [7] Kumar, A.V.S., Diagnosis of heart disease using Advanced Fuzzy Resolution Mechanism, *International Journal of Science and Applied Information Technology*, 2(2), pp. 22-30, 2013.
- [8] Neshat, M. and Yaghobi, M., Designing a Fuzzy Expert System of Diagnosing the Hepatitis B Intensity Rate and Comparing in with Adaptive Neural Network Fuzzy System, *Proceeding of the World Congress on Engineering and Computer Science*, 2009.
- [9] Jyh-Shing and Jang, R. ANFIS: Adaptive-Neuro-Based Fuzzy Inference System, *computer methods and programs in biomedicine*, IEEE Transactions on Systems, 1993.
- [10] Vipul, A.S., Adaptive Neuro-Fuzzy Inference System for Effect of Wall Capacitance in a Batch Reactor, *Advances in Fuzzy Mathematics*, Vol:4, pp. 69-75, 2009.
- [11] Sungging, H.W., Sylvia Ayu, P., Santoso, M.Y., Arifin, S., Application of Adaptive Neuro Fuzzy Inference System (ANFIS) for Lung Cancer Detection Software, *Nominator TICA Cluster II*, 2011.
- [12] Askerbeyli.I. & S. Abduljabar, J., Using fuzzy logic methods for carbon dioxide control in carbonated beverages, *International Journal of Electric & Computer Sciences*, No:03, pp. 114-116, 2011.
- [13] Askerbeyli.I. & Aybike, F., Neuro-Fuzzy approach for solving communication, *International Journal of Electric & Computer Sciences*, Vol.11, No: 01, pp. 37-43, 2011.
- [14] Askerbeyli.I. & Aybike, F., Effects of fuzzy logic methods over ATM networks, *International Journal of Research and Reviews in Applied Sciences*, Vol.9, pp. 295-305, 2011.
- [15] Malhotra, R. & Malhotra, D.K., Differentiating between good credits and bad credits using neuro-fuzzy systems, *European Journal of Operational Research*, Vol. 136, pp.190-211, 2001.
- [16] Soyguder, S. & Alli, H., An expert system for the humidity and temperature control in HVAC systems using ANFIS and optimization with Fuzzy Modeling Approach, *Energy and Buildings*, Vol. 41, pp. 814-822, 2009.
- [17] Zarandi, M.H.F, Zolnoori, M. M., Heidarnejad, H., A fuzzy-rule based expert system for diagnosing asthma, *Transaction E: Industrial Engineering*, 17(2), pp. 129-142, 2010.
- [18] Ucar, T., Karahoca, A., Karahoca, D., Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets, *Neural Comput&Applic*, 2012.
- [19] Wei, M., Bai, B., Sung, A., Liu, Q., Wang, J. & Cather, M. Predicting injection profiles using ANFIS, *Information Sciences Journal*, 2007.
- [20] Kahramanli, H. & Allahverdi, N. Design of a hybrid system for the diabetes and heart disease, *Expert systems with applications*, pp. 82-89, 2008.
- [21] Patil, B.M., Joshi, R.C. & Toshnival, D., Hybrid prediction model for type-2 diabetic patients, *Expert systems with applications*, pp. 8102- 8108, 2010.
- [22] <http://www.fizika.umk.pl/projects/datasets.html>
- [23] Shouman, M. Tuner, T. & Stocker, R. Using data mining techniques in heart disease diagnosis and treatment, *International Conference on Communications and Computers*, pp.173-177, 2012.