Introduction of a Method to Diabetes Diagnosis According to Optimum Rules in Fuzzy Systems based on Combination of Data Mining Algorithm (D-T), Evolutionary Algorithms (ACO) and Artificial Neural Networks (NN)

Mohammad Fiuzy [1, 2], Azam Qarehkhani[3], Javad Haddadnia [1,2,*], Javad vahidi[4], Hadi Varharam [5]

[1] Biomedical Engineering Department, Electrical and Computer Faculty, Hakim Sabzevari University, Sabzevar, Iran.

[2] Research Center for Advanced Medical Technologies, Sabzevar University of Medical Sciences, Sabzevar, Iran.

[3] Department of Computer, Khorasan Razavi, Neyshabur,Science and Research Branch, Islamic Azad University, Neyshabur, Iran.

[4] Department of Applied Mathematics, Iran University of Science and Technology, Behshahr, Iran.

[5] Ministry of Science, Research and Technology, Tehran, Iran.

Mohammad.fiuzy@yahoo.com, az_qarehkhani@yahoo.com, Haddadnia@sttu.ac.ir, Jvahidi@iust.ac.ir, hadi_varahram@yahoo.com

Abstract

In time diagnosis of diabetes significantly reduces damages and inconveniences of this disease in society. It may be said that one of the most important problems of diagnosis methods of this disease, particularly in early phases, is not to pay attention to proper features in order to diagnose the disease and as a result weakness in disease diagnosis. This research endeavors to introduce a new method for accurate diagnosis of this disease through usage of a combination of artificial intelligent methods such as fuzzy systems for immediate and accurate decision making, Evolutionary Algorithms (ACO[1]) for choosing best rules in fuzzy systems, and artificial neural networks for modeling, structure identification, and parameter identification. The proposed system relying on features of database in the form of combination and interaction succeeded in reaching an accuracy of 95.852% which in comparison to current methods on the one hand and to artificial methods in foresaid references on the other hand, has a proper and very faster performance than other intelligent methods and you can see its accuracy and excellence as an intelligent system.

Keywords: Diabetes, Diagnosis, Optimal Rule, Fuzzy Systems, Data Mining, Artificial Processing

---

[1] Ant Colony Optimization

1. Introduction

Diabetes or Diabetes Mellitus is one of the most prevalent diseases in the world. Reports declare that there are approximately 220 million diabetic patients in the world [1, 2, and 3]. Diabetes is diagnosed primarily through blood glucose tests (Plasma, Glucose, and Fasting). The disease is classified into 3 categories. Type 1 diabetes, or Insulin-Dependent Diabetes Mellitus (IDDM), is mostly diagnosed in young age and children. Type 2 Diabetes, or Non-Insulin-Dependent Diabetes Mellitus (NIDDM) is seen in 90% to 95% of diabetes patients [3]. In this category insulin is secreted, but body resists consuming it. Type 3 Diabetes is mostly seen in pregnant women and usually is changed into type 2 Diabetes after pregnancy [4]. In recent few years a significant attention is given to medical data mining methods. It is proven in theory that a group of classifiers yield a better result than the best of them [5-7]. It is understandable that every group of data and methods has its own weaknesses and capabilities and that it may act well in a data mining system and doesn't yield a desirable result somewhere else [8]. Therefore, there is an increasing tendency to combine intelligent methods in order to reach a long lasting and desirable answer. Up to now many methods have been studied and introduced in the form of many researches each of which has been done at best through using RBF[1] Neural Network with accuracy of 73.32% [9], MLP[2] Neural Networks with accuracy of 76.89% [10], K-NN[3] with accuracy of 75.55% [11], Evolutionary Algorithms with accuracy of 76.17% [3], and SVM[4] Algorithm with accuracy of 82.5% [3]. In this paper it is endeavored first to explain the problem and seriousness of the issue concerning diabetes mellitus, then the method is explained generally, and finally main points are discussed considering Flow chart (1). We hope this research will be helpful in improving methods, treatment, and diagnosis of this disease.

2. Application Steps in the Proposed System

First data is extracted from the database [12]. Then, it is explored through process of decision tree so that a space for research and a path to reach from input space (patient's features) to output space (relative rate of being sick or healthy) is provided through symbolic drawing of a tree. Considering drawing paths of the tree, somelogical provisions for reaching from beginning of the tree to its end (percentage of individual's being sick or healthy) are obtained. These provisions can be used as sources of Bank of Fuzzy System Rules. Afterwards, using ACO Exploration (Evolutionary) Algorithm, best possible Rules (provisions) are obtained by using Rules of decision tree. Then, Fuzzy System, considering the bank of optimum rules of inputs (features of each person), determines the individual's rate of being sick or healthy so that identification and final modeling of the patient is done through input of the data into artificial nervous system. Flow chart (1) shows all steps followed in this research.
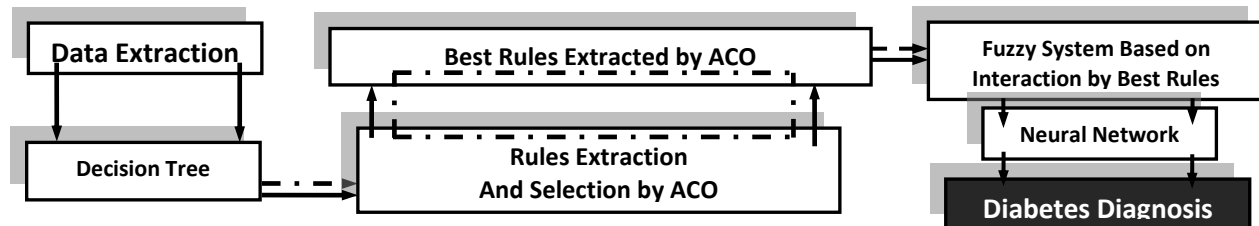
Flow chart (1): Application process of the Algorithm proposed in this research in order to diagnose diabetes

---

[1] Radial Basis Function

[2] Multy Layer Perceptron

[3] K – Nearest Neighbor

[4] Support Vector Machine

3. Importance of Feature Selection or the Issue of Pattern Recognition

First the issue of selecting proper features in transform space must be dealt with. Feature selection is actually selection of those features which have the maximum power in foreseeing the output [13]. Defining what the optimum subset is, depends on problems you are going to solve [14]. In discussed methods, amounts of vectors show their importance. According to this, paying attention to all features is not a proper thing to do. In addition, sometimes feature vectors don't contain useful information. As proven in [15], it is expected to get better results when you have more information; however, in practice this causes a reduction in accuracy [16]. Therefore, to reach a proper representation, it is essential to select the best feature vector or simplify it. A current Pattern Recognition System Consists of 4 parts: Feature Extraction, Feature Selection, Design and Instruction of Classifier, and Finally Testing. In this research Decision-Making Tree is used for Exploring and Extracting Features (Rules), Ants Colony Algorithm is used for selecting Effective rules, Fuzzy system is used for Simplifying Features, and Artificial Neural Network is used for Classifying or Testing the Data of [12] for diagnosis.

4. Database

Data used in this research was taken from PID[1] database belonging to references of [12] which contained 8 Features of 768 Women whom were at least 21 years old. From among these 500 women were healthy and 268 suffered diabetes. The 8 features, According to WHO[2] , are these: A. Number of Pregnancy, B. Concentration of Glucose in Plasma During 2 hours, C. Blood Pressure (Hg mm), D. Skin Fold Thickness of Triceps (mm), E. 2-h Serum Insulin (mu U/m1), F. Body Mass Index (kg/m2), G. Diabetes Pedigree, H. Age.

5. Decision Tree

Decision trees were first popularized in 1986 by Quinlan with ID3[3] Algorithm which uses Entropy Standard to evaluate differentiation capacity in each Feature of data [17]. The most important characteristic of decision trees is their ability in dividing a complex problem into smaller problems and as a result yielding an understandable solution [17, 18]. The closer the numbers of members of the two categories and the more diverse the data, the closer the entropy standard to 1.The Less the number of members of one category compared to that of the other, the closer the entropy standard to 0 [18]. Figure (1) shows changes in entropy according to probability of data of one category for a two-category problem.
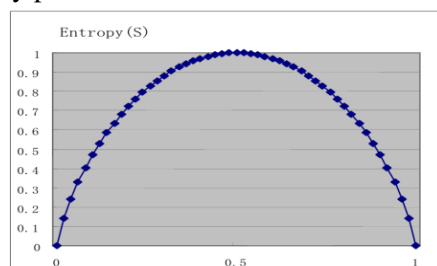

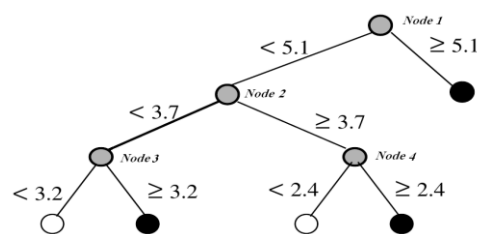| Figure 1: Changes in entropy according to probability of occurrence [16, 18] | Figure 2: Schema of a sample CART tree [16, 18] |

---

[1] Pima Indian Diabet database

[2] World Health Organization

[3] Iterative Dichotomies 3

CART[1] Tree (the tree used in this research) was introduced by Breiman in 1984. Primary idea of this method: in each step, from among a combination of all possible modes to divide data into two parts, the mode which results in yielding purer data is chosen [18]. Suitability of division is dependent on just one former step and optimum division is not examined according to former observances. If discretely-displayed input data is shown by L, then 2^{L-1}-1 modes will be available for selection of categories. If data is uninterrupted and there are K categories of different data, then there will be K-1 possibilities for dividing input data. Calculation of purity of data is done through Criterion Tuning Standard. Figure (2) shows a sample tree with Four Nodes [18]. In the End, it can be said that decision tree is used in order to form a bank of accurate and optimum rules (considering inputs (features)).

6. Ant Colony Optimization Algorithm (ACO)

Ant ColonyAlgorithm was first introduced by "Dorigo" and his Colleagues in 1991 to solve complicated optimization problems such as Travelling Salesman Problem (TSP[2]). Then in 1996 and 1997 Dorigo and his colleagues introduced Ants Society Algorithm [19, 20]. Ants are small Insects that live in groups and Societies. These insects can't see, so they find their way from their nest to food by secreting a chemical substance named Pheromone. In real world Ants first go accidentally this way and that way to find food. Then they return to their nest leaving a trace of Pheromone on the ground. These Traces are changed into white and become visible after rain.When other ants find these traces, sometimes they stop wandering and follow the traces. Then if they reach the food, they return to their nest and leave a new trace of their own beside the old one; in other words they strengthen the old trace. It can be concluded that secretion of the chemical substance and its density are important factors in finding the shortest routes by ants. In other words, according to density of the chemical substance on the routes, ants recognize amount of traffic on the routes and take the busiest [21]. Proper Pattern to solve optimum problems is obtained on the basis of ants' behavior in finding the shortest route - figure (3).
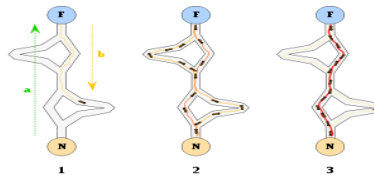
Figure 3: Navigation of ants through Ants Colony Algorithm

First Ants must be put in N cities. Naturally, the most proper method in the first place is accidental placement of Ants. Equal amounts of primary pheromone, whose proper amount will be discussed later, are attributed to all routes. Then, ants move on the routes. Probability of the Ants going from Node i to node j during time period of t is calculated through Equation (1).

Equation (1)

$$\tau_{ij}(t+1) = (1-\rho)\tau(ij) + \sum_{\substack{K \in Colony \\ that \\ used \\ edge(i,j)}} \frac{Q}{L_K} \qquad P_{ij} = \frac{\tau_{ij}{}^\alpha (\frac{1}{d_{ij}})^\beta}{\sum_{j \in allowednodes} \tau_{ij}{}^\alpha \left(\frac{1}{d_{ij}}\right)^\beta}$$

In this equation τ_{ij} is amount of pheromone on the routes, d_{ij} is the distance between the two nodes. α, β Are amounts which are determined at the beginning and remain the same up to the end.When ants have finished a complete tour, comprehensive updating of pheromone is done through Equation (2) by the ant which has taken the best route. α Is a parameter between zero and one, ρ is evaporation

---

[1] Classification And Regression Tree

[2] Travel Salesmen Person

coefficient, Q is a constant, and L_K , which is attributed in the beginning of traition and solving, is the length of the route [22]. Repetition of Ants Algorithm is stopped when convergence condition is fulfilled through 2 rules of pheromone density of the route and number of route repetition. In this way Ants Algorithm results in selection of best rules. First, a series of numbers is created, then considering the application process of the Algorithm, created numbers, place of rule storage, and satisfaction of fitness function the best and of course smallest numbers of rules which result in most accurate diagnosis of the disease are defined for fuzzy system with usage of bank of rules.

7. Fuzzy System

Fuzzy logic is a fairly new technology compared to current methods of design and modeling of systems which need fairly advanced probabilities and mathematics. Fuzzy logic defines correct propositions of Boolean accurately. Membership function of a fuzzy system, in special conditions, is a curve which shows mapping of each spatial point to membership degree (a number between 0 and 1) [23, 24]. Each fuzzy system has rules in bank of rules in the form of phrases IF_THEN. In this research we have used decisions of the tree to obtain logical and correct rules. Processing step which is called decision making step acts according to bank of rules. There are 2 types of Mamdani and Sugeno fuzzy inference systems which are different from each other in their methods of output definition [25]. In this system we have used Mamdani inference. Figure (4) shows the designed Fuzzy System with 8 inputs (people's features) and 2 outputs (Fuzzy expression of rate of People's sickness or health).
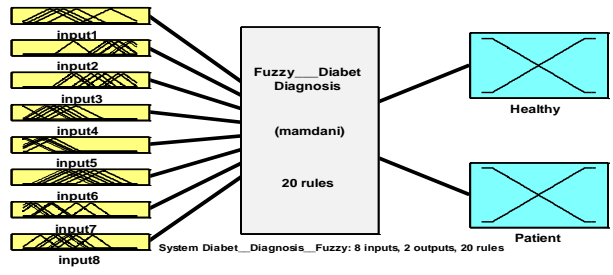
Figure (4): Designed fuzzy system with 8 inputs (features), 2 outputs (rate of fuzziness of people's sickness and health), and 7 rules

7.1 Input and Output Parameters of Fuzzy System

Input parameters (for each person) in this system are the same 8 features of the database [12]. Membership functions and their intervals should be determined for each input; number of membership functions is determined by their ranges and ranges of membership functions are determined by decisions of the Tree. Type of membership functions is determined through consideration of domain and range of membership functions. For example, membership functions (in input parameters) which have proper domains are of "triangular" type (like figure (5) for second (input) feature and equation (3)) and functions which have very small domains are of "trapezius" type (like figure (6) for the first output and equation (4)).
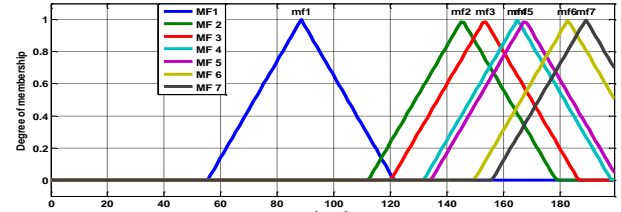
Figure5: Membership functions for second (input) feature
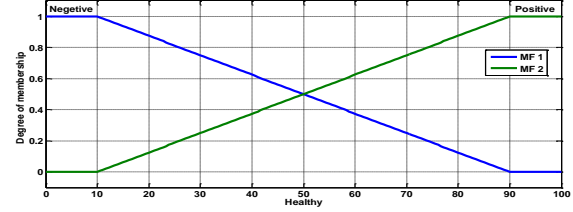
Figure6: Membership functions for first (output) feature

$$f(x,a,b,c) = \max\left(\min(\tfrac{x-a}{b-a}, \tfrac{c-x}{c-b}), 0\right); \quad (3) \quad f(x,a,b,c,d) = \max\left(\min(\tfrac{x-a}{b-a}, 1, \tfrac{d-x}{d-c}), 0\right); \quad (4)$$

Equations (3 and 4): Equations of triangular membership functions (3) in input parameters (figure 5) and trapezius membership functions in output parameters (figure 6).

Output parameters show membership rate of each person (in fuzzy expression) in 2 healthy and sick output parameters which were showed by 2 trapezius membership functions. The most important part of fuzzy system is bank of rules. In bank of rules, considering input, output, membership functions, and inference system, correct decisions are made to determine Crisp output. All parameters of the proposed fuzzy system are shown in table 1.

Table 1: Characteristics of the designed fuzzy system

Implication	Number of Membership Functions	Number of parameters	And Method	Type of membership functions	Defuzzification	Aggregation	Or Method
Mamdani	Variable	8in / 2out	min	Trap&Gbell	Centerior	Max	Min

7.2. Decision Tree, Bank of Rules, and Physical Surface

Decision Trees can present an understandable description of the relations existing in a data set and can be used for categorization and forecasting [26]. Decision structure can also be presented as computational techniques which help describe, categorize, and General Making of a set of data [27]. Flow chart (2) shows the relation between the decision tree and the fuzzy system in the proposed method.
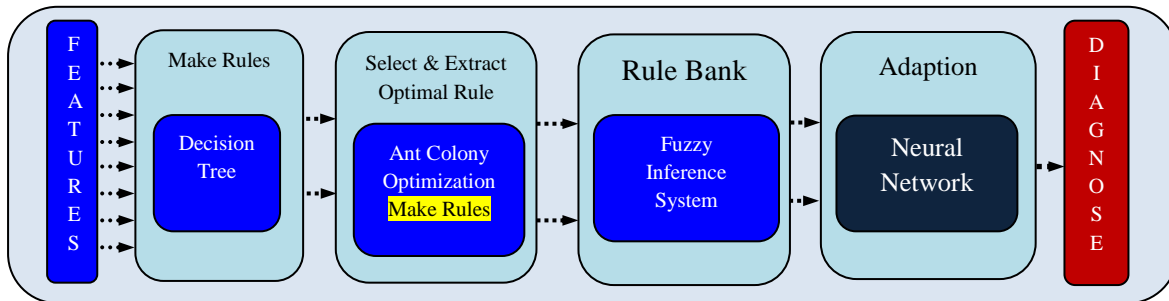
Flow chart 2: Application process of the proposed Algorithm in structure, selection, and usage of best rules for diagnosis of diabetes

By applying the Decision Tree to the database some relations and rules are obtained (like figure (7)) and finally by using these rules (table (2)) inputs of next step are fed i.e. with rules that are extracted in this step, by using Ants Colony Optimization Algorithmbest rules are obtained provided that the proposed fitness function is satisfied. A sample of decisions and rules obtained from the decision tree is shown in figure (7) and table (2).

The figure shows a decision tree branch with the following rules:

$$
\left.\begin{array}{l}
\text{if } X2 >= 143.5 \text{ and } X2 <= 154.5 \text{ Then } Patirnt = 0.5 \\[2pt]
\text{if } X2 >= 143.5 \text{ and } X2 > 154.5 \text{ and } X6 < 29.85 \text{ Then } Patient = 0.52381 \\[2pt]
\text{if } X2 >= 143.5 \text{ and } X2 > 154.5 \text{ and } X6 >= 29.85 \text{ and } X5 >= 544 \text{ Then } Patient = 0.4 \\[2pt]
\text{if } X2 >= 143.5 \text{ and } X2 > 154.5 \text{ and } X6 >= 29.85 \text{ and } X5 < 544 \\
\text{and } X7 < 0.3005 \text{ Then } Patient = 0.75 \\[2pt]
\text{if } X2 >= 143.5 \text{ and } X2 > 154.5 \text{ and } X6 >= 29.85 \text{ and } X5 < 544 \\
\text{and } X7 >= 0.3005 \text{ and } X4 >= 44 \text{ Then } Patient = 0.78937 \\[2pt]
\text{if } X2 >= 143.5 \text{ and } X2 > 154.5 \text{ and } X6 >= 29.85 \text{ and } X5 < 544 \\
\text{and } X7 >= 0.3005 \text{ and } X4 < 44 \text{ Then } Patient = 1
\end{array}\right\}
$$

Figure 7: A branch of the first tree (inputs and outputs of fuzzy expression of disease)

After creating the rules by the decision tree (application of the tree to input space (people's feature) and to output space (patients and healthy people)), some rules are obtained, but the main problem is existence of lots of rules. If the rules are applied to fuzzy system on this basis, it will show itself up in speed and operation of fuzzy system in the form of slowness and low accuracy. That is why Ants Colony Optimization Algorithm is used. First, a Binary String which accidentally contains 20 bits of number of rules is taken in order to choose at most 20 better rules, this process is taken to satisfy fitness function. Then, this process is repeated so many times that at most 20 better rules are chosen.

10	125	1	14	118	98	25	78	60	24	36	91	18	12	148	45	24	78	75	20

For example, some rules of this fuzzy system which has more than 248 rules are shown in table (2).

Table 2: Some of the rules extracted from the branch of the tree in figure (7)

Some Of first Rules in Fuzzy System	
1 1 0 0 0 1 0 1, 3 0 (1) : 1	1 1 0 0 0 1 0 1, 0 1 (1) : 1
2 1 0 0 0 1 0 1, 1 0 (1) : 1	2 1 0 0 0 1 0 1, 0 3 (1) : 1
5 1 1 0 0 2 1 1, 3 0 (1) : 1	5 1 1 0 0 2 1 1, 0 1 (1) : 1
0 1 2 0 0 0 1 1, 3 0 (1) : 1	0 1 2 0 0 0 1 1, 0 1 (1) : 1
0 1 0 0 0 2 2 1, 3 0 (1) : 1	0 1 0 0 0 2 2 1, 0 1 (1) : 1
0 1 0 0 0 3 0 1, 1 0 (1) : 1	0 1 0 0 0 3 0 1, 0 3 (1) : 1
0 1 0 0 0 4 0 2, 3 0 (1) : 1	0 1 0 0 0 4 0 2, 0 1 (1) : 1
0 2 0 0 0 5 0 2, 1 0 (1) : 1	0 2 0 0 0 5 0 2, 0 3 (1) : 1
0 3 0 0 0 5 0 2, 3 0 (1) : 1	0 3 0 0 0 5 0 2, 0 1 (1) : 1
0 4 0 0 0 5 3 2, 3 0 (1) : 1	0 4 0 0 0 5 3 2, 0 1 (1) : 1
3 4 0 0 0 5 4 2, 1 0 (1) : 1	3 4 0 0 0 5 4 2, 0 3 (1) : 1
0 4 0 0 0 5 7 2, 1 0 (1) : 1	3 4 0 0 0 5 4 2, 0 2 (1) : 1
0 5 0 0 0 7 0 0, 3 0 (1) : 1	0 4 0 0 0 5 7 2, 0 3 (1) : 1
0 6 0 0 0 7 0 0, 2 0 (1) : 1	0 5 0 0 0 7 0 0, 0 1 (1) : 1
0 7 0 0 0 6 0 5, 4 0 (1) : 1	0 6 0 0 0 7 0 0, 0 4 (1) : 1
0 7 0 0 0 6 5 6, 2 0 (1) : 1	0 7 0 0 0 6 0 5, 0 2 (1) : 1
0 7 0 0 0 6 8 6, 1 0 (1) : 1	0 7 0 0 0 6 5 6, 0 4 (1) : 1
0 8 0 0 1 6 0 0, 1 0 (1) : 1	0 7 0 0 0 6 8 6, 0 3 (1) : 1
0 8 0 0 1 6 6 3, 1 0 (1) : 1	0 8 0 0 1 6 0 0, 0 3 (1) : 1
0 8 0 0 2 6 0 0, 3 0 (1) : 1	0 8 0 0 1 6 6 3, 0 3 (1) : 1
0 8 0 0 2 6 0 0, 0 1 (1) : 1	0 8 0 0 2 6 0 0, 0 1 (1) : 1

Order of numbers in each rule is as follows: consider the first law, 1 1 0 0 0 1 0 1, 3 0 (1) : 1, this rule is interpreted in this way, the first part from the left and before the comma i.e. digits that are highlighted with yellow color, shows the membership function in each input, in other words, antecedent parts of each rule or if. So, in this rule it must be said that if in the first input, the first membership function exists and in the second input, the first membership function exists and in the third, fourth, and fifth inputs no membership functions exist and in the sixth input the first membership function exists and in the seventh input no membership function exists and in the eighth input the first membership function

exists (up to now, antecedent part of each rule is expressed); after that is the place of consequent or thered part. In consequent part we must have the same number of rules as the number of outputs. So, this part is as follows: then the first output is the third membership function and the second output doesn't have a membership function. The figure in parentheses is the weight of this rule or the blue-color figure. The number after: is either the figure of parentheses 1 or the number of connections of this rule with other rules. The first rule is connected to one rule (green color).Therefore rules by the order show rule (a, 1: end,'indexed') and of course the interpretation of this rule is in this way that: a,the storage of the designed fuzzy system, expresses rules of 1 to the end with the Index structure. In each attempt from among 248 rules, an accidental array full of an accidental number between 1 to 248, which shows the number of line of rules, is created accidentally byAnts Colony Optimization Algorithm. In each attempt of Algorithm, 20 ants propose 20 search paths which are the same 20 accidental rules. Of course this process is repeated so many times that the best rules are obtained. Cost function of Ants Algorithm is taken as flow chart (3). The above vector is an accidental sample of the vectors that are studied.

Objective Function which is taken in this research is:

$$Fithness = \alpha.Accuracy \ + \ \beta.\frac{|n|-|s|}{|n|}$$

7.3. Objective Function

The objective function is the equation that gets merits for each Iteration. It seems that accuracy in identifying is more important than the small selected subsets. Although the two series have the same accuracy, the smaller set is preferred, so we suggest the following fitness function in Equation (5).

Equation (5)
$$Fithness = \alpha.Accuracy \ + \ \beta.\frac{|n|-|s|}{|n|}$$

Which |n| is the number of the total rules and |s| is the number of selective rules. First sentence is the carefully identified coefficient and the second sentence is the rate coefficient of reduced features. Accuracy is defined by the following equation or Equation (6).

Equation (6)
$$Accuracy \ = \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$$

Tp = True Positive = Cytological and suspicious positive diagnosis, which are positive in pathological tests.

Fp = False Positive = Cytological and suspicious positive diagnosis, which are negative in pathological tests.

Tn = True negative = Cytological and suspicious negative diagnosis, which are negative in pathological tests.

Tp = True positive = Cytological and suspicious positive diagnosis, which are positive in pathological tests.

We supposed that the sum of α and β coefficient is fixed and equal to 100. Now, in regard to importance of the identified Accuracy and fewer features used in the diagnosis, α and β coefficients are set. Undoubtedly, detection Accuracy is more important in this problem, and therefore, α (here 99) will be greater than β (here 1) [26].The suggested Fitness Function located in the ACO Algorithm cost function is observed in Flowchart (3).
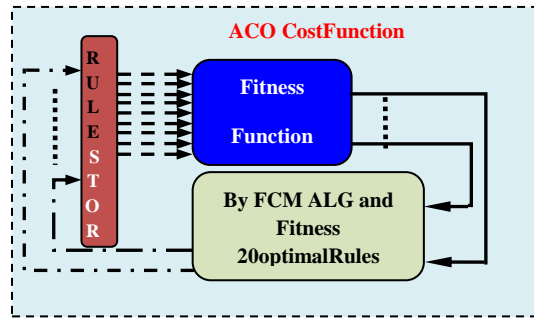
Flow chart (3): ACO Cost Function

However, to determine the parameters in Accuracy we use FCM Algorithm that is described below. We use FCM, because the FCM Algorithm will classify Without Unsupervised Algorithm and Target Vector. The FCM tries to make the data classification, also in the second class. Data from the fuzzy system must be classified into two classes to determine the Accuracy of Fuzzy System. Firstly we are trying to introduce the FCM Algorithms, because to obtain the rate of accuracy and classification of the data, we need it.

7.4. FCM or Fuzzy C-Means Algorithm

In 1969, "Baraldi", proposed the first clustering model with fuzzy idea [30]. In this method, the amount of any membership or data belonging to each data to any cluster in matrix member is determined.

$$U = \left[u_{i,j} \right]_{c \times n} = (\vec{u}_1, \vec{u}_2, ..., \vec{u}_n)$$ Equation (7)

Which, c is number of clusters and n is the data number. This method suffers from two main limitations: the first limitation was that no cluster should be empty.

$$(\sum_{j=1}^{u} u_{ij} > 0 \quad \forall_i \in \{1, ..., c\})$$ Equation (8)

The second limitation, called normalization constraints, was that the total membership of all clusters must be equal to "1" in each data class.

$$(\sum_{j=1}^{u} u_{ij} = 1 \quad \forall j \in \{1, ..., n\})$$ Equation (9)

FCM tries for any data set, finds the parts that minimize the following cost Equation 1 or objective function Equation 10. $J_f(X, U_f, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2$ Equation (10)

Where in it, d_{ij} is data distance between X_j and the ith cluster center. $m \in [1, \infty)$ is degree of fuzziness. If "M" goes to one, clustering will be more difficult or crisp. On the contrary, if "M" goes to infinity, clustering will be fuzzier. Supposedly as demonstrated in Figure 4 we supposedly classified 1000 random data in 2 classes.

Fuzzy –C- Means or "FCM" methods consist of 4 stages:

1. If function cannot be minimized directly, repeat Algorithm could be used. To solve this problem, the optimal replacement scheme was used as follows: select the proper values for m, C and small positive number for "ε". The matrix C is randomly filled (middle or center of clusters) finally, set t =0.

2. In (t = 0) membership matrix is calculated, and in (t > 0) updated membership matrix is determined. This means that the degree of membership for fixed parameters of clusters is optimized such as Equation 11 in the flowing:

$$u_{ij}^{(t+1)} = \frac{d_{ij}^{-2/(m-1)}}{\sum_{i=1}^{c} d_{ij}^{-2/(m-1)}} = \frac{1}{\sum_{l=1}^{c} (\frac{d_{lj}}{dij})^{1/(1-m)}} \quad for \quad i = 1,...c \quad and \quad j = 1,.....N$$

Equation (11)

3. The final step is updating the center of the clusters with optimized membership matrix. In addition to these parameters, how the distance is measured is of prime importance.

4. Repeating Steps 2 and 3 until $\left\|C^{(t+1)} - C^{(t)}\right\| < \varepsilon$; or $\left\|U^{(t+1)} - U^{(t)}\right\| < \varepsilon$ by all above steps are applied on the 100 sample data that Outside From Fuzzy System the First (Diabetic) cluster and second (Health) cluster is created. The result can be observed in Figure 8.
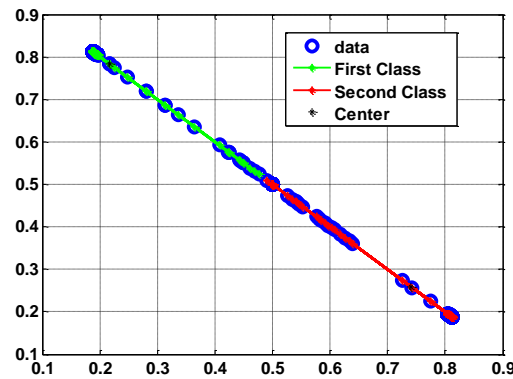

Figure (8): FCM Application on data that extracted of Fuzzy System

This shows sum of squares error that output of fuzzy system, considering the rules proposed by Ants Colony Algorithm, expresses versus real diagnosis results of patients. Final goal is minimization of error so that it isknown which best rules can result in minimum error. In the end _**20**_ rules were obtained which are the final rules of the system. In figure (9) process of optimization by Ants Colony Optimization Algorithm is shown and in table (3) Optimal rules are shown.
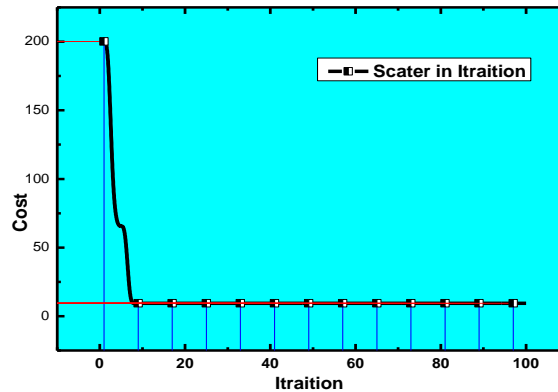

Figure 9: Process of optimization by Ants Colony Optimization Algorithm

Table 3: Optimum rules in fuzzy system that are identified by Ants Colony Optimization Algorithm

Optimal Rules that Extracted By ACO Algorithm
62. If (Preg(X1) is mf11) and (Glucose(X2) is mf20) and (BP(X3) is mf15) and (BMI(X6) is mf23) and (pedigree(X7) is mf19) and (Age(X8) is mf10) then (output1 is mf3) (1)
21. If (Glucose(X2) is mf9) and (BMI(X6) is mf8) and (pedigree(X7) is mf6) and (Age(X8) is mf3) then (output1 is mf1) (1)
42. If (Preg (X1) is mf9) and (Glucose(X2) is mf10) and (BP(X3) is mf8) and (Two Hour(X5) is mf4) and (BMI(X6) is mf8) and (pedigree(X7) is mf16) and (Age(X8) is mf5) then (output1 is mf1) (1)
58. If (Glucose(X2) is mf20) and (BP(X3) is mf13) and (BMI(X6) is mf22) and (Age(X8) is mf11) then (output1 is mf3) (1)
74. If (Glucose(X2) is mf20) and (BP(X3) is mf16) and (Two Hour(X5) is mf8) and (BMI(X6) is mf22) and (pedigree(X7) is mf24) and (Age(X8) is mf12) then (output1 is mf1) (1)
79. If (Glucose(X2) is mf24) and (Two Hour(X5) is mf10) and (BMI(X6) is mf22) and (pedigree(X7) is mf27) and (Age(X8) is mf16) then (output1 is mf3) (1)
45. If (Preg(X1) is mf9) and (Glucose(X2) is mf10) and (BP(X3) is mf8) and (Triceps(X4) is mf6) and (Two Hour(X5) is mf5) and (BMI(X6) is mf8) and (pedigree(X7) is mf16) and (Age(X8) is mf6) then (output1 is mf1) (1)
12. If (Glucose(X2) is mf5) and (BP(X3) is mf3) and (BMI(X6) is mf4) and (pedigree(X7) is mf5) and (Age(X8) is mf1) then (output1 is mf3) (1)
13. If (Glucose(X2) is mf6) and (BP(X3) is mf3) and (BMI(X6) is mf2) and (pedigree(X7) is mf5) and (Age(X8) is mf1) then (output1 is mf1) (1)
22. If (Glucose(X2) is mf9) and (BMI(X6) is mf8) and (pedigree(X7) is mf7) and (Age(X8) is mf3) then (output2 is mf1) (1)
69. If (Glucose(X2) is mf23) and (BMI(X6) is mf27) and (pedigree(X7) is mf22) and (Age(X8) is mf12) then (output1 is mf1) (1)
21. If (Glucose(X2) is mf9) and (BMI(X6) is mf8) and (pedigree(X7) is mf6) and (Age(X8) is mf3) then (output1 is mf1) (1)
67. If (Glucose(X2) is mf21) and (BMI(X6) is mf27) and (pedigree(X7) is mf22) and (Age(X8) is mf12) then (output1 is mf1) (1)
20. If (Glucose(X2) is mf25) and (BMI(X6) is mf8) and (Age(X8) is mf2) then (output1 is mf3) (1)
77. If (Glucose(X2) is mf24) and (Two Hour(X5) is mf10) and (BMI(X6) is mf22) and (pedigree(X7) is mf25) and (Age(X8) is mf15) then (output1 is mf2) (1)
29. If (Glucose(X2) is mf12) and (Triceps(X4) is mf3) and (BMI(X6) is mf12) and (pedigree(X7) is mf10) and (Age(X8) is mf2) then (output2 is mf3) (1)
17. If (Glucose(X2) is mf1) and (BMI(X6) is mf5) and (Age(X8) is mf1) then (output1 is mf3) (1)
21. If (Glucose(X2) is mf9) and (BMI(X6) is mf8) and (pedigree(X7) is mf6) and (Age(X8) is mf3) then (output1 is mf1) (1)
51. If (Glucose(X2) is mf17) and (BMI(X6) is mf19) and (Age(X8) is mf7) then (output2 is mf1) (1)
39. If (Preg (X1) is mf7) and (Glucose(X2) is mf10) and (BP(X3) is mf8) and (BMI(X6) is mf14) and (pedigree(X7) is mf11) and (Age(X8) is mf2) then (output2 is mf1) (1)

By Entering input parameters (features of each patient) and considering membership functions and so-called The Fire (activation of rules), application of inference and finally use of defuzzification (center of gravity) for creating figures of defuzzification, outputs are obtained so that from this step on obtained outputs enter adaptive Neural Network. Relation between inputs with an output is shown in figure (10). Smoothness of the figure represents correct implementation of fuzzy system.
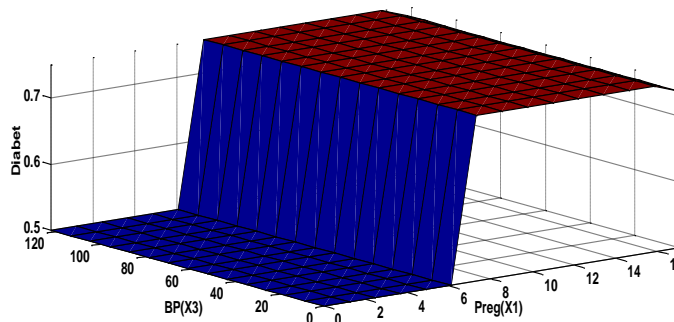
Figure 10: Physical Surface created in system. Of course in this figure relation is between first and sixth input with first output. ([1 6], 1)

8. Feed Forward Neural Network

Multilayer forward neural network that is shown in Figure (11) consists of a large number of processing parts and interrelated signals named neurons [28]. In general, input neurons add and multiply given figures of each input (p) mane and by applying bias (b) (width from the source), give the product to an activation function (f) and transfer output of the result to the next layer (a). Error Back-Propagation Algorithm is used as instructor of these networks in this model (and in this research as well). In this project a 3-layer feed forward network with Sigmoid activation function in middle layer, output layer and linear function for input layer are used for each patient. A NN schematic is shown below
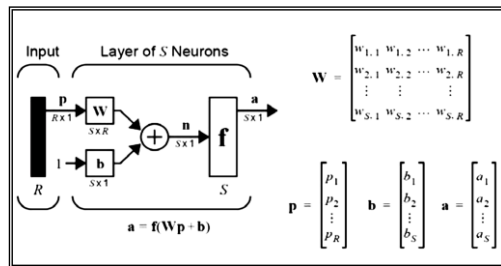

Figure (11): Neural Network Stretcher

At first (in full model) number of input layer neurons of this network was taken 2 neurons (considering output of fuzzy system for each patient). To choose the best structure, considering references used in [29] and considering the best result, 4 neurons were taken for the middle layer. In this project number of repetition is taken 500 and rate of Error Global is taken 0.02. Results of precision of coincidence of network accuracy and also histogram of network error are shown in figure (12) and Regression between Target and Output is shown in Figure (13).
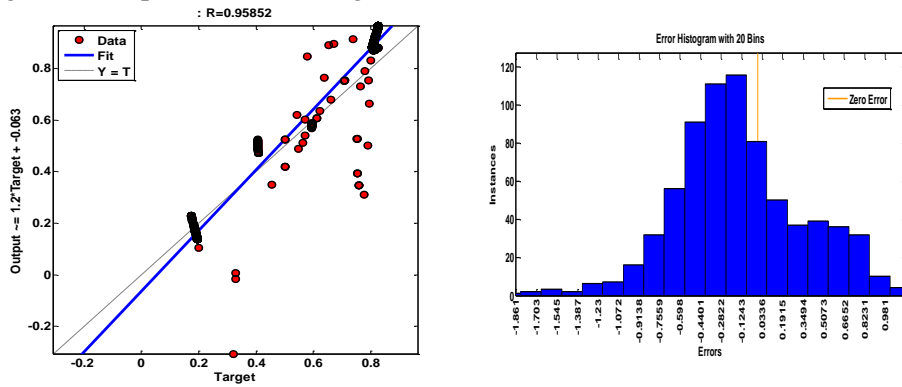
Figures 12, 13: Performance of the proposed neural system and Regression by output and Target

In the right hand graph, rate of network performance error is shown that is numeral amount of the errors that the network has obtained. The horizontal axis shows error rate whose center is zero, and the vertical axis shows the number of errors which is obtained. More Gaussian the distribution of the graphs, more the number of errors of zero and close to zero (this shows a good performance) and less the number of errors which are farther from zero (this shows a bad performance).Of course in this network close-to-zero errors were many and farther-from-zero errors were few. In the left hand axis there is a graph in which estimated figures are shown according to real figures. Of course closer the estimated figure to the real figure, closer the calculated regression figure to 1, and of course the regression figure calculated in this network is a proper and acceptable figure.

283

9. Test Accuracy and Final Evaluation

In general all process of feature (optimum rules) selection, which resulted in formation of a model for disease diagnosis, can be expressed in this way: considering the number of rules (formed by the decision tree) for disease diagnosis, vectors full of numbers of rules are produced by Ants Algorithmin storage of bank of rules, the string is multiplied by storage of the patient's bank of rules and considering the features (rules), in each effort, an array in cost function of Ants Algorithm satisfies the fitness function. This process is repeated so many times that the best rules, while the number of rules is small, have the best diagnosis and the highest accuracy. Then, optimum rules are applied to fuzzy system and as a result of application of optimum rules to fuzzy system, fuzzy expression is made for worsening of the diseaseor healthiness of each person. In the end, to form adaptation in the designed system and for generalization forward neural network was used to universalize this system. To prove capabilities of the proposed system, 10-group cross-validation (by K-fold) was done for results of database of 768 samples which were divided into 10 groups. 9 groups of 77 data and 1 group of 75 data so that each part had a mixture of all (sick and healthy) people. In this kind of validation, 9 groups of data are used for instruction and the remaining group is used for testing. This process of instruction and testing is done sequentially-rotationally for 10 times so that every time a different group is left aside for testing.

10. Results of Simulations and Evaluations

MATLAB® software was used for simulation. Instruction of neural network was done simultaneously. For each input or one line of data set all figures (13) change, and to determine the Accuracy we use Equation (6) or Accuracy Definition. Results are shown in table (4). In evaluation by [8] 202, in evaluation by [3] 200, in evaluation by [3] 217, in evaluation by [9] 192, in evaluation by [11] 197, and in the proposed method 247 patients were diagnosed. So, you can easily decide on accuracy and efficiency of the proposed method.

Table 4: Results of simulations

Method under study	Accuracy	Best obtained result	Diagnosed patients
MLPNN[8]	76.89%	not available	202 patients
Evolutionary [3]	76.17%	---	200 patients
SVM[3]	82.50%	---	217 patients
RBF network [9]	73.13%	---	192 patients
K-NN[11]	75.55%	---	197 patients
Method proposed in this research	%92	92.45%	247 patients

11. Discussion and Conclusion

Diabetes is one of the most prevalent diseases in the world which can even result in mutilation. In this research considering the necessity of early and on time diagnosis of this disease, a new method based on combining intelligent systems is presented. Many researchers like to use this tool, yet the challenge of instruction of neural networks is their main concern. Combination of Evolutionary Algorithms, fuzzy logic, and data mining systems can be a good idea for improving efficiency of these networks (in this research diagnosis and selection of features (rules)) which in the end result in correct grouping and performance of data. Results of simulations show that optimum rules in fuzzy logic can improve accuracy and efficiency of the designed system; even in interaction of results of this step with neural network, it will have a better, faster, and more accurate output for diagnosis in generalization of the designed system. Of course the main goal in this type of research is to reach a 100 percent accuracy and precision so that you can trust artificial intelligent systems and give them a practical dimension. Reaching this goal is possible just when you have a great knowledge and accurate understanding of

intelligent systems on the one hand and of disease and data mining on the other hand that may be achieved in a close future.

Reference:

[1]   World Health Organization, Diabetes Center, Fact Sheet N°312, www.who.int/mediacentre/factsheets/fs312/en **(2011)**.

[2]   Centers for Disease Control and Prevention, National Diabetes, Fact Sheet, www.cdc.gov/diabetes **(2011)**.

[3]   M.R. NaiemAbadi, N. A.A Chamachar, E. Tahami, H. Rabbani, Diabet Diagnosis By SVM . Proceeding of the 14th Iranian student Conference on Electrical Engineering, **(2011)** September 6-8; Kermanshah, Iran.

[4]   American Diabetes Association, Diabetes Basics, www.diabetes.org/diabetes-basics **(2011)**.

[5]   L. I. Kuncheva, J. C. Bezdek, R. P. W. Duin, Decision templates for multiple classifier fusion: an experimental comparison, Pattern Recognition Journal, Vol 34, No 2, pp 299–314, **(2001)**.

[6]   B. Yaxin, B. David, W. Hui, G. Gongde, G. Jiwen, Combining Multiple Classifiers Using Dempster's rule for text Caregorization, Applied Artificial Intelligence Journal, Vol 21, No:3, pp 211- 239 ,**(2007)**.

[7]   L. I. Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, New York, NY: Wiley Interscience Publisher, **(2005)**.

[8]   S. Tulyakov, S. Jaeger, V. Govindaraju, D. Doermann, Review of Classifier Combination Methods, Studies in Computational Intelligence Journal, Vol 90, No 2, pp 361-386, **(2008)**.

[9]   J. Siti Fahanah Bt and A. Darmawaty Mohd, Diabetes Mellitus Forecast using artificial neural network (ANN), Life Science Journal, Vol 10, No 4,pp 1025- 1031, **(2005)**.

[10]   S.E. Thami, M.A. Khalil Zade, Intelligent Diabet Diagnosis by Artificial neural Network. Proceeding of the 12 th Iranian Society for Biomedical Engineering (ISBME), **(2005)** Nov 16-18; Sahand University of Thecnology, Iran.

[11]   Y. Jiang and Z. Zhou, Editing Training Data for KNN Classifiers with Neural Network Ensemble. Journal of Advances in Neural Networks Lecture Notes in Computer Science,Vol 25, No 4, pp 356-361 **(2004)**.

[12]   D.J. Newman, S. Hettich, C.L.S. Blake and C.J. Merz, repository of machine learning database. Irvine, UCI, CA: University of California, Dept. of Information and Computer Science, archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes **(1998)**.

[13]   R. Jensen, Combining rough and fuzzy sets for feature selection. Ph.D. Thesis. School of informatics, Univertsity of Edinburgh, Engeland **(2005)**.

[14]   M. Alipoor, J. Haddadnia, Introduction of an intelligent system for accurate diagnosis of breast cancer. Iranian Journal of Breast Disease (IJBD), Vol 2, No 2, pp 33-40, **(2009)**.

[15]   T. Xiang, and S. Gong, Spectral clustering with eigenvector selection. Pattern Recognition journal. Vol. 41, No. 3, pp 1012-1029, **(2008)**.

[16]   A. Qarehkhani, M. Fiuzy, J. Haddadnia. A Anovel For Diabet Diagnosis based on Combining Intelligent System such as Fuzzy System ,Decission Tree, Adaptive Neuro Fuzzy System. Proceedings of the 4th Conference on Information & Knowledge Technology (IKT), **(2012)** , Babaol, Iran.

[17]   S. Mitra, K. M. Konwar and S. K. Pal. Fuzzy Decision Tree. Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation. Journal of IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, Vol. 32, No 14, pp. 328-339, **(2002)**.

[18]   S.M.M. Ziaberi, Ebadzadeh. Fuzzy System and Method. B. Sc Thesis, Amirkabir University of Technology, July **(2009)**.

[19]   M. Dorigo, and L.M. Gambardella. Ant Colony System: A Cooperative Learning approach to the traveling salesman Problem. Journal of IEEE Transactions Evolutionary Computation, Vol 1, No 1, pp :53-66, **(1996)**.

[20]   M. Dorigo and L.M. Gambardella, Ant Colonies for the Traveling Salesman Problem. Journal of Bio systems,Vol 43, No 18, pp 73-81 **(1997)**.

[21]   Frederick Ducatelle. Adaptive Routing in Ad Hoc Wireless Multi-hop Networks . Università Della Svizzera Italiana Publisher, Milan, Italy, **(2007)**.

[22]   A. Ebrahimi. Algorithms for the imperfections of the social ants entering and leaving the power plant units. Proceeding of 24ᵗʰ International Conference of Power System, **(2009),** November, Tehran, Iran.

[23]   M. Sugeno. Industrial Applications of Fuzzy Control. Elsevier, Book Publisher, New York, **(1985)**.

[24]   F. Bergh and A. Engel Brecht. A new locally convergent particle swarm optimizer, IEEE, Fundamental Informatics Journal, Vol 18, No 4, **(2002)**.

[25]   F. Herrera and M. Lozano. Fuzzy adaptive genetic algorithm: design Taxonomy and future directions. Journal of Soft Computing vol 7, No 3, pp 545-562, **(2003)**.

[26]   J. Mahjobi, A.F. Shahidi. Wave Height Estimation from Wind in Neka by Decision Tree. Proceeding of 7th Iranian Data Mining Confrence, **(2007)**, Nov 20-21, Amirkabir University of Technology, Iran.

[27]   M. Kantardzic. Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons Publisher, **(2003)**.

[28]   Madan M. Gupta, Liang Jin, and Noriyasu Homma. Static and Dynamic Neural Networks from fundamental to advanced theory. Forwarded by Lotfi A. Zadeh. IEEE Publisher, **(2003)**.

[29]   M. Fiuzy, A. Qarehkhani, J. Haddadnia. An Intelligent Combined Method for Diabet Diagnosis. Proceeding of 20th Iranian Confrence of Electrical Engineering, **(2012)**,May 15-17, Tehran University.

[30]   A. Baraldi, and P. Blonda. A Survey of Fuzzy Clustering Algorithms for Pattern Recognition. Journal of IEEE Tracsactions on systems. Main and cybernetics part b: cybernetics. Vol 29, No 6: pp 786-801, **(1999)**.