International Journal of Engineering

Journal Homepage: www.ije.ir

# Diagnosis of Coronary Artery Disease via a Novel Fuzzy Expert System Optimized by Cuckoo Search

S. Moameri[a], N. Samadinai*[b]

[a] *Department of Computer Engineering, University of Bojnord, Bojnord, Iran*
[b] *Department of Computer Engineering, Kosar University of Bojnord, Bojnord, Iran*

*PAPER INFO*

*Paper history:*
Received 08 June 2018
Received in revised form 26 July 2018
Accepted 17 August 2018

*Keywords*:
*Coronary Artery Disease*
*Fuzzy System*
*Cuckoo Search*
*Heart Disease*
*Decision Tree*

*ABSTRACT*

In this paper, we propose a novel fuzzy expert system for detection of Coronary Artery Disease, using cuckoo search algorithm. This system includes three phases: firstly, at the stage of fuzzy system design, a decision tree is used to extract if-then rules which provide the crisp rules required for Coronary Artery Disease detection. Secondly, the fuzzy system is formed by setting the intervals for fuzzy variables and extracted rules. Finally, Cuckoo Search algorithm is used to optimize fuzzy membership functions. The accuracy of our proposed system is evaluated using Cleveland Cardiac Patient Database. The detection rate is 93.48% employing optimized membership functions. Also, 85.76% accuracy is obtained for predicting the risk of coronary artery disease. The superiority of proposed system is obvious by comparing it to the previously methods; it is more accurate and is also easier to implement.

***doi**: 10.5829/ije.2018.31.12c.06*

## 1. INTRODUCTION

Coronary artery disease (CAD) as the most common type of heart disease is the cause of a growing incidence of death in both women and men. This disease affects 17 million people worldwide and is the leading cause of death among other cardiovascular diseases [1] as 23.6 million deaths in 2030 are estimated by the World Health Organization (WHO). In coronary artery disease the coronary arteries of the heart are clogged by fat deposits (atherosclerosis). This condition limits the blood and oxygen needed in the heart especially when doing physical activity. For many people the first sign of heart dysfunction is a heart attack that occurs when the blood clot in the coronary arteries blocks the flow of blood into a part of the heart muscle [2].

There are several methods to detect coronary artery disease such as Exercise Tolerance Test (ETT), electrocardiography (ECG), angiography, or cardiac catheterization. But patient pains and inadequate accuracy in diagnosis limit using all these methods; therefore, doctors are pursued to apply computer aided systems [3]. Computer aided methods which extract effective features and use them in the classifications for the early detection of the diseases, overcome these problems.

In general, some risk factors are responsible for coronary artery disease. Reviewing various sources points the effective risk factors for coronary artery disease are: smoking, high blood pressure, high fat (high cholesterol, high triglycerides, high LDL and low HDL), diabetes, physical inactivity, obesity, age, gender and family history [4, 5]. Based on these risk factors, a variety of algorithms such as decision tree [6, 7], linear support vector machine [3-8] and various neural networks have been proposed and developed for detecting and preventing coronary artery disease.

A computer aided system has been proposed with 87% accuracy to diagnose coronary artery disease using artificial intelligence recognition system and K-nearest neighbor (KNN) [9]. Dennis and Muthukrishnan have proposed a genetic fuzzy system [10]. This fuzzy model is trained based on Genetic algorithm and the highest accuracy (76.67%) is obtained after repeating 20 times. An automatic classifying method has been presented to

*Corresponding Author Email: *nsamadiani@kub.ac.ir* (N. Samadiani)

Please cite this article as: S. Moameri, N. Samadinai, Diagnosis of Coronary Artery Disease via a Novel Fuzzy Expert System Optimized by Cuckoo Search, International Journal of Engineering (IJE), IJE TRANSACTIONS C: Aspects Vol. 31, No. 12, (December 2018) 2028-2036

detect the medicine data using wavelet transform and type-2 fuzzy logic [11].

Various hybrid methods have been developed to address the problems of detecting coronary artery disease. Nguyen et al. [12] have proposed a standard additive model (SAM) combined by Genetic algorithm and called it GSAM. This model used discriminative features in big data sets and achieved to 77.48% accuracy on evaluating of Cleveland data. The risk of heart disease in the patients has been investigated by Anooj via designing a decision support system based on fuzzy rules [13]. The best accuracy was obtained 62% from classification of Cleveland data set. Marateb and Goudarzi have proposed a fuzzy rule-based system based on neuro-fuzzy classifier optimized by conjugate gradient algorithm [14]. The proposed system has been achieved 84% accuracy for predicting Cleveland dataset. In another work, combining C4.5 algorithm and fuzzy expert system could diagnose coronary artery diseases patients with 81.82% acuuracy [15]. Safdarian et al. [16] have applied different classification methods such as probabilistic neural network, KNN, MLP and Naïve Bayes (NB) to diagnose myocardial infraction (MI). The highest accuracy of 76.67% was obtained from NB by studying 549 records. A hybrid method of prediction of coronary artery disease was proposed by Arabasadi et al. [17]. This method can improve the learning power of neural network approximately 10% by optimization of weights using Genetic algorithm. The 88.25% of accuracy was obtained from the proposed network. Recently, an automatic fuzzy system has been proposed based on Genetic algorithm and modified dynamic multi-swarm particle optimization (MDMS-PSO) which is able to diagnose coronary artery disease with 92.1% accuracy on Cleveland dataset [18].

In this paper we propose a novel method to diagnose the coronary artery disease. After extracting the rules of fuzzy system by training decision tree and removing unimportant features, the fuzzy system is designed and the proper fuzzy intervals are specified. Finally, these fuzzy values are optimized by cuckoo search algorithm. This optimized fuzzy system by cuckoo search can predict coronary artery disease with high accuracy.

In the rest of paper, the proposed method and different steps are introduced and explained in section 2. In section 3 the results of evaluating the proposed method are reported and the conclusion finishes the paper.

## 2. METHODOLOGY

In this paper, an expert fuzzy system is proposed based on cuckoo search algorithm that consists of four stages: first, the missing data is removed using K-means clustering data imputation. A decision tree is constructed and the rules are extracted in the second phase. In the third step, the crisp rules are converted to fuzzy rules set by fuzzy membership functions. Finally, in the last stage, the variables of fuzzy rules are tuned employing cuckoo search algorithm. The fuzzy model with optimized intervals are formed the final fuzzy system. Figure 1 shows the details of proposed method.

### 2. 1. UCI Heart Data Set

In this paper, University of California heart data set is applied. These data was collected by Hungarian Cardiac Disease Center, University Hospital of Switzerland and Cleveland Clinical Data [19]. They include 76 features which only 13 of them are used as input features in published experiments as mentioned in literature [19]. The input attributes are age, blood pressure, fasting blood sugar, resting ECG, exercise induced angina, old peak, slope, fluoroscopy, serum cholesterol, maximum heart rate, sex, chest pain type, and thallium scan. A feature shows the output and there are four possible values (0 to 4) for output variable that show angiographic status. The study concentrates on two various problems: the first is attempting to recognize healthy (value 0) and unhealthy individuals (value 1) (namely, binary classification) and the second detects healthy peoples and who suffer from coronary artery disease (values 1, 2, 3, 4) (multi-class problem).

### 2. 2. Preprocessing

When data value is absent from data set, Missing data or missing value occurs [20]. Missing data is an inevitable part of data acquisition and collection which is caused by many reasons such as poor performance of equipment and lack of tendency to fill gaps of questionnaires in traditional data collection methods. There are many databases such as the introduced data set which include mising data. Figure 2 shows some missing values in the used data set. Missing data can negatively affect the overall performance of any intelligent system. Morover, as we deal with patients' data and the accuracy of system is crucial, eliminating these corrupted or missing data from dataset can virtually increase the diagnosis quality of the proposed method.

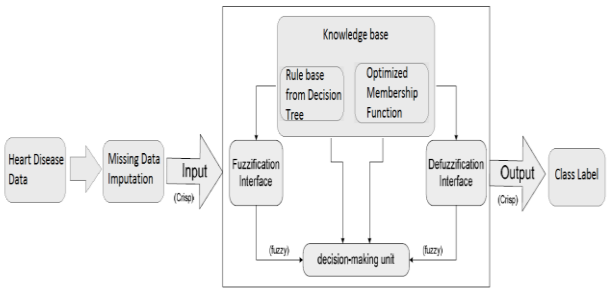



**Figure 1.** The details of proposed method

| Age | Sex | CP | TB | CH | Fbs | RC | TC | Exang | OP | Slope | Ca | TS | Label |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57 | 0 | 2 | 130 | 236 | 0 | 2 | 174 | 0 | 0 | 2 | 1 | 3 | 1 |
| 38 | 1 | 3 | 138 | 175 | 0 | 0 | 173 | 0 | 0 | 1 | ? | 3 | 0 |
| 47 | 1 | 4 | 110 | 0 | 1 | 1 | 149 | 0 | 2.1 | 1 | 2 | ? | 2 |
| 47 | 1 | 4 | 160 | 0 | ? | 0 | 124 | 1 | 0 | 2 | 1 | 7 | 1 |
| 59 | 1 | 4 | 164 | 176 | 1 | 2 | 90 | 0 | 1 | 2 | 2 | 6 | 3 |

**Figure 2.** Some samples of Missing values in UCI heart data set. They are marked by red rectangular.

In this paper, we employ K-means clustering data imputation method to estimate the missing values. Imputation with K-means clustering method is comprised of three stages:

1. We have N data X= {$x_1$, $x_2$, …, $x_N$} where $x_{ij}$ represents record $x_i$ ($1 \le i \le N$) with S features ($1 \le j \le S$). If {$x_{ij} = \varphi \mid \forall\ 1 \le j \le S$}, record $x_i$ is an *incomplete* object which has missing values and we apply R= {$j \mid x_{ij} \neq \varphi$, $1 \le j \le S$} to show a set of features with available values and these are called *reference* features. We aim to estimate the values of non-reference features for incomplete records. K random points are chosen as cluster centres in where C= {$c_1$, $c_2$, …, $c_k$} is cluster centroids and $c_i$ ($1 \le i \le K$) presents the centre of cluster i and is a S-dimension vector.
2. Similarity between all data samples and each cluster centre are calculated using similarity measures like Euclidean distance as in Equation (1). Samples are assigned to a cluster with the minimum distance. For missing data, we ignore the non-reference features and calculate the distance by Equation (2). The process of assigning data samples to clusters and updating the cluster center based on the average of data samples assigned to particular clusters continues until the summation distance is less than a user-defined threshold.

$$d(c_k, d_i) = \Big(\sum_{j=1}^{S} \left|d_{i,j} - c_{k,i}\right|^2\Big)^{\frac{1}{2}} \qquad (1)$$

$$d(c_k, d_i) = \Big(\sum_{j=1}^{R} \left|d_{i,j} - c_{k,i}\right|^2\Big)^{\frac{1}{2}} \qquad (2)$$

3. The final stage of the process is filling the missing data fields (non-reference attributes). The missing fields of samples are filled by using the nearest neighbor in the cluster that the missing data sample belongs to Patil et al. [21].

In this paper, we set number of clusters 2 and 5 respectively according to the binary and five-class problmes which we aim to solve them. We repeated K-means clustering 15 times and computed summation of distances in each iteration. Finally, by choosing minimum value of summations as threshold, the K-means data imputation is applied and missing values are estimated.

## 2. 3. Decision Tree

The decision tree is one of the most widely used data mining algorithms. The decision tree is a learning algorithm based on training data and the advantages are easiness, clarity and the ability to extract rules. A decision tree is a classifier that divides the data area into sub part based on maximum information gain [22]. The learning system of a common decision tree obeys a top-down strategy. Each inner or non-leaf node is characterized by a feature. Also, the leaves represent a class or a set of possible solutions. Every path to leaf nodes shows a classification rule in decision tree [23].

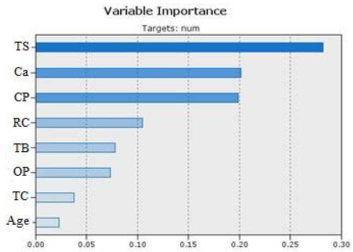
The rules play a key role in fuzzy system. In this paper decision tree is used to extract crisp rule base. Firstly, we apply decision tree to classify healthy and unhealthy people included 13 features. Therefore, the classification accuracy is obtained 83.87%. Because the accuracy of this step is critical for the next stages of the proposed method, we have to improve the performance of decision tree. Consequently, by showing more important features we start to remove the less momentous attributes. According to Figure 3 (a) there are only 8 features which are effective in classification. So, we remove five less important variables and apply decision tree to data included 8 attributes. Finally, the classification accuracy is obtained 93.07%. Figure 4 (a) shows the extracted crisp rules from tree. We do these steps to predict the risk of CAD in multi-class problem and achieve to 81.25% accuracy using data contained 13 features. By illustrating the more important attributes in Figure 3(b), it is seen that removing four less important features makes improving the accuracy to 81.85%. Figure 4(b) reveals the extracted rules for this problem.

## 2. 4. The Fuzzy Inference System

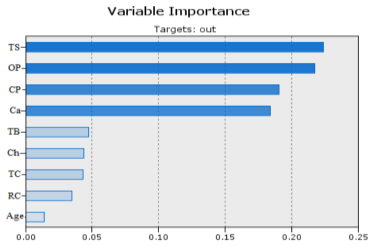
Fuzzy modeling is one of the special arithmetic ways to solve complicated, ambiguous real problems that are nearer to humans' natural language. Three fundamental aspects form a fuzzy model: the fuzzy processing, fuzzy inference system (FIS) or decision unit and deffuzication. The process of designing fuzzy model includes steps as follows:

1. Setting input and output variables;
2. Introducing fuzzy membership functions per variable;
3. Extracting fuzzy rules;
4. Tuning parameters of (2) and (3).

The steps 1, 2 and 3 organize the structure of fuzzy model but the task 4 involves in tuning and optimizing the model parameters [24, 25].




(a)




(b)

**Figure 3.** The importance of features in (a) binary problem and (b) multi-class problem. Thallium scan is the most important attribute and age is the least important feature. (TS= Thallium scan, CP= Chest pain type, OP= Old peak, Ca= number of major vessels colored by fluoroscopy, RC = resting electrocardiographic results, TB= resting blood pressure, TC= maximum heart rate achieved, Ch= Cholesterol)

In this paper, we obtained If-Then rules by decision tree and it is necessary to turn them into fuzzy variables. As a result, the crisp rules set are mapped to a fuzzy model using a triangular membership function in the fuzzification step. The triangular membership is defined as following (Equation (3)):

$$\text{Triangle}(y: a, b, c) = \begin{cases} 0 & y < a \\ \dfrac{y-a}{b-a} & a \le y \le b \\ \dfrac{c-y}{(c-b)} & b \le y \le c \\ 0 & y > c \end{cases} \tag{3}$$

The designing of fuzzy membership functions for every variable is an important phase of fuzzy modeling. There are 9 input variables and one output in the proposed fuzzy model (Table 1 shows the corresponding fuzzy values); They are as following:

**Age:** This variable is divided to four fuzzy set included "Very old", "Old", "Middle", and "Young".

**Chest pain type (CP):** This input reveals four types of chest pain: 1 = Typical Angina, 2 = Atypical Angina, 3 = Non-anginal pain, 4 = Asymptomatic. Each one is a fuzzy set and fuzzy sets do not overlap together.

**Resting blood pressure (TB):** This variable shows blood pressure in resting time that is modeled to "Low", "Medium", "High" and "Very high" fuzzy sets.

**Resting electrocardiographic results (RC):** This input has three values 0, 1 and 2 which correspond to normal, ST-T abnormality and left ventricular hypertrophy, respectively.

| | | |
|---|---|---|
| If TS <=3 and Ca <=0 and TB >156 and Age <= 62 | Then | CAD |
| If TS <=3 and Ca <=0 and TB >156 and Age > 62 | Then | Healthy |
| If TS <=3 and Ca > 0 and CP <=3 and Age > 55 and Age<= 59 | Then | CAD |
| If TS <=3 and Ca > 0 and CP > 3 and RC<= 1 and Age <=63 | Then | CAD |
| If TS <=3 and Ca > 0 and CP > 3 and RC<= 1 and Age > 63 | Then | Healthy |
| If TS >3 and CP>3 and OP>0.3 | Then | Healthy |
| If TS > 3 and CP>3 and OP<=0.3 and RC<= 1 and TB >136 | Then | CAD |
| If TS > 3 and CP >3 and OP<=0.3 and RC<= 1 and TB <=136 and TC<=233 | Then | Healthy |
| If TS > 3 and CP <= 3 and Ca <= 0 and RC <= 1 and TB <=130 and OP >0. 9 | Then | CAD |
| If TS > 3 and CP <= 3 and Ca <= 0 and RC <= 1 and TB > 130 | Then | Healthy |
| If TS > 3 and CP <= 3 and Ca > 0 and RC <= 1 and TC <= 141 | Then | CAD |
| If TS > 3 and CP <= 3 and Ca > 0 and RC > 1 | Then | CAD |

(a)

| | | |
|---|---|---|
| If TS > 3 and CP<=3 and OP <=1.9 and TS<=6 and Ca<=0 | Then | Healthy |
| If TS > 3 and CP<=3 and OP <=1.9 and TS<=6 and Ca>0 | Then | CAD (value 2) |
| If TS > 3 and CP<=3 and OP <=1.9 and TS>6 and TC<=109 | Then | CAD (value 2) |
| If TS > 3 and CP<=3 and OP <=1.9 and TS<=6 and TC>109 | Then | Healthy |
| If TS > 3 and CP<=3 and OP <=1.9 and TS<=6 and TC>109 and Age <=52 | Then | CAD (value 1) |
| If TS > 3 and CP<=3 and OP <=1.9 and TS<=6 and TC>109 and Age>52 | Then | Healthy |
| If TS > 3 and CP<=3 and OP >1.9 and Age<=58 and Ca<=1 | Then | CAD (value 4) |
| If TS > 3 and CP<=3 and OP >1.9 and Age<=58 and Ca>1 | Then | CAD (value 3) |
| If TS > 3 and CP<=3 and OP >1.9 and Age>58 | Then | CAD (value 2) |
| If TS > 3 and CP>3 and Ca<=2 and OP<=0.7 and RC<=1 and Ca<=0 | Then | Healthy |
| If TS > 3 and CP>3 and Ca<=2 and OP<=0.7 and RC<=1 and Ca>0 | Then | CAD (value 2) |
| If TS > 3 and CP>3 and Ca<=2 and OP<=0.7 and RC>1 | Then | CAD (value 1) |
| If TS > 3 and CP>3 and Ca<=2 and OP>0.7 and TS>6 and Age<=58 and Ch>302 | Then | CAD (value 2) |
| If TS > 3 and CP>3 and Ca<=2 and OP>0.7 and TS>6 and Age>58 | Then | CAD (value 3) |
| If TS > 3 and CP>3 and Ca>2 and TB >130 | Then | CAD (value 4) |
| If TS > 3 and CP>3 and Ca>2 and TB <130 | Then | CAD (value 3) |
| If TS > 3 and CP>3 and Ca<=2 and OP>0.7 and TS>6 and Age<=58 and Ch<=302 and Ca>1 | Then | CAD (value 3) |
| If TS > 3 and CP>3 and Ca<=2 and OP>0.7 and TS>6 and Age<=58 and Ch<=302 and Ca <=1 and RC<= 109 | Then | CAD (value 1) |
| If TS > 3 and CP>3 and Ca<=2 and OP>0.7 and TS>6 and Age<=58 and Ch<=302 and Ca <=1 and RC> 109 | Then | CAD (value 3) |

(b)

**Figure 4.** The crisp rules set extracted from decision tree: a) binary problem, b) multi-class problem (the output has values (0, 1, 2, 3, 4)). (TS = Thallium scan, CP = Chest pain type, OP = Old peak, Ca= number of major vessels colored by flourosopy, RC= resting electrocardiographic results, TB= resting blood pressure, TC= maximum heart rate, Ch= cholesterol)

**Maximum heart rate (TC):** This variable is divided to three fuzzy sets "Low", "Medium", and "High".

**Old peak (OP):** This input is modeled by "Low" and "High" fuzzy sets.

**Number of major vessel colored by fluoroscopy (Ca):** This variable shows the number of blood vessels that is highlighted by fluoroscopy and has four values 0, 1, 2, and 3.

**Thallium scan (TS):** This input shows patient's defect by three values (3, 6, and 7) matched to normal, fixed defect, and reversable defect, respectively.

**Cholesterol (Ch):** This field is designed in four fuzzy sets: "Low", "Medium", "High" and "Very High".

Moreover, we need to apply a fuzzy operator for combining and modifying the fuzzy sets. Three fundamental operators are AND, OR and NOT which we have used AND(min) operator. This operator results the smallest membership value between elements on the sets involved.

In this paper, the Takagi-Sugno-Kang (TSK) fuzzy inference system is applied which has been introduced by Takagi-Sugeno-Kang in 1985 [26]. This method is less time-consuming than another fuzzy inference technique (Mamdani) and system output isn't represented by fuzzy set; in fact, it yields a constant or a consequence in the form of linear equations which

**TABLE 1.** Input variables, the ranges and fuzzy sets

| Input variable | Range | Fuzzy set |
|---|---|---|
| **Age** | 0-34 | Young |
| | 30-35 | Middle |
| | 40-58 | Old |
| | 52-77 | Very Old |
| **Resting Blood Pressure** | 0-134 | Low |
| | 126-154 | Medium |
| | 142-172 | High |
| | 154-354 | Very High |
| **Maximum Heart Rate** | 0-141 | Low |
| | 111-194 | Medium |
| | 153-353 | High |
| **Old Peak** | 0-4.2 | Low |
| | 2.55-7 | High |
| **Cholesterol** | 0-198 | Low |
| | 188-250 | Medium |
| | 217-307 | High |
| | 281-681 | Very High |

determines the order of method. Generally, a rule in a zero-order Sugeno fuzzy model is in the following form:

$$\text{If } (x_1=A_1)\ (x_2=A_2)\ (x_3=A_3) \ldots (x_N=A_N) \quad \text{Then } z=k \tag{4}$$

where $A_i$, and k represent i-th fuzzy set, fuzzy operator and output constants, respectively. In the proposed system, the output demonstrates the patient's angiography status. This variable has five values (0 to 4) which 0 and values (1 to 4) correspond to healthy person with less than 50% diameter narrowing and heart disease conditions (more than 50% diameter narrowing), respectively. Therefore, because the output is a constant, we have selected zero-order TSK fuzzy inference method to design the fuzzy diagnosis system. Finally, weighted average is employed to calculate the crisp output in the defuzzification stage [27]. Also, it is noticeable that we solved two problems: a binary problem to distinguish healthy person and patient and a five-class problem to detect the risk of CAD determined by values greater than 1.

**2. 5. Cuckoo Search Optimization of Fuzzy Membership Values** A suitable fuzzy rule base plays a critical role in designing fuzzy system. Experts usually generate rules and membership functions for a specific area because their definition is influenced by particular decisions. Although extracting fuzzy rules is relatively easy, but finding membership values of functions is considered as a difficult task. Also, optimization of values is a time-consuming process. As mentioned above, a fuzzy system is designed by own particular membership functions and the performance of system depends on the type of fuzzy membership function and the parameters. Against their importance, there is no accurate method accessible for determining them.

The fuzzy systems can be considered as a space search problem where each point in the space shows fuzzy rules and membership values. Therefore, using evolutionary algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), and cuckoo search (CS) may be the best choices for probing the spaces [28].

However, evolutionary algorithms perform similarly but CS does not rely on the operators like crossover and mutation in GA; moreover, it results more accurate and robust than PSO. Also, the statistics analysis reveals the CS success of solving problem is completely more than PSO and the known stable problem of PSO limits the success rate rather than CS [29]. Other CS advantages are modernity and better convergence in comparison to GA.

Xin-She Yang and Suash Deb have introduced one of the latest evolutionary algorithm named Cuckoo Search (CS) in 2009 [30]. CS is based on brood parasitism of cuckoo birds. Laying eggs of some species of cuckoo in the nest of other host nest is considered as the primary characteristic of them. If a host cuckoo recognizes the eggs are not their own, it will throw the unfamiliar eggs or abandon its nest and create a new one elsewhere [31]. As a result, the cuckoo is extremely expert in mimicking the color and model of their host eggs with their own. An optimization algorithm has been idealized such reproducing behavior. A standard CS algorithm always follows three rules:

- cuckoo chooses a nest in a random manner and depletes an egg in it.
- The best nests included high quality of eggs will organize the next generations.
- A fixed number shows the number of available host nests and a probability $p_d \in (0,1)$ demonstrates the identifying chance of cuckoo's egg by the host bird. There are two various solutions for host bird: dropping the egg or leaving the nest to build a novel nest.

A solution is shown by each egg in the nest while a new solution is presented by a cuckoo egg. CS aims to apply the strong better solutions when it displaces them with fairly good solutions in the nests. The quality of a nest is assessed by fitness function. In this paper, the designed fuzzy system is used as fitness function. In fact, the solutions resulted by CS are considered as new membership function parameters and a fuzzy system is constructed based on these values. This system would then be evaluated by applying data and all records would be labled as healthy or unhealthy people. By measuring the performance of system built, the importance of each solution would be computed whether it could be replaced by not-so-good solutions or not. Therefore, we face with a maximization problem which aims to increase the performance of patients' diagnosis. The initial values of solutions and nests are randomly determined. The fitness function evaluates the initial solution and Equation (5) creates the current best solution:

$$\sigma_u = \left\{ \frac{r(1+\beta)\sin\left(\frac{\pi\beta}{2}\right)}{r\left[\frac{1+\beta}{2}\right]\beta 2^{\frac{\beta-1}{2}}} \right\}^{\frac{1}{\beta}}, \quad \sigma_v = 1 \tag{5}$$

where, Levy distribution parameter is shown by β=1.5 and gamma function is represented by r. A Levy flight [32] is responsible for producing the new solution $x^{(t+1)}$ for i[th] cuckoo (Equation (6)):

$$x_i^{(t+1)} = x_i^{(t)} \alpha . S \tag{6}$$

where α>0 presents the step size and in this paper α has value of 1 such as the most cases. Also, the equation shows that new solution $x^{(t+1)}$ for i[th] cuckoo is dependent to its current solution $x^{(t)}$. S represents the length of random walk with Levy flights of Mantegna's algorithm with step size computed by Equation (7) [33]:

$$S^{(t+1)} = \frac{u^{(t+1)}}{|v^{(t+1)}|^{\frac{1}{\beta}}}(v - x_{best}) \tag{7}$$

where normal distribution approximates the values of u and v by Equation (8):

$$u \sim N(0, \sigma_u^2), v \sim N(0, \sigma_u^2) \tag{8}$$

The current best nest is generated by evaluating the new solution. If a nest has the probability less than $p_d$, it is ignored and other nests are carried to the next generation. This algorithm is iterated until arriving at maximum generation number or the minimum error rate is obtained [34, 35].

There are three parameters for every membership function shown in Figure 5: C, L, and R that shows center, first, and the last point of interval, respectively. C', L', and R' implies center, first and the last point of optimized interval. We try cuckoo search to optimize the C, L, and R points of the initial intervals mentioned in Table 1. This optimization does not influence on the shape of triangular membership functions and they just shrink or stretch. Table 2 shows the values of necessary cuckoo search parameters in optimization of membership values. Figure 6 illustrates the membership functions of age, resting blood pressure, old peak, and

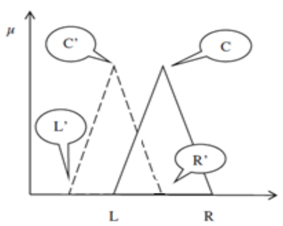



**Figure 5**. The parameters of membership functions

**TABLE 2.** Cuckoo Search parameters

| Parameter | Value |
|---|---|
| Population size | 25 |
| Max generation | 1000 |
| $p_d$ | 0.25 ( recommended in [33]) |
| $\beta$ | 1.5 ( recommended in [33]) |

maximum heart rate variables before and after optimization in binary problem.

## 3. RESULTS

We prepared the fuzzy rule base and used fuzzy toolbox of Matlab 2017 to create fuzzy inference system. As said before, rules were obtained from decision tree trained by UCI data set. We started to optimize fuzzy membership values after initializing features by values mentioned in Table 1 and set cuckoo search parameters based on values of Table 2. Then the algorithm began to find the optimum membership values and obtain the best classifying accuracy. In fact, the accuracy of classification is calculated by fuzzy inference system with new membership values obtained by cuckoo search in every repetition. This algorithm repeated 1000 times and the best results were stored. To evaluate the proposed method, we used confusion matrix of its performance and calculated accuracy, specificity and sensitivity. In this matrix, four values (TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative)) are measured. Sensitivity and specificity measure accuracy of the proposed method to find healthy people and who are suffered from coronary heart disease, respectively. Moreover, accuracy demonstrates the number of correct classified people to all people. Table 3 shows these criterions and the results. As seen in Table 3, there are the results of evaluating the proposed method using Cleveland, Hungarian and Switerzland data sets for two different binary and multi-class problems.

We have compared the proposed method with similar methods in the literature to demonstrate the superiority of it. Table 4 shows this comparison. The mentioned results have been obtained by Cleveland data set which missing values are eliminated or estimated by different methods. It is clear the proposed method has been obtained the highest accuracy in detecting the healthy and unhealthy people. Moreover, since the result is more precise due to applying fuzzy expert system to estimate the patient status, it is more reliable for the doctors. Using the proposed method as a predictor can extremely be impressive for the medicine because it is capable of evaluating the risk of CAD with high accuracy (multi-class problem), while the existence researches have solved only the binary problem. It is noticeable that implemention of the proposed method is simpler and needs less memory in comparison to other works especially study of Paul et al. [18] which estimates the number of swarms in a dynamic way; therefore, it is more time-consuming and requires more memory to predict the disease in running time. Also, we automatically extract the rules by decision tree while other researches employ the rules suggested by experts.

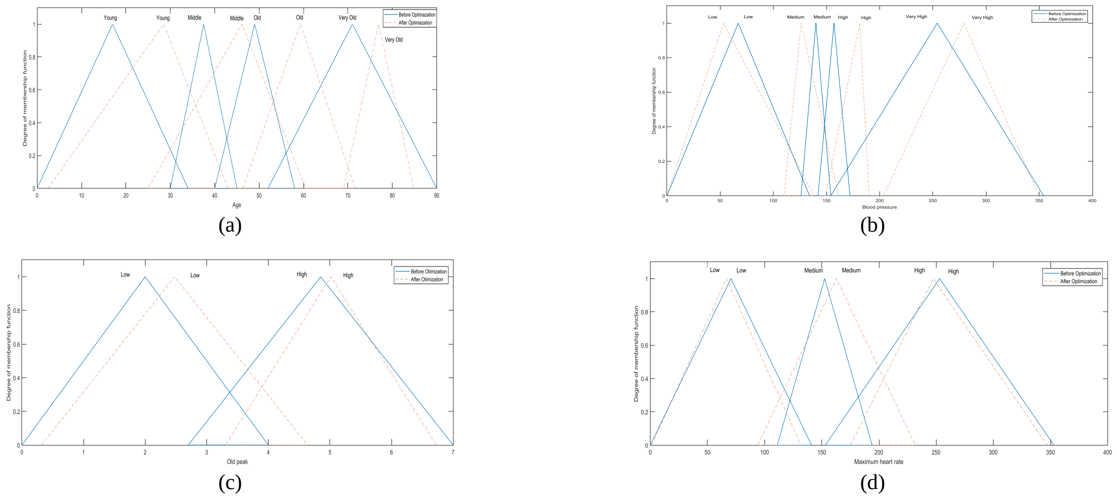



**Figure 6.** The membership functions of variables: a) Age, b) Resting blood pressure, c) Old peak, d) Maximum heart rate before (-) and after optimization (---) in binary problem.

**TABLE 3.** The results of evaluating the proposed method

| Data sets | Evaluation criterion | Recognizing healthy and unhealthy people | Multi-class CAD detection |
|---|---|---|---|
| | Sensitivity= (TP/(TP+FN)) | 93.40% | 90.02% |
| Cleveland dataset | Specificity= (TN/(FP+TN)) | 93.45% | 82.31% |
| | Accuracy= (TP/ (TP+FN+FP+TN)) | **93.48%** | **85.76%** |
| | Sensitivity | 78.45% | 72.67% |
| Hungarian dataset | Specificity | 92.33% | 83.54% |
| | Accuracy | **89.4%** | **81.28%** |
| | Sensitivity | 74.93% | 76.78% |
| Switzerland dataset | Specificity | 86.64% | 71.49% |
| | Accuracy | **81.41%** | **74.12%** |

**TABLE 4.** The comparison of the proposed method with other researches

| work | Method | Accuracy % |
|---|---|---|
| [9] | Combining artificial algorithms and KNN | 87 |
| [10] | Fuzzy system + Genetic algorithm | 76.67 |
| [11] | Automatic complex wavelet transform and type-2 fuzzy logic | 78.78 |
| [12] | Standard additive model + Genetic algorithm | 77.48 |
| [13] | rule-based decision support system | 62 |
| [14] | Neuro-fuzzy + conjugate gradient algorithm | 84 |
| [15] | C4.5 algorithm+ fuzzy expert system | 81.82 |
| [17] | Neural network + GA algorithm | 88.25 |
| [18] | Fuzzy system + particle swarm optimization and dynamic multi-swarm (MDMS-PSO) | 92.1 |
| The proposed method | Fuzzy logic optimized by cuckoo search algorithm | **93.48** |
| The proposed method | Fuzzy logic optimized by cuckoo search algorithm (Multi-class problem) | **85.26** |

## 4. CONCLUSION

Coronary artery disease is one of the most common heart diseases which many people lose their lives due to heart failer. Using computer aided systems have become common to diagnose this disease because the real methods such as angiography are difficult and do not usually result accurately. In this paper, we could predict the disease with 93.48% accuracy by fuzzy logic and cuckoo search algorithm. Also, the risk of CAD is measured by the proposed method with 85.26% accuracy on Cleveland data set. Comparing the proposed method with other researches reveals the superiority of it.

## 5. ACKNOWLEDGMENT

The corresponding author was supported by Kosar University of Bojnord with the grant number (9704121468).

## 6. REFERENCES

1. Wong, N.D., "Epidemiological studies of CHD and the evolution of preventive cardiology", ***Nature Reviews Cardiology***, Vol. 11, No. 5, (2014), 276–289.
2. Buchan, K., Filannino, M., and Uzuner, Ö., "Automatic prediction of coronary artery disease from clinical narratives", ***Journal of biomedical informatics***, Vol. 72, (2017), 23–32.
3. Davari Dolatabadi, A., Khadem, S.E.Z., and Asl, B.M., "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM", ***Computer Methods and Programs in Biomedicine***, Vol. 138, , (2017), 117–126.
4. Ramezankhani, A., Bagherzadeh-Khiabani, F., Khalili, D., Azizi, F., and Hadaegh, F., "A new look at risk patterns related to coronary heart disease incidence using survival tree analysis: 12 Years Longitudinal Study", ***Scientific Reports***, Vol. 7, No. 3237, (2017), 1-11.
5. Cagle, S.D., and Cooperstein, N., "Coronary Artery Disease: Diagnosis and Management", ***Primary care***, Vol. 45, No. 1, (2018), 45–61.
6. Shouman, M., Turner, T., and Stocker, R., "Using decision tree for diagnosing heart disease patients", In Proceedings of the Ninth Australasian Data Mining Conference - Volume 121, Australian Computer Society, (2011), 23–30.
7. Leach, H.J., O'Connor, D.P., Simpson, R.J., Rifai, H.S., Mama, S.K., and Lee, R.E., "An exploratory decision tree analysis to predict cardiovascular disease risk in African American women", ***Health Psychology***, Vol. 35, No. 4, (2016), 397–402.
8. El-Bialy, R., Salamay, M.A., Karam, O.H., and Khalifa, M.E., "Feature Analysis of Coronary Artery Heart Disease Data Sets", ***Procedia Computer Science***, Vol. 65, (2015), 459–468.
9. Polat, K., Şahan, S., and Güneş, S., "Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing", ***Expert Systems with Applications***, Vol. 32, No. 2, (2007), 625–631.
10. Dennis, B., and Muthukrishnan, S., "AGFS: Adaptive Genetic Fuzzy System for medical data classification", ***Applied Soft Computing***, Vol. 25, (2014), 242–252.
11. Nguyen, T., Khosravi, A., Creighton, D., and Nahavandi, S., "Medical data classification using interval type-2 fuzzy logic system and wavelets", ***Applied Soft Computing***, Vol. 30, No. C, (2015), 812–822.
12. Nguyen, T., Khosravi, A., Creighton, D., and Nahavandi, S., "Classification of healthcare data using genetic fuzzy logic system and wavelets", ***Expert Systems with Applications***, Vol. 42, No. 4, (2015), 2184–2197.
13. Anooj, P.K., "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", ***Journal of King Saud University - Computer and Information Sciences***, Vol. 24, No. 1, (2012), 27–40.
14. Marateb, H.R., and Goudarzi, S., "A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system", ***Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences***, Vol. 20, No. 3, (2015), 214–223.
15. Baihaqi, W.M., Setiawan, N.A., and Ardiyanto, I., "Rule extraction for fuzzy expert system to diagnose Coronary artery disease", In 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, (2016), 136–141.
16. Safdarian, N., Dabanloo, N., and Attarodi, G., "A New Pattern Recognition Method for Detection and Localization of Myocardial Infarction Using T-Wave Integral and Total Integral as Extracted Features from One Cycle of ECG Signal", ***Journal of Biomedical Science and Engineering***, Vol. 07, No. 10, (2014), 818–824.
17. Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., and Yarifard, A.A., "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm", ***Computer Methods and Programs in Biomedicine***, Vol. 141, (2017), 19–26.
18. Paul, A.K., Shill, P.C., Rabin, M.R.I., and Murase, K., "Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease", ***Applied Intelligence***, Vol. 48, No. 7, (2018), 1739–1756.
19. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Sandhu, S., and Guppy, K., "International application of a new probability algorithm for the diagnosis of coronary artery disease", ***The American journal of cardiology***, Vol. 64, No. 5, (1989), 304–310.
20. Schafer, J.L., and Graham, J.W., "Missing Data: Our View of the State of the Art", ***Psychological Methods***, Vol. 07, No. 2, (2002), 147–177.
21. Patil, B.M., Joshi, R.C., and Toshniwal, D., "Missing Value Imputation Based on K-Mean Clustering with Weighted Distance", In International Conference on Contemporary Computing, Springer, Berlin, Heidelberg, (2010), 600–609.
22. Rokach, L., and Maimon, O., Data mining with decision trees: theory and applications, Vol. 69. World scientific, (2008).
23. Wang, W., Li, Q., Han, S., and Lin, H., "A Preliminary Study on Constructing Decision Tree with Gene Expression Programming", In First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), IEEE, (2006), 222–225.
24. Mahdizadeh, M., and Eftekhari, M., "A Novel Cost Sensitive Imbalanced Classification Method based on New Hybrid Fuzzy Cost Assigning Approaches, Fuzzy Clustering and Evolutionary Algorithms", ***International Journal of Engineering - Transactions B: Applications***, Vol. 28, No. 8, (2015), 1160–1168.
25. Moaref, A., and Sattari-Naeini, V., "Fuzzy-rough Information Gain Ratio Approach to Filter-wrapper Feature Selection",

***International Journal of Engineering - Transactions C: Aspects***, Vol. 30, No. 9, (2017), 1326–1333.

26. Takagi, T., and Sugeno, M., "Fuzzy identification of systems and its applications to modeling and control", ***IEEE Transactions on Systems, Man, and Cybernetics***, Vol. SMC-15, No. 1, (1985), 116–132.
27. Yulianto, T., Komariyah, S., and Ulfaniyah, N., "Application of fuzzy inference system by Sugeno method on estimating of salt production", In AIP Conference Proceedings, AIP Publishing LLC, (2017).
28. Muthukaruppan, S., and Er, M.J., "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease", ***Expert Systems with Applications***, Vol. 39, No. 14, (2012), 11657–11665.
29. Bazoobandi, H.A., and Eftekhari, M., "A Differential Evolution and Spatial Distribution based Local Search for Training Fuzzy Wavelet Neural Network", ***International Journal of Engineering - Transactions B: Applications***, Vol. 27, No. 8, (2014), 1185–1194.
30. Yang, X.-S., and Suash Deb, S., "Cuckoo Search via Lévy flights", In World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE, (2009), 210–214.
31. Payne, R., and Sorensen, M., The cuckoos, Vol. 15. Oxford University Press, (2005).
32. Yang, X.-S., Nature-inspired metaheuristic algorithms, Luniver Press, (2008).
33. Yang, X.-S., and Deb, S., "Engineering Optimisation by Cuckoo Search", ***International Journal of Mathematical Modelling and Numerical Optimisation***, Vol. 01, No. 4, (2010), 330–343.
34. Samadiani, N., and Moameri, S., "Diagnosis of Coronary Artery Disease using Cuckoo Search and genetic algorithm in single photon emision computed tomography images", In 7th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE, (2017), 314–318.
35. Yildiz, A.R., "Cuckoo search algorithm for the selection of optimal machining parameters in milling operations", ***The International Journal of Advanced Manufacturing Technology***, Vol. 64, No. 1–4, (2013), 55–61.

# Diagnosis of Coronary Artery Disease via a Novel Fuzzy Expert System Optimized by Cuckoo Search

S. Moameri[a], N. Samadinai[b]

[a] *Department of Computer Engineering, University of Bojnord, Bojnord, Iran*
[b] *Department of Computer Engineering, Kosar University of Bojnord, Bojnord, Iran*

*P A P E R  I N F O*

*Paper history:*
Received 08 June 2018
Received in revised form 26 July 2018
Accepted 17 August 2018

*Keywords*:
*Coronary Artery Disease*
*Fuzzy System*
*Cuckoo Search*
*Heart Disease*
*Decision Tree*

**چکیده**

در این مقاله، یک سیستم خبره فازی جدید برای تشخیص بیماری عروق کرونر با استفاده از الگوریتم تکاملی فاخته (CS) پیشنهاد شده است. این سیستم در سه مرحله این بیماری را تشخیص می‌دهد: ابتدا در مرحله‌ی طراحی سیستم فازی، یک درخت تصمیم برای استخراج قوانین اگر-آنگاه مورد نیاز برای پیش‌بینی بیماری عروق کرونر، استفاده می‌شود. سپس در مرحله‌ی دوم، سیستم فازی با مقداردهی اولیه‌ی بازه‌های مقادیر فازی و قوانین مستخرج از مرحله‌ی قبل، ساخته می‌شود. در پایان، الگوریتم فاخته برای بهینه کردن توابع عضویت فازی مورد استفاده قرار می‌گیرد. دقت سیستم پیشنهادی با رکوردهای پایگاه داده کلیولند ارزیابی می‌شود. سیستم پیشنهادی قادر است با دقت ۹۳/۴۸٪، بیماری عروق کرونر را تشخیص دهد. همچنین، دقت ۸۵/۷۶٪ برای پیش‌بینی میزان ریسک این بیماری به دست آمد. برتری سیستم پیشنهادی با پیاده‌سازی ساده‌تر و دقت بیشتر در مقایسه با سایر پژوهش‌های موجود در این حوزه، کاملاً آشکار است.

***doi****: 10.5829/ije.2018.31.12c.06*