

## Research Article

# Coronary Artery Disease Detection Using a Fuzzy-Boosting PSO Approach

**N. Ghadiri Hedeshi and M. Saniee Abadeh**

*Faculty of Electrical and Computer Engineering, Tarbiat Modares University, P.O. Box 14115-143, Tehran, Iran*

Correspondence should be addressed to M. Saniee Abadeh; [saniee@modares.ac.ir](mailto:saniee@modares.ac.ir)

Received 23 October 2013; Revised 22 February 2014; Accepted 16 March 2014; Published 10 April 2014

Academic Editor: Saeid Sanei

Copyright © 2014 N. Ghadiri Hedeshi and M. Saniee Abadeh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the past decades, medical data mining has become a popular data mining subject. Researchers have proposed several tools and various methodologies for developing effective medical expert systems. Diagnosing heart diseases is one of the important topics and many researchers have tried to develop intelligent medical expert systems to help the physicians. In this paper, we propose the use of PSO algorithm with a boosting approach to extract rules for recognizing the presence or absence of coronary artery disease in a patient. The weight of training examples that are classified properly by the new rules is reduced by a boosting mechanism. Therefore, in the next rule generation cycle, the focus is on those fuzzy rules that account for the currently misclassified or uncovered instances. We have used coronary artery disease data sets taken from University of California Irvine, (UCI), to evaluate our new classification approach. Results show that the proposed method can detect the coronary artery disease with an acceptable accuracy. Also, the discovered rules have significant interpretability as well.

## 1. Introduction

Accumulation of atherosclerotic plaques in coronary arteries leads to the coronary artery disease (CAD) which results in clogging of the coronary lumen, and, consequently, occlusion, and then leads to myocardial infarction (MI) or sudden cardiac death. The CAD is the leading cause of death in the United States. Understanding the pathophysiology of coronary artery disease, the prevention of its progression, the identification and efficient modification of cardiovascular risk factors, its diagnosis, and remedy in early and reversible phases is of great significance [1]. The “gold standard” method for the diagnosis of CAD, which is widely used, is coronary angiography (CA). However, CA is a costly and invasive procedure and needs technology and high-level technical experience; therefore, it cannot be used to screen large populations or close followup of treatment [2]. Hence, in the clinical setting, for the detection of CAD, other noninvasive methods are being used. The most important of those include exercise electrocardiogram (ECG) [3] testing, single photon emission computed tomography (SPECT or scintigraphy)

[4], and stress echocardiography (ECHO), while multislice spiral computerized tomography (MSCT) or electron-beam computerized tomography (EBCT) and coronary magnetic resonance angiography (CMRA) are also being now used [2].

While many people with heart disease have symptoms such as angina, fatigue, and chest pain, many people have no symptoms until a heart attack happens. According to the American Heart Association (AHA), CAD is one of the most important killers of American men and women, reported as the cause for more than one of every five deaths in 2001 [5].

There are many risk factors related to CAD. Some factors such as family history, gender, and age cannot be controlled. However, other risk factors that are associated with lifestyle can often be controlled [6]. For example, physical inactivity, high cholesterol, high blood pressure, and smoking are all considered as risk factors for this disease that can be modified and even, in some cases, eliminated by modifying daily life routines and taking medication. Early changes in lifestyle can significantly prevent diabetes and obesity. The large number of factors that have to be analyzed for diagnosing CAD makes the physician’s work even more difficult. In general,

physicians make decisions by evaluating the existing test results of the patients. The earlier diagnoses made on other patients with the same condition are also considered by the physicians. These complicated procedures are not easy to perform when considering many factors that the physician has to evaluate. So, the decision about presence or absence of the disease depends on the physician's experience and skill to compare his patient with his previous ones. This procedure is a challenging task regarding the large number of factors that has to be considered. In this complex stage, the doctor may need an accurate tool that lists his earlier decisions about patient having the same (or close to same) factors [7].

During the past decades, the level of interest in the use of data mining and artificial intelligent tools in medical fields and the provision of healthcare has undergone a significant increase. Several sections of the researches in this area are related to developing the diagnostic tools that are used to help physicians in a diagnosis. As an advanced data mining technique, PSO has been applied to many tasks in medicine.

In PSO, particles are available to be adjusted by the learning process. In the research area of rule extraction and pattern recognition, this approach has been widely used. In this paper, we have applied PSO and fuzzy logic with a proposed boosting algorithm for the diagnosis of coronary artery disease.

The boosting mechanism adapts the distribution of training instances in a way that the previously misclassified or uncovered instances are further considered by the PSO algorithm.

The Cleveland, Hungarian, Switzerland, and VA Long Beach data sets, which are taken from Data Mining Repository of University of California, Irvine (UCI), have been used for testing this method [8].

The results show that this method can classify these data sets with acceptable accuracy or even better than the results achieved by previous works. This method is also superior to other methods in terms of interpretability.

The rest of this paper is organized as follows: a brief overview on related work, data set description, fuzzy rule-based classification system, particle swarm optimization, and ensemble based methods, is presented in the Background section. The proposed method involving a new boosting algorithm, fuzzy rule extraction with PSO, and, finally, En-PSO approach is discussed in Section 3. Experimental results are reported in Section 4, and the paper ends by some concluding remarks in Section 5.

## 2. Background

*2.1. A Brief Overview on Related Works.* Up to now, various researches have been done for the diagnosis of heart disease. These researches have used different methods for the detection of heart disease and have achieved relatively high accuracies, of 77% or higher on UCI machine learning repository data sets. Some examples are presented here.

Detrano et al. [9] applied a logistic-regression-derived discriminant function and achieved a correct classification accuracy of approximately 77%.

The accuracy achieved by John Gennari's CLASSIT conceptual clustering system was 78.9% on the Cleveland data set [10].

Gamboa et al. [11] proposed a fuzzy support vector clustering system to diagnose heart disease. In this method, a kernel induced metric was used to assign each piece of data. Experimental results were achieved with a well-known benchmark of heart disease.

Using SAS base software 9.1.3, an ensemble method with three neural networks was introduced by Das et al. [12]. The classification accuracy obtained by this method was 89.01% on Cleveland well-known data set.

Chau et al. [13] proposed the use of decision tree C4.5 algorithm, bagging with decision tree C4.5 algorithm, and bagging with Naïve Bayes algorithm to identify the heart disease of a patient and compare the effectiveness and correction rate among them. Results showed that bagging algorithms, especially bagging with Naïve Bayes, have the best performance among the tested methods.

In Rani's study [14], heart diseases data set is analyzed using neural network approach. To increase the efficiency of the classification process, the parallel approach is also adopted in the training phase. The experimental results showed that neural networks technique provides satisfactory results for the classification task.

A data set which has been called Z-Alizadeh Sani is introduced in [15]. This data set contains 303 patients and 54 features and introduces several effective features.

Acharya et al. [16] employed grayscale features from left ventricle echocardiographic images to classify patients with coronary artery disease.

*Data Set Description.* Coronary artery disease data sets are taken from Data Mining Repository of University of California, Irvine (UCI). The CAD data sets contain 920 instances collected from Cleveland, Hungarian, VA Long Beach, and Switzerland. Coronary angiography determines the result of CAD diagnosis. These data sets have 14 attributes of CAD data. These attributes are listed in Table 1. Following is a brief description of each of these data sets.

*Cleveland Data.* Cleveland data set was collected by Robert Detrano, M.D. and Ph.D. degrees holder at V.A. Medical Centre. All the papers are related to the use of a subset of 14 of the 76 features that are presented in the processed Cleveland Heart Disease data set. The "end" field indicates the existence of coronary artery disease in the patients. This field includes an integer constant that can take any value from 0 to 4. Value 0 is for nonexistence and values 1, 2, 3, and 4 are for disease existence. In fact, these values indicate the number of blocked vessels. Six of the examples have missing values. Class distributions are 54% heart disease absent and 46% heart disease present.

*2.1.1. Hungarian Data.* Andras Janosi, M.D. degree holder, collected this data set at the Hungarian Institute of Cardiology, Budapest. The format of this data set is the same as that of the Cleveland data. Class distributions are 37.5% heart disease present and 62.5% heart disease absent.

TABLE 1: Summary of attributes (UCI heart disease data base).

Attribute	Description	Value description
age	Age	Numerical
sex	Sex	1: if male; 0: if female
cp	Chest pain type	1: typical angina 2: atypical angina 3: nonanginal pain 4: asymptomatic
trestbps	Resting systolic blood pressure on admission to the hospital (mmHg)	Numerical
chol	Serum cholesterol (mg/dL)	Numerical
fbs	Fasting blood sugar over 120 mg/dL?	1: if yes 0: if no 0; normal
restecg	Resting electrocardiographic results	1: having ST-T wave abnormality 2: having LV hypertrophy
thalach	Maximum heart rate achieved	Numerical
exang	Exercise induced angina?	1: if yes 0: if no
oldpeak	ST depression induced by exercise relative to rest	Numerical
slope	The slope of the peak exercise ST segment	1: upsloping 2: flat 3: downsloping
ca	Number of major vessels colored by fluoroscopy	Numerical 3: normal
thal	Exercise thallium scintigraphic defects	6: fixed defect 7: reversible defect
num	Diagnosis of heart disease (angiographic disease status/presence of coronary artery disease (CAD))	0: if less than 50% diameter narrowing in any major vessel (CAD, no) 1: if more than 50% (CAD, yes)

*2.1.2. Switzerland Data.* This data set was collected at the University Hospital, Zurich, Switzerland, by William Steinbrunn, M.D. degree holder. Among four data sets related to CAD, the maximum number of missing value is related to Switzerland data set. It has 123 instances and class distributions in it are 6.5% heart disease absent and 93.5% with heart disease.

*2.1.3. VA Long Beach.* VA Long Beach data set which has 200 instances was collected by Matthias Pfisterer, M.D. degree holder, at the University Hospital, Basel, Switzerland. This data set is in second rank in terms of the number of missing values. Class distributions in it are 25.5% heart disease absent and 74.5% heart disease present.

*2.2. Fuzzy Rule-Based Classification System.* Assume that we have a classification problem with  $c$  classes in the  $n$ -dimensional space with continuous attributes. Also, assume that  $M$  real vectors  $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$ ,  $p = 1, 2, \dots, m$ , are given as training samples from the  $c$  classes ( $c \ll M$ ).

The pattern space is  $[0, 1]^n$  and attribute values of each pattern are  $x_{pi} \in [0, 1]$  for  $p = 1, 2, \dots, m$  and  $i = 1, 2, \dots, n$ . All the attribute values of each sample are in  $[0, 1]$ .

Fuzzy if-then rules, in this method, are expressed with the following form:

rule  $R_j$ : if  $x_1$  is  $A_{j1}$  and  $\dots$  and  $x_n$  is  $A_{jn}$ , then class  $C_j$  with  $CF = CF_j$ .

Here,  $R_j$  is the label of the  $j$ th fuzzy if-then rule,  $A_{j1}, \dots, A_{jn}$  are antecedent fuzzy sets in the range of  $[0, 1]$ ,  $C_j$  is the result class (i.e., one of the given  $c$  classes), and  $CF_j$  is the degree of confidence of the fuzzy if-then rule  $R_j$ .

In previous work [18], we used one uniform fuzzy partition for all attributes, while we know that these attributes may be of different kinds (e.g., continues, ordinal, and ratio). Therefore, it is more appropriate to apply several fuzzy partitions for various types of attributes. For example, in a binary attribute, only two fuzzy amounts are required, while, for a continuous attribute, more fuzzy amounts (like small, medium small, medium, medium large, and large) must be used.

In computer implementations, we have used four fuzzy partitions evenly separated with symmetric triangular fuzzy sets in Figure 1.

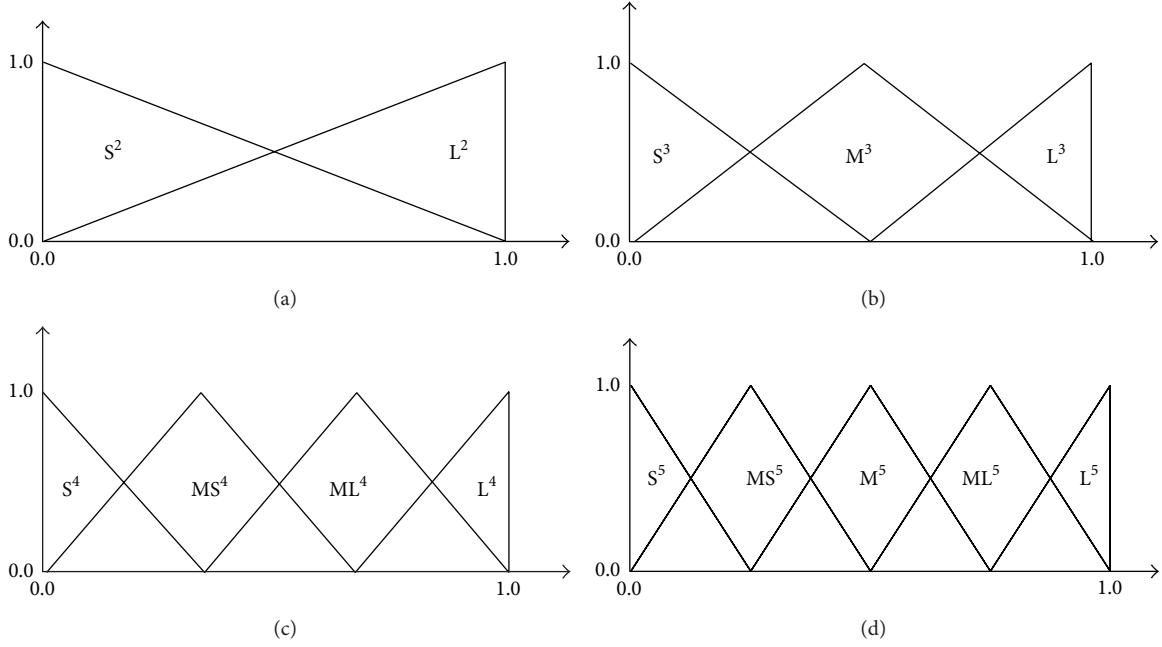


FIGURE 1: Four fuzzy partitions used in computer implementations. The superscript of each part shows the granularity of the fuzzy partition.

To illustrate the high performance of our fuzzy classifier system, we use such simple specification in computer implementations.

However, we can use any adapted membership functions in our fuzzy classifier system for a specific pattern classification problem. When a rule is extracted with PSO, the following steps are applied to calculate the level of confidence of each fuzzy if-then rule.

*Step 1.* The compatibility of the fuzzy if-then rule  $R_j$  with each training instance  $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$  is obtained as follows:

$$\mu_j(x_p) = \mu_j(x_{p1}) \times \dots \times \mu_{jn}(x_{pn}), \quad p = 1, 2, \dots, m, \quad (1)$$

where  $\mu_{ji}(x_{pi})$  is the membership function of  $i$ th attribute of  $p$ th instance and “ $m$ ” indicates the total number of instances.

*Step 2.* For each of the classes, calculate the relative sum of the compatibility grades of the training instances with the fuzzy if-then rule  $R_j$ :

$$\beta_{\text{Class } h}(R_j) = \sum_{x_p \in \text{Class } h} \frac{\mu_j(x_p)}{N_{\text{Class } h}}, \quad h = 1, 2, \dots, c, \quad (2)$$

where  $\beta_{\text{Class } h}(R_j)$  denotes the sum of the compatibility grades of the training instances in Class  $h$  with the fuzzy if-then rule  $R_j$  and  $N_{\text{Class } h}$  is the number of training instances whose corresponding class is Class  $h$ .

*Step 3.* The certainty factor  $CF_j$  is calculated as follows:

$$CF_j = \frac{\left[ \beta_{\text{Class } \hat{h}_j}(R_j) - \left( \sum_{h \neq \hat{h}_j} \beta_{\text{Class } h}(R_j) / (c - 1) \right) \right]}{\sum_{h=1}^c \beta_{\text{Class } h}(R_j)}. \quad (3)$$

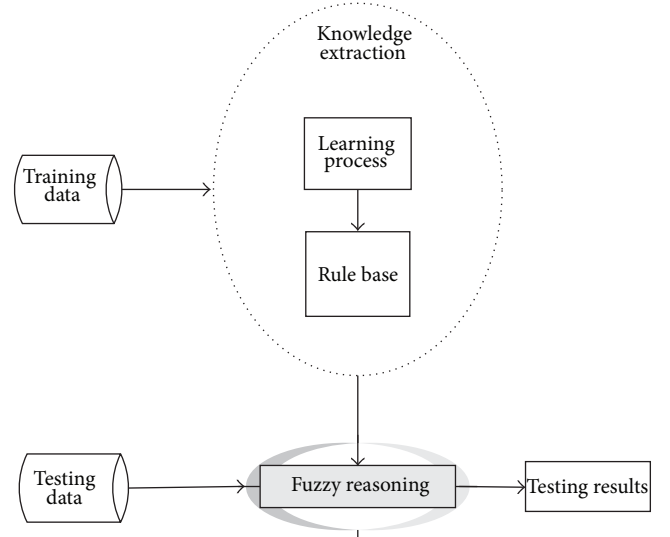


FIGURE 2: The basic model of rule-based fuzzy classifiers [17].

Now, we can determine the certainty factor for all combinations of antecedent fuzzy sets.

A fuzzy classification system (FCS) includes two basic steps [17]: (1) knowledge extraction which contains a set of fuzzy rules as extracted knowledge and (2) inference engine which classifies the input samples according to fuzzy rule set and reasoning method. Figure 2 presents the basic model of rule-based fuzzy classification systems.

The goal of our fuzzy classification system (FCS) is to produce combinations of antecedent fuzzy sets for extracting

a rule set  $S$  with more power to classify. With a rule set  $S$ , an input instance  $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$  is classified by a single winner rule  $R_j$  in  $S$ , which is determined by the following equation:

$$\mu_j(x_p) \cdot CF_j = \max \{ \mu_j(x_p) \cdot CF_j R_j \}. \quad (4)$$

Thus, the winner rule has the maximum product of the compatibility and the certainty factor  $CF_j$ .

Each fuzzy rule is encoded into a specific string. Fifteen symbols are used for representing the 14 antecedent fuzzy sets in Figure 1 plus *don't care*. These symbols are as follows:

0 = do not care (DC), 1 = small ( $S^2$ ), 2 = large ( $L^2$ ), 3 = small ( $S^3$ ), 4 = medium ( $M^3$ ), 5 = large ( $L^3$ ), 6 = small ( $S^4$ ), 7 = medium small ( $MS^4$ ), 8 = medium large ( $ML^4$ ), 9 = large ( $L^4$ ), a = small ( $S^5$ ), b = medium small ( $MS^5$ ), c = medium ( $M^5$ ), d = medium large ( $ML^5$ ), and e = large ( $L^5$ ).

For example, fuzzy rule "0150b0  $\rightarrow$  0" denotes that the fuzzy rule "If  $x_2$  is  $S^2$  and  $x_3$  is  $L^3$  and  $x_5$  is  $MS^5$ , then class is 0."

### 2.3. Particle Swarm Optimization

**2.3.1. RAM Concept.** At first, the RAM concept, which has been used in PSO algorithm, must be explained. RAM term in this work is the same as what is defined in [19] by Saniee et al. Here, we describe it briefly.

If  $R_A$  is a fuzzy rule with  $n$  antecedents, then  $R_A$  is as follows:

$$R_A = (a_i), \quad i = 1, 2, \dots, n. \quad (5)$$

Rule antecedent modification (RAM) operator or  $RAM(k, A_j)$  is defined as follows.

RAM operator stands for substitution of the existing linguistic value of the rule  $k$ th precondition with the  $j$ th linguistic value. Thus, by applying RAM operator, a new fuzzy rule is made according to the following equation:

$$R'_A = R_A + RAM(k, A_j). \quad (6)$$

So, the symbol "+" in (6) has a new concept. For example, if we have a classification problem with five inputs, a fuzzy rule for this problem is like the following equation:

$$R_A = (S^5, ML^4, L^2, M^3, DC). \quad (7)$$

Suppose that the RAM operator is  $RAM(3, A_1)$ . The result of applying this operator is shown in the following equation:

$$R'_A = R_A + RAM(3, A_1) = (S^5, ML^4, S^2, M^3, DC). \quad (8)$$

Rule antecedent modification sequence (RAMS) is built with more than one RAM operators which are applied, respectively, as follows:

$$RAMS = (RAM_1, RAM_2, \dots, RAM_n). \quad (9)$$

Note that the order of RAM operators in RAMS is very important.

When it is said that one RAMS acts on a solution, it means that all the RAM operators of it act on the solution in order.

Operator  $\oplus$  is defined for merging two RAM sequences into a new RAM sequence.

Consider that RAMS1 and RAMS2 are two RAM sequences that operate on  $R_A$ , respectively. With this act, a new rule  $R'_A$  is obtained. Now, suppose that there is another RAMS sequence  $RAMS'$  operating on solution  $R_A$  with the same result,  $R'_A$ . Under these circumstances,  $RAMS'$  is described as follows:

$$RAMS' = RAMS1 \oplus RAMS2. \quad (10)$$

$RAMS1 \oplus RAMS2$  and  $RAMS'$  are in the same equivalent set.

In an equivalent set, the RAMS having the lowest number of RAM operators is called basic RAM sequence (BRAMS).

Consider that  $R_N$  and  $R_M$  are two solutions. Next definition is about  $BRAMS_{M \rightarrow N}$ ; that is,

$$BRAMS_{M \rightarrow N} = R_N - R_M. \quad (11)$$

$BRAMS_{M \rightarrow N}$  in previous equation is a basic RAM sequence which acts on  $R_M$  to obtain the solution  $R_N$ .

**2.3.2. PSO Algorithm.** Particle swarm optimization (PSO) is an evolutionary computation method for optimization that was developed by Kennedy and Eberhart in 1995 [20]. It is inspired by social behaviour of bird flocking or fish schooling and swarm theory.

This algorithm works by simultaneously maintaining a number of candidate solutions in the search space. Each candidate solution is considered as a particle "flying" in the  $n$ -dimensional search space to find the best solution. In all of the iterations of algorithm, each candidate solution, or, in other words, each particle, is appraised by the objective function, and the fitness of that solution is calculated. At first, like the GA, the PSO algorithm is initialized with a population of random solutions in the search space. PSO also needs only the information about the fitness values of the particles in the population. This algorithm simply calculates the fitness values of the individuals by applying the objective function. In comparison with genetic algorithm, individuals in PSO have memory so that the information of the particles with better solutions is maintained by all individuals. In other words, it creates a constructive cooperation between the individuals, and the individuals share information between themselves.

All the individuals have a location vector  $x_i$  and a velocity vector  $V_i$ . Location of each individual is updated at each iteration considering its personal best position ( $P_i$ ) and the best position among all individuals ( $P_g$ ) at each iteration.

The PSO algorithm only consists of three steps which are repeated until some stopping condition is met [21].

- (1) Calculate the fitness value of each particle.
- (2) Update local and global best positions and fitness.

- (3) Calculate the new velocity and position of each particle with regard to

$$\begin{aligned} V_i(t+1) &= \omega V_i(t) + c_1 \times (P_i(t) - x_i(t)) \\ &\quad + c_2 \times (P_g(t) - x_i(t)), \quad (12) \\ x_i(t+1) &= x_i(t) + V_i(t+1), \end{aligned}$$

where  $\omega$  is the inertia weight that is applied to control the influence of the previous history of the velocities on the new velocity.

$c_1$  and  $c_2$  are acceleration coefficients;  $r_1 \sim U(0, 1)$  and  $r_2 \sim U(0, 1)$  are random constants in the range  $[0, 1]$  and are uniformly distributed. This velocity-updating method allows the particles to search around their individual best position  $P_i$  and the global best position  $P_g$ .

**2.4. Ensemble Based Methods.** Machine learning is a set of nonparametric statistical methods. By simulating human idea process, it gives a smart structure. In comparison to other statistical models, in the field of machine learning, a model is made without any assumption. The most important concept in this field is the generalization ability. It confirms how well the result learned by a training data set can be applied to the unsighted test data set. Ensemble learning is a class model in the field of machine learning. Due to its stable and accurate performance, this method is popular for prediction and classification. In contrast with ordinary methods, in machine learning, which tries to construct only one learner from the training data, ensemble methods try to construct several learners and to combine them. In general, the generalization ability of an ensemble learner is better than a single learner.

In fact, ensemble techniques are more popular because of their ability to strengthen weak learners [22].

Bagging, Adaboost, Logitboost, Random Forest, and so forth are various ensemble techniques, whereas Random Forest and bagging act in parallel and the rest act in sequential act in parallel are boosting acts in sequential.

An ensemble technique has two steps: (1) generating individual members and (2) right combining individual members' outputs to find a new output [12].

### 3. The Proposed Methodology and Implementation

**3.1. Boosting Algorithm.** Boosting algorithm was offered by Freund and Schapire [23, 24]. It is a method of producing very accurate prediction rules by combining several "weak" learners that can only be moderately accurate. Boosting algorithm obviously modifies the distribution of the training data that has been given to each single classifier by the weights of training samples. At first, all training samples are given the same weight. During the boosting process, those weights are changed based on the error that current classifier has on the training data. The weights of those samples that are correctly classified are reduced, while the weights of those that are misclassified or not covered by the classifier remain unchanged. The weight  $w_k$  indicates the importance of the  $k$ th

instance in current training data set. When a weak classifier is run, the weights  $w_k$  of the correctly classified instances are decreased; thus, the next weak classifier pays more attention to the misclassified or not covered instances due to their higher weights.

The initial weights of all examples are the same and are equal to value 1. The PSO algorithm is repeatedly called on the current training data by the boosting algorithm. The error  $ER_{R_T}$  of the fuzzy rule  $R_T$ , produced by PSO algorithm in  $T$ th run, is calculated by the boosting algorithm. The parameter  $ER_{R_T}$  is evaluated by measurement matching  $\mu(x_k)$  which determines the measure of matching between the  $K$ th training instance and the fuzzy rule  $R_T$  antecedent as well as its weight  $w_k$ . Consider

$$ER_{R_T} = \frac{\sum_{k|C_k \neq C_i} w_k \mu(x_k)}{\sum_k w_k \mu(x_k)}. \quad (13)$$

The goal of this intelligent system is to find classification rules that classify accurately the current distribution of the training data.

In normal form of Adaboost algorithm, only for the correctly classified instances, the weights are reduced. The weight of a correctly classified instance is decreased by factor  $\alpha^k$ . Consider

$$w_k(t+1) = \begin{cases} w_k(t), & C_i \neq C_k \\ w_k(t) \times \alpha^k, & C_i = C_k, \end{cases} \quad (14)$$

$$\alpha^k = \left( \frac{1}{1 + \text{EXP}(\mu(x_k)/(ER_{R_T} + w_k))} \right)^{\mu(x_k)}. \quad (15)$$

The factor  $\alpha^k$  is obtained based on three parameters  $\mu(x_k)$ ,  $ER_{R_T}$ , and the weight  $w_k$ . According to (14) and (15), as the instance and the rule antecedent are more matched, the weight of the instance is more reduced. Also, the weights of the instances, which are correctly classified by a rule with less error or are better known in previous iteration, are more reduced. The weights of the misclassified or uncovered instances have been left unchanged.

**3.2. Fuzzy Rule Extraction with PSO Algorithm.** PSO algorithm has been used in various data mining problems such as clustering and classification [25–27]. In this paper, PSO algorithm extracts appropriate rules for a classification problem.

Population  $P$  with  $L$  particles is defined as follows [28]:

$$P = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_h \\ \vdots \\ p_L \end{bmatrix} = \begin{bmatrix} r_1 & g_1 \\ r_2 & g_2 \\ \vdots & \vdots \\ r_h & g_h \\ \vdots & \vdots \\ r_L & g_L \end{bmatrix}, \quad (16)$$

where  $p_h = [r_h \ g_h]$  is a particle of population that defines a set of fuzzy rules.  $r_h = [r_1^h \ r_2^h \ r_3^h \ \dots \ r_j^h \ \dots \ r_B^h]$

```

While ( $\sum_{\text{for all training instances}} \text{weight}(\text{instance}) > k$ ) {
  (1) Run the PSO algorithm
  (2) For each rule exist in best solution in PSO algorithm do {
    Reweight (instances) according to new rule has acquired by PSO algorithm:
    (a) Calculate the error  $ER_{R_T}$  of the current Fuzzy Rule according to (14).
    (b) Calculate factor  $\alpha^k$  according to (16).
    (c) For each of training instances in heart data set calculate new weight
    (d)  $w_k(T + 1)$  according to (15).
  } // end of For
} // end of while loop

```

PSEUDOCODE 1: The pseudocode of En-PSO2 approach.

includes a set of candidate fuzzy rules. It defines the position of a particle where  $B$  is a positive integer variable to set the maximum number of fuzzy rules that can be produced. The velocity vector defines as  $V_h = [v_1^h \ v_2^h \ v_3^h \ \dots \ v_j^h \ \dots \ v_B^h]$  that  $v_j^h$  is a BRAMS. The parameter vector  $g_h = [g_1^h \ g_2^h \ \dots \ g_j^h \ \dots \ g_B^h]$  is used in order to reduce the number of fuzzy rules. If the parameter  $g_j^h$  is less than or equal to a user-defined maximum threshold, the rule  $r_j^h$  remains in fuzzy rule set; otherwise, it is eliminated. In initialization step, this parameter is random in  $[0, 1]$ ; then, it is calculated according to (17) in each iteration. Consider

$$g_j^h = \frac{NMP_{r_j^h}(w)}{NCP_{r_j^h}(w)}, \quad (17)$$

$$NMP_{r_j^h}(w) = \sum_{k|c_k \neq C_j} w_k, \quad (18)$$

$$NCP_{r_j^h}(w) = \sum_{k|c_k = C_j} w_k. \quad (19)$$

In fact,  $NMP_{r_j^h}(w)$  is the total weight of the instances that have been incorrectly diagnosed and  $NCP_{r_j^h}(w)$  is the total weight of the instances that have been correctly diagnosed by the rule  $r_j^h$ .

Suppose that  $r_h$  is the number of acceptable fuzzy rules; then, index of these rules is  $I_r \in \{1, 2, \dots, B\}$ ,  $r = 1, 2, \dots, r_h$ , and the acceptable fuzzy rule set is as  $\{r_{I_1}^h, r_{I_2}^h, \dots, r_{I_r}^h, \dots, r_{I_{r_h}}^h\}$ .

The position of a particle in iteration  $T$  is considered as  $r_h(T)$ ; in this case, its next position will be as follows:

$$\begin{aligned} r_h(T+1) &= r_h(T) + V_h(T) \\ &= \left[ r_{I_1}^h + v_{I_1}^h, r_{I_2}^h + v_{I_2}^h, \dots, r_{I_r}^h + v_{I_r}^h, \dots, r_{I_{r_h}}^h + v_{I_{r_h}}^h \right]. \end{aligned} \quad (20)$$

The next velocity of each particle is calculated by the following equation:

$$\begin{aligned} V_h(T+1) &= \omega \times V_h(T) \oplus \alpha \times (L_h(T) - r_h(T)) \\ &\oplus \beta \times (L_g(T) - r_h(T)). \end{aligned} \quad (21)$$

$L_h(T)$  is the best ever existing position experienced by the particle and is called local best position. Also,  $L_g(T)$  is called global best position and is the best position ever experienced by all the particles.

**3.3. The En-PSO2 Approach.** In a number of researches, combinations of nature-inspired algorithms and ensemble methods have been used for rule extraction. For example, Ant-Miner method can be expressed. In Ant-Miner, a series covering method is followed to find a set of classification rules covering all, or approximately all, the training instances [29]. Another example is the study that is referred to in [30] in which a boosting algorithm is combined with genetic algorithm to extract classification rules.

In this section, we will describe the details of En-PSO2 approach for generating the classification fuzzy rules for detecting coronary artery disease.

In this method, the PSO algorithm that has been described in Sections 2 and 3.2 runs several times. In previous works, in En-PSO algorithm, each time that PSO algorithm was run, only one fuzzy rule was added to the rule set, but, in En-PSO2, each time that PSO is run, several fuzzy classification rules are extracted and added to the fuzzy rule set. Maximum number of the fuzzy rules that may be extracted in each run is equal to parameter  $B$  (defined in previous section). The new boosting method, described in Section 3.1, considers the collaboration between the fuzzy classification rules that are extracted from PSO algorithm. After a rule is added to rule set, the weights of those instances that have been correctly classified with a recently extracted rule are reduced by boosting algorithm and, for the remaining instances, the weight will not be changed to increase their chances in next runs. Thus, the En-PSO2 is biased to find those fuzzy classification rules that complete the existing fuzzy rule set and rectify its shortages. The pseudocode

TABLE 2: Parameter setting for the proposed En-PSO2.

Parameter	Value
Swarm size	25
Number of iteration	Until the global best does not change after 5 iterations
$\omega$	1/iteration number
$c_1$	Random (between 0 and 1)
$c_2$	Random (between 0 and 1)
$K$	12
$B$	20

of En-PSO approach is shown in Pseudocode 1. In this pseudocode,  $K$  is the allowable amount for total weights of training examples. Figure 3 presents a flowchart showing how En-PSO2 algorithm works.

#### 4. Experimental Result

In this research, we had two classes: 1 and 0; 1 refers to the healthy people and 0 is for patients who are subject to possible CAD. To test this method, we used UCI coronary artery data set which was described in Section 2.

This data set has 920 instances: 509 instances with coronary artery disease and 411 instances without coronary artery disease.

We have run En-PSO2 method with various parameters and at last we set them as in Table 2.

To achieve more reliable results, in partitioning training and test sets,  $K$ -fold cross-validation technique was used. In the mentioned method, the main data set is randomly divided into  $K$  partitions. Here, the classification method is run ten times. Each time, the next subset is considered as the validation and the remaining  $k - 1$  partitions are used as training data. In fact, in  $K$ -fold cross-validation, each of the  $K$  partitions is used once as validation data. The  $K$  results are averaged to a single estimation. Different tests on various data sets have shown that 10 is almost the best number of folds to get the best approximation of error [31].

Different evaluation criteria have been used in data mining and machine learning to test methods performances. The classification accuracy is the most common evaluation criterion used in data mining field. We have considered five criteria to evaluate the performance of the proposed method: accuracy, specificity, sensitivity, precision, and  $F$ -measure.

To calculate these measures, a well-known matrix called confusion matrix (contingency table) is formed. This matrix represents the classification results.

When the confusion matrix was constructed, the accuracy, specificity, sensitivity, precision, and  $F$ -measure can be easily calculated as the following [15]:

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})},$$

$$\text{specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})},$$

TABLE 3: The obtained confusion matrix.

		Actual result		
Patient	Healthy			
90	419	Healthy		Classifier result
370	41	Patient		

$$\text{sensitivity} = \text{recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})},$$

$$\text{precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})},$$

$$F\text{-measure} = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})}. \quad (22)$$

*True Positive (TP)*. It is the number of correct predictions with CAD that is diagnosed as patient by the angiography.

*True Negative (TN)*. It is the number of correct predictions as normal with CAD that is labelled as healthy by the angiography.

*False Negative (FN)*. It is the number of incorrect predictions as normal with CAD that is diagnosed as patient by the angiography.

*False Positive (FP)*. It is the number of incorrect predictions as a patient that is labelled as a healthy person by the angiography.

Classification accuracy is calculated by the ratio of the number of the instances correctly classified to the total number of samples.

Precision measures the percentage of the actual patients (i.e., true positive) between the patients that got declared CAD; recall measures the percentage of the actual patients that were discovered;  $F$ -measure balances between precision and recall. Specificity measures the percentage of patients without disease that can be correctly eliminated.

Table 3 shows the confusion matrix obtained by the proposed system. We achieved accuracy of 85.76 on UCI data set of coronary artery disease. All the five performance measures, accuracy, sensitivity, specificity, precision, and  $F$ -measure, for En-PSO2 have been shown in Table 4. This method has generated in average 21.2 rules. The comparison between rule-based methods is presented in Table 5. As can be seen among all the rule-based methods for detection of CAD, the minimum number of rules is allocated to En-PSO2. Also, the accuracy of this method is in second place after En-PSO. In addition, the average length of rules is 3.01. These results suggest that the interpretability of the proposed method is very high since the comprehensibility of a rule-based expert system with a few if-then rules is much more than a system with large number of rules.

In Figure 4, two factors, time and accuracy of the method, are compared with the increasing amount of initial number of rules for each particle ( $B$ ). As can be seen maximum accuracy is achieved by setting 30 for  $B$ . This negligible increase in accuracy (approximately 0.03) is obtained with a



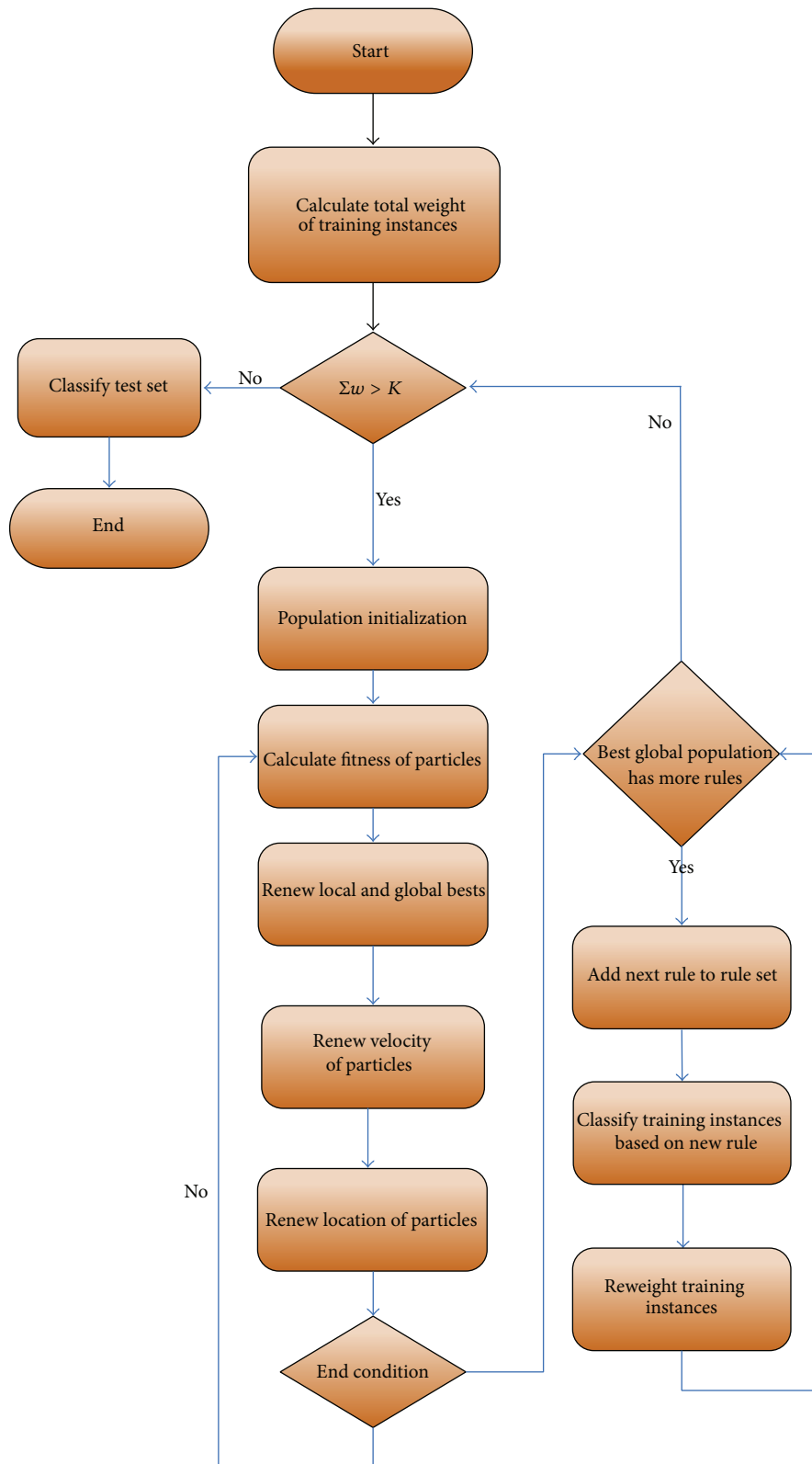


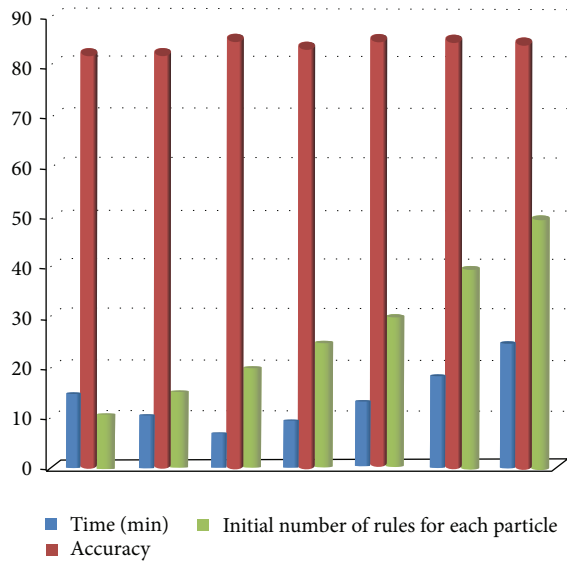
FIGURE 3: Total flowchart of the expert system of coronary artery detection.  $K$  is the allowable amount for total weights of training examples.

TABLE 4: Performance measures of En-PSO2 according to confusion matrix.

Measure name	Accuracy	Sensitivity	Specificity	$F$ -measure	Precision
Amount (%)	85.76	90.02	82.31	86.48	91.08

TABLE 5: Comparison between rule-based methods.

Rule-based methods	Accuracy	Number of rules
Decision tree [33]	85.6	83
Support based [33]	84.4	27
Pearson [34]	84.5	27
RST [33]	85.2	27
En-PSO [18]	85.97	25.3
<b>En-PSO2 (this study)</b>	<b>85.76</b>	<b>21.2</b>

FIGURE 4: Time and accuracy changes based on increasing amount of  $B$ .

time difference of about 6.5 minutes. Therefore, we have set 20 for  $B$ .

Figure 5 illustrates the average accuracy after adding each rule to rule set. Given that the method which we have chosen to test the proposed method is 10-fold cross-validation, we do not have a single rule set, and we have 10 rule sets. Therefore, in Figure 5, we have calculated the average of accuracy. As can be seen, the average accuracy on the train data set is always ascending, but the slope of graph is declining. In the earlier runs, the extracted rules cover the samples easier and reduce their chances for selection in the next runs. So, with increasing the number of runs, the rules are more specific and cover the less number of samples; thus, increasing accuracy is less.

Table 6 compares the classification accuracy, sensitivity, specificity, and number of rules between proposed method and previous methods. This comparison shows that En-PSO2, both in terms of performance and the interpretability, is better.

TABLE 6: Comparison between classification algorithms according to accuracy, sensitivity, specificity, and number of rules.

Method	Accuracy	Sensitivity	Specificity	Number of rules
Decision tree [35]	78.91	72.01	84.48	—
LTF-C [36]	81.2	—	—	—
Bagging [35]	81.41	74.93	86.64	—
$k$ -NN [33]	81.5	—	—	—
Bayesian model [37]	82	87	—	—
Decision tree (C4.5) [37]	82.5	87.17	—	—
SVM [37]	82.5	88	—	—
RST [33]	85.2	—	—	27
ANN [38]	85.53	—	—	—
Decision tree [33]	85.6	—	—	83
NN-Alizadeh [15]	85.43	90.2	73.5	—
En-PSO2 (this study)	85.76	90.02	82.31	21.2

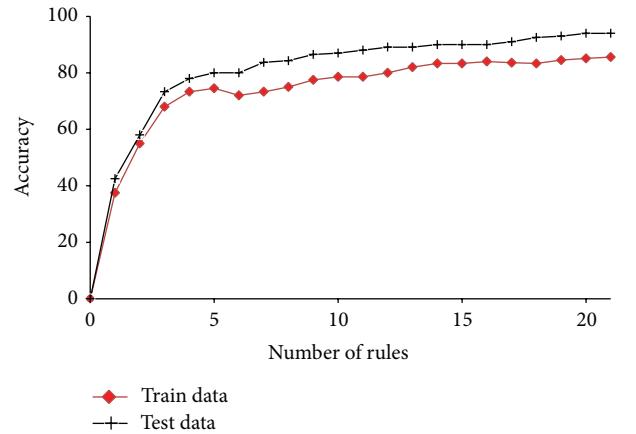


FIGURE 5: Average accuracy obtained on train and test data sets from adding each rule to the rule set.

It is important to note that the computational complexity of our method is  $O(P * R * F)$ ;  $P$  denotes number of particles;  $R$  addresses the number of patients records; and  $F$  is number of features. In comparison to other works, it seems that our algorithm is more complex and needs more time to extract its knowledge base (fuzzy if-then rule). However, since the testing time is very fast ( $O(F)$ ) and the extracted rule set is very accurate and interpretable, we can accept the final performance of this complex learning algorithm.

## 5. Conclusion

In a number of researches, diagnosis of CAD has been considered. These researches have applied different techniques to

the given problem and attained high classification accuracies on the UCI data sets. In a number of these researches, only determination of patient or normal has been considered while, in some other, the extraction of appropriate rules is also considered.

In this research, a fuzzy boosting PSO approach has been proposed to generate appropriate rules for detection of CAD. The proposed boosting algorithm helps to produce optimal rules to cover more instances.

In our previous works [18, 32], in each run, one rule was extracted and added to rule set. In this work, each individual in PSO algorithm includes a set of rules. Therefore, each time the PSO algorithm is run, several rules are added to rule set. This has two results. (1) Algorithm will be completed faster if appropriate values for the parameters are set. This is because each time PSO algorithm is run, several rules can be added to rule set and this makes the process of the instance covering faster. (2) There is more coordination between production rules. In previous method, the extracted rule in each run may be optimal in terms of covering instances but may not be coordinate between existing rules in rule set. Consequently, the accuracy of the method may be reduced. In En-PSO2, in each run, several rules are extracted which are compatible. So, in this method, rules are more coordinate.

This method, in addition to being competitive with other approaches in terms of accuracy, produces an average of 21.2 rules that is the minimum number of rules for the detection of the mentioned disease. Moreover, the average length of rules is 3.01. These results show that the interpretability of the proposed method is very high (in comparison to [18, 33, 34], whose methods produce more number of if-then rules to detect CAD).

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] G. De Backer, E. Ambrosioni, K. Borch-Johnsen et al., "European guidelines on cardiovascular disease prevention in clinical practice: third joint task force of European and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of eight societies and by invited experts)," *European Journal of Cardiovascular Prevention and Rehabilitation*, vol. 10, supplement 4, pp. S1-S10, 2003.
- [2] E. Escolar, G. Weigold, A. Fuisz, and N. J. Weissman, "New imaging techniques for diagnosing coronary artery disease," *Canadian Medical Association Journal*, vol. 174, no. 4, pp. 487-495, 2006.
- [3] L. Goldberger and E. Goldberger, *Clinical Electrocardiography*, Mosby, Saint Louis, Mo, USA, 1977.
- [4] R. J. Gibbons, G. J. Balady, J. T. Bricker et al., "ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association task force on practice guidelines (committee to update the 1997 exercise testing guidelines)," *Circulation*, vol. 106, no. 14, pp. 1883-1892, 2002.
- [5] American heart association (AHA), 2011, <http://www.americanheart.org/>.
- [6] 2011, <http://yourtotalhealth.ivillage.com/heart-disease-fast-facts.html>.
- [7] K. Polat, S. Güneş, and S. Tosun, "Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing," *Pattern Recognition*, vol. 39, no. 11, pp. 2186-2193, 2006.
- [8] C. L. Blake and C. J. Merz, "UCI Repository of Machine Learning Databases," 1996, <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [9] R. Detrano, A. Janosi, W. Steinbrunn et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *American Journal of Cardiology*, vol. 64, no. 5, pp. 304-310, 1989.
- [10] J. H. Gennari, P. Langley, and D. Fisher, "Models of incremental concept formation," *Artificial Intelligence*, vol. 40, no. 1-3, pp. 11-61, 1989.
- [11] A. L. Gamboa, M. G. Mendoza, J. M. Vargas, N. H. Gress, and R. E. Orozco, "Hybrid fuzzy-SV clustering for heart disease identification," in *Proceedings of the International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA/IAWTIC '06)*, Sydney, Australia, November 2006.
- [12] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675-7680, 2009.
- [13] M. Chau, D. Shin, and D. K. Shin, "A comparative study of medical data classification methods based on decision tree and bagging algorithms," in *Proceedings of the 8th IEEE International Symposium on Dependable, Autonomic and Secure Computing (DASC '09)*, pp. 183-187, Chengdu, China, December 2009.
- [14] K. U. Rani, "Analysis of heart disease dataset using neural network approach," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 1, no. 5, 2011.
- [15] R. Alizadehsania, J. Habibia, M. J. Hosseinia et al., "A data mining approach for diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 1, pp. 52-61, 2013.
- [16] U. R. Acharya, S. V. Sree, M. M. R. Krishnan et al., "Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 3, pp. 624-632, 2013.
- [17] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: current framework and new trends," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 5-31, 2004.
- [18] N. G. Hedeshi and M. S. Abadeh, "An ensemble PSO-based approach for diagnosis of coronary artery disease," in *Proceedings of the International Symposium on Artificial Intelligence and Signal Processing (AISP '11)*, pp. 77-82, Tehran, Iran, June 2011.
- [19] M. Saniee, J. Habibi, and S. Aliari, "Using a particle swarm optimization approach for evolutionary fuzzy rule learning: a case study of intrusion detection," in *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU '06)*, 2006.
- [20] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the 1995 IEEE International Conference on Neural Networks. Part 1 (of 6)*, pp. 1942-1948, December 1995.
- [21] F. van den Bergh, *An analysis of particle swarm optimizers [Ph.D. thesis]*, University of Pretoria, Pretoria, South Africa, 2001.

- [22] X. Wu and V. Kumar, "The top ten algorithm in data mining," International Standard Book 13: 978-1-4200-8964-6, 2009.
- [23] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Tech. Rep., AT&T Bell Laboratories, Murray Hill, NJ, USA, 1995.
- [24] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceeding of the International Conference on Machine Learning*, pp. 148–156, Bari, Italy, July 1996.
- [25] K. Srinivas, B. K. Rani, and A. Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *International Journal on Computer Science and Engineering (IJCSSE)*, vol. 2, no. 2, pp. 250–255, 2010.
- [26] A. Wang, J. Wu, X. Zhang, and R. Tao, "A novel medical image registration algorithm based on PSO and wavelet transformation combined with 2v-SVM," in *Proceedings of the 2nd International Conference on Innovative Computing, Information and Control (ICICIC '07)*, IEEE, Kumamoto, Japan, September 2007.
- [27] Y. W. Chen, C. L. Lin, and A. Mimori, "Multimodal medical image registration using particle swarm optimization," in *Proceedings of the 8th International Conference on Intelligent Systems Design and Applications (ISDA '08)*, pp. 127–131, IEEE, Kumamoto, Japan, November 2008.
- [28] C. C. Chen, "Design of PSO-based fuzzy classification systems," *Tamkang Journal of Science and Engineering*, vol. 9, no. 1, pp. 63–70, 2006.
- [29] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 321–332, 2002.
- [30] F. Hoffmann, "Combining boosting and evolutionary algorithms for learning of fuzzy classification rules," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 47–58, 2004.
- [31] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, Calif, USA, 2nd edition, 2005.
- [32] N. G. Hedeshi and M. S. Abadeh, "An expert system working upon an ensemble PSO-based approach for diagnosis of coronary artery disease," in *Proceedings of the 18th Iranian Conference of Biomedical Engineering (ICBME '11)*, pp. 249–254, IEEE, Tehran, Iran, December 2011.
- [33] N. A. Setiawan, P. A. Venkatachalam, and M. H. Fadzil, "Rule selection for coronary artery disease diagnosis based on rough set," *International Journal of Recent Trends in Engineering*, vol. 2, pp. 198–202, 2009.
- [34] T. Agotnes, *Filtering Large Propositional Rule Sets While Retaining Classifier Performance*, Department of Computer and Information Science: Norwegian University of Science and Technology, Trondheim, Norway, 1999.
- [35] M. C. Tu, D. Shin, and D. Shin, "Effective diagnosis of heart disease through bagging approach," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, Tianjin, China, October 2009.
- [36] J. G. Bazan, M. S. Szczuka, A. Wojna, and M. Wojnarski, "On the evolution of rough set exploration system," in *Proceedings of the 4th International Rough Sets and Current Trends in Computing Conference (RSCTC '04)*, pp. 592–601, June 2004.
- [37] K. Srinivas, G. Raghavendra Rao, and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," in *Proceedings of the 5th International Conference on Computer Science and Education (ICCSE '10)*, pp. 1344–1349, Hefei, China, August 2010.
- [38] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 8, no. 8, pp. 343–350, 2008.