Received: 1 April 2019 | Revised: 23 July 2019 | Accepted: 4 October 2019
DOI: 10.1111/exsy.12485

Expert Systems WILEY

SPECIAL ISSUE PAPER

Hybrid particle swarm optimization for rule discovery in the diagnosis of coronary artery disease

Mariam Zomorodi-moghadam[1] | Moloud Abdar[2] | Zohreh Davarzani[3] | Xujuan Zhou[4] | Pawel Pławiak[5] | U.Rajendra Acharya[6]

[1]Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Département d'Informatique, Université du Québec à Montré al, Montré al, Quebec, Canada
[3]Department of Computer Engineering, PNU University, Tehran, Iran
[4]School of Management and Enterprise, University of Southern Queensland, Queensland, Australia
[5]Institute of Telecomputing, Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Krakow, Poland
[6]Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore, Singapore

Correspondence
Mariam Zomorodi-moghadam, Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
Email: m_zomorodi@um.ac.ir

Abstract

Coronary artery disease (CAD) is one of the major causes of mortality worldwide. Knowledge about risk factors that increase the probability of developing CAD can help to understand the disease better and assist in its treatment. Recently, modern computer-aided approaches have been used for the prediction and diagnosis of diseases. Swarm intelligence algorithms like particle swarm optimization (PSO) have demonstrated great performance in solving different optimization problems. As rule discovery can be modelled as an optimization problem, it can be mapped to an optimization problem and solved by means of an evolutionary algorithm like PSO. An approach for discovering classification rules of CAD is proposed. The work is based on the real-world CAD data set and aims at the detection of this disease by producing the accurate and effective rules. The proposed algorithm is a hybrid binary-real PSO, which includes the combination of categorical and numerical encoding of a particle and a different approach for calculating the velocity of particles. The rules were developed from randomly generated particles, which take random values in the range of each attribute in the rule. Two different feature selection methods based on multi-objective evolutionary search and PSO were applied on the data set, and the most relevant features were selected by the algorithms. The accuracy of two different rule sets were evaluated. The rule set with 11 features obtained more accurate results than the rule set with 13 features. Our results show that the proposed approach has the ability to produce effective rules with highest accuracy for the detection of CAD.

KEYWORDS

classification, coronary artery disease (CAD), hybrid particle swarm optimization, rule discovery

1 | INTRODUCTION

Investigations show that coronary artery disease (CAD) is one of the main causes of mortality throughout the world. In the recent years, the number of deaths due to the CAD is increasing, and, therefore, it is necessary to find the important factors contributing to the disease. Nowadays, data mining is becoming increasingly popular and important in healthcare. Medical data is different from other data sets. The patient records sometimes lack specific data, and such incomplete data sets should be considered in preprocessing steps (Kononenko, 2001). Other aspects of medical data include ethical issues, data security, and uncertainty and errors (Cios & Moore, 2002; Kononenko, 2001).

The heart disease is a general term used for a diverse type of diseases that affect the heart (Srinivas et al., 2010). Among them, CAD is the most common cardiovascular condition (Arabasadi et al., 2017). The data mining techniques can be utilized to diagnose heart disease from the patient's clinical data, which may contain data about patients that is interrelated or incomplete (Kavitha & Kannan, 2016).

Initially, the data repository is preprocessed, which has a direct impact on the quality of the final results. The preprocessing phase, which is the first step in the association of rule mining is used to deal with missing values. Data mining techniques are used for different goals, which depends

Mariam Zomorodi-moghadam and Moloud Abdar contributed equally to this work.

on the modelling objective. Two of the most common modelling objectives are classification and prediction (Palaniappan & Awang, 2008). Classification models predict classes for data set elements, and numeric prediction models are used for modelling continuous valued functions.

The classification rule discovery is one of the important tasks in data mining, which aims at finding a small set of rules from the training data set. The purpose of classification is to predict the class c_i for an intended data input. Various classification techniques exist including decision trees, rough set methods, statistical methods, logistic regression, naive Bayes, support vector machines, and neural networks (Abdar & Zomorodi-Moghadam, 2018; Apté & Weiss, 1997; Bazan et al., 2000; Mazurowski et al., 2008; Safavian & Landgrebe, 1991). In particular, decision trees run a hierarchical procedure to recursively partition the data set until all subsets fit a particular class (Friedl & Brodley, 1997).

The classification problem is a difficult task owing to the large number of parametres involved. The robustness of evolutionary algorithms and the fact that they perform a global search in the space makes them to extract rules and mine data. New evolutionary algorithms, such as ant colony optimization algorithm, particle swarm optimization (PSO) algorithm, and also genetic algorithm (GA) have been employed to extract interesting rules and patterns from data sets (Mukhopadhyay et al., 2014; Sousa et al., 2004). For example, GA has been applied to classification tasks (Fidelis et al., 2000; Hemanth & Anitha, 2019; Yang & Honavar, 1998). Moreover, the PSO algorithm, which is a meta-heuristic algorithm derived from the social behaviour of some animals and insects like birds, bees, and so forth, has grown significantly and used in many applications (Mishra, 2016). The PSO utilizes few operators and also its relative simplicity and information-sharing mechanism related to this method have made it a good candidate in which can be extended from single to multi-objective optimization (Tahyudin & Nambo, 2017). Many optimization problems can be solved by PSO algorithm, but applications in data mining is a relatively new area of research (Wang et al., 2007).

In this work, a modified PSO algorithm is applied on the CAD data set to discover the classification rules. The classification rules are encoded in particles of the PSO algorithm, and particles are updated according to PSO equations. As the CAD data set consists of attributes of real, integer, and binary type, particles have these data types in their encoding. We proposed modified velocity and position calculation for binary attributes. To the best our knowledge, this is the first study for applying PSO algorithm for rule discovery in diagnosis of CAD. Moreover, in this study, a hybrid of real and binary PSO is proposed. In other words, our proposed methodology combines two types of PSO in diagnosing CAD.

The rest of the article is organized as follows: the background about CAD and method employed are presented in Section 2. Then in Section 3, the related works using evolutionary algorithms to produce the rules are discussed. The brief explanation about the PSO algorithm is provided in Section 4. In Section 5, the proposed methodology will be discussed. Experimental results of the proposed algorithm are reported in Section 6, and finally in Section 7, we conclude the paper with a few proposals for future work.

2 | BACKGROUND

The task of classification has been extensively studied in data mining, and over decades, many approaches have been proposed to do it. Assuming a data set $\mathcal{D}$ consisting of d items and each having k attributes or features and a respective outcome or class c_i, the classification is the process of predicting a certain outcome or class for an unknown item by examining the features of each item in $\mathcal{D}$. The outcome of the classification can be a model or a function, which describes the class of each data item (Han et al., 2011).

One of these models is *classification rules*, which is represented as *IF-THEN* rules (Abdar et al., 2015; Han et al., 2011), and this model is used in this paper for classification.

Definition 1. Rule discovery algorithm is essentially the algorithm that is used to produce some IF-THEN rules for classification.

The input to the rule discovery algorithm is the data from the training set of a given data set consisting of samples $(x_i, y_i)_{i=1,\dots,n}$. x_i is the input and in our case; it is a vector with the size k, equal to the number of attributes in the data set, and y_i is the label that we try to predict.

3 | RELATED WORK

Nowadays, with the increasing development in data mining applications in human life, working with large amount of data is possible. Data mining has many applications in different areas such as marketing, banking, business, healthcare, and so forth. (Abdar et al., 2018; Berry & Linoff, 2004; Bigus, 1996; Chye & Gerry, 2002). With the help of data mining, useful patterns of information can be extracted from the data to classify automatically and reduce the manual labor.

Data mining tasks have been successfully utilized in healthcare, where there are many applications in this sector due to huge amount of data being generated (Abdar, 2015; Acharya et al., 2003; Acharya et al., 2012; Ksiażek et al., 2019). In one of the earliest works, data mining technique has been applied to detect acute myocardial infarction detection (Baxt, 1990). The neural network showed better performance than other approaches.

The performance of different data mining techniques for medical application has been investigated in Lavra c (1999). The authors have considered decision tree algorithms and Bayesian classifiers in their work. Compared with decision trees, authors believed that the explanation of decisions made by the naive Bayesian classifier is more appealing to physicians.

Data mining techniques have also been applied for multi-disease prediction modelling (Chang et al., 2011) and common risk factors of two diseases including hypertension and hyperlipidemia. A two-stage procedure has been applied. In the first stage, six data mining approaches have

been used, and the risk factors for hypertension and hyperlipidemia have been identified. In the second stage, multivariate adaptive regression splines (Friedman, 1991) method has been used for the construction of a predictive model based on the common risk factors of two diseases.

In artificial intelligence, many computational intelligence systems have been proposed. Many of these algorithms have applications in machine learning. The GA, ant colony optimization, artificial immune system, and PSO are some of the well-known evolutionary algorithms in machine learning. Many attempts have been made to apply evolutionary computation algorithms as machine learning techniques (Zhang et al., 2011) and these algorithms yielded excellent performances using medical data sets (Freitas, 2003). The GA has been applied on real-world data sets in the medical domain to discover the classification rules (Fidelis et al., 2000). In Pławiak (2018b), GA has been used at different stages of cardiac health recognition, and also in another work, the author has applied GA using ensemble technique for myocardium dysfunction recognition (Pławiak, 2018a). In another work (Tayefi et al., 2017), GA was used to produce the initial weights of neural network and then classified using neural network. The PSO has been applied for data clustering and has showed comparable performance to k-means clustering (Van der Merwe & Engelbrecht, 2003). The authors have also extended the algorithm to use k-means clustering algorithm as the initial step to seed the swarm. The PSO has been also applied for classification task (Sousa et al., 2003; Sousa et al., 2004). In these two works, three PSO variants have been compared. The experiments have been conducted using three different data sets. The role of the PSO algorithm for rule discovery is to find and return the most appropriate rules for the classification task. The selection criterion for the rules is related to the number of instances in the training set, which were classified by those rules correctly. The rules discovered from the classification algorithm go through a pruning process, and unnecessary attribute tests are removed. The accuracy of the rule set has been carried out by tenfold cross-validation method.

Rule discovery has been studied in Hassoon et al. (2017) using GA for liver disease prediction. The algorithm does not use GA to produce rules from scratch, but instead, it applies the rules released from Boosted C5.0 classification as input to the algorithm. The GA improves the rules to obtain higher accuracies.

The PSO-based rule mining has been used in Wang et al. (2007) and tested with few data sets including Wisconsin breast cancer, Ljubljana breast cancer, heart disease, dermatology, tic-tac-toe, and hepatitis. Its performance has been compared with Ant-Miner and organizational coevolutionary algorithm classification (OCEC), which is a genetic classifier algorithm where evolutionary operations act on the given data directly (Liu et al., 2003). It has been shown that the accuracy of the PSO is better than Ant-Miner and OCEC. Also, the number of rules are discovered by PSO, and also its run time has been less than both Ant-Miner and OCEC.

In Alkeshuosh et al. (2017), PSO has been used for producing the rules for diagnosing heart disease. They have encoded the rules using the attributes of the data set. Each particle is divided into three parts. The first part is a bit, which can be either zero or one, indicating that the corresponding attribute is inactive or active, respectively. The second part represents the operator used for connecting the attributes in the rule, and the third part is the value of the attribute itself. The fitness function consists of the number of the accuracy parametre and the number of attributes in the data set. Then the authors compared their work with C4.5 algorithm and showed that the PSO has obtained better accuracy as compared with the C4.5 decision tree algorithm.

Regarding the feature selection approaches, authors in Wosiak and Zakrzewska (2018) have suggested a feature selection algorithm based on three different algorithms including reverse correlation algorithm, classical floating search method, and ReliefF algorithm. Then two existing clustering algorithms, expectation–maximization, and k-means were applied, and the clusters for patients were created. It was shown that the performance of reverse correlation algorithm algorithm is like two other feature selection approaches and is even better. It is an unsupervised learning and has more flexible applications.

4 | PSO ALGORITHM

The PSO is based on the simulation of social behaviour of a population. It was first introduced in Eberhart and Kennedy (1995) and modified in Shi and Eberhart (1998). It is simple and easy to implement. Hence, it has been used in different fields such as function optimization, signal processing, machine study, the model classification, and so forth (Bai, 2010).

The algorithm is based on the idea of swarm intelligence (Bai, 2010). The basic elements of the PSO algorithm are particles. Each particle has some properties and can communicate with other particles in the system. Each particle is equivalent to an individual in genetic-based representation, and the swarm is a population of particles. The PSO is some kind of the population-based optimization algorithm and checks for the solution in the search space. Each particle has a position, and the position of each particle shows a potential solution to the problem in the hand. A swarm of n particles are distributed in the D-dimensional search space. Each particle i has a current position x_i, velocity v_i, and its personal best position p_i, which is the best experience of that particle so far, according to the given fitness function. There is also a best position encountered by the whole swarm denoted as p_{gb}. The direction for the movement of a particle is based on four components: current velocity of the particle, current position of the particle, its best position so far, and position of the best particle or particles in the swarm. This direction is calculated and placed in a parametre called *velocity*. Hence, all iterations of the algorithm update the position of the particles in the swarm based on their calculation of velocity as Equations (1) and (2; Shi, 2001):

$$X_i(t+1) \leftarrow X_i(t) + V_i(t), \tag{1}$$

$$V_i(t+1) \leftarrow WV_i(t) + C_1 r_1 (pbest_i(t) - X_i(t)) + C_2 r_2 (gbest(t) - X_i(t)). \tag{2}$$

In these equations, $X_i(t)$ shows the current position of particle i at time t, and $V_i(t)$ is the velocity of particle i at time t, which is based on two best values. The first one is $\text{pbest}_i(t)$, which is the best position found by particle itself so far, and the second is gbest(t), which is the best position found so far by the whole swarm. W is an inertia coefficient, C_1 and C_2 are positive constants called acceleration coefficients, which determine the relative acceleration towards $pbest_i$ and gbest. r_1 and r_2 are two random variables within the range of 0 and 1. So the next position of the particle $X_i(t+1)$ is obtained by Equations (1) and (2).

5 | PSO-BASED RULE DISCOVERY FOR CAD

In this work, a PSO-based algorithm is proposed to discover the classification rules in the real-world data set. A set of random classification rules are encoded in the particles using PSO representation, and then according to their accuracy, they are changed and improved until some classification rules with high accuracy are achieved. In this section, first, the components of the PSO algorithm for rule construction are described, and then the proposed algorithm for rule discovery problem using PSO as the base algorithm is presented.

As each particle includes the values of different attributes in the data set and the data set consists of different data types, the proposed algorithm has some differences with the original PSO algorithm. Different data types affect particle encoding and position calculation of particles, which is already affected by velocity calculation of the particles.

5.1 | Data set evaluation for different data types

Our data set contains numerical and categorical data. The numerical data is further divided into real and integer data types. The categorical data is also divided into binary and multivalued data types. The most relevant features or attributes are selected by an appropriate feature selection algorithm, and then the algorithm is applied to the new data set.

5.2 | Particle encoding and rule representation

Like other population-based optimization algorithms, there should be a representation for particles in the system. Two different approaches have been considered in the literature for encoding a rule (Casillas et al., 2007; Freitas, 2013).

The first one encodes each individual of the population as a set of rules, which is an entire candidate solution. This approach is called Pittsburgh method. The solution for the problem of rule discovery is essentially a set of rules rather than a single rule. Thus, encoding a complete set of rules in a single particle seems a natural encoding in evolutionary based rule discovery algorithms. But in fact, it results in a big and complicated encoding. In the second method, an individual represents a single rule. This approach is known as the Michigan method. In this work, Michigan method is used. So each particle shows a single rule for the given data set. At the end of the execution of the algorithm, the best generated rule is selected. Therefore, the algorithm should be executed several times to obtain a set of rules. It is also possible to select a set of best rules for each execution of the algorithm. This approach is useful when the number of classes is noticeable, and, actually, it is not possible to run the algorithm many times.

Encoding just one rule in a particle is appropriate for data sets with two or limited number of classes because as discussed, the algorithm is executed once for each class of data.

5.2.1 | Particle structure

A rule has an IF-THEN form, where IF part or antecedent is a conjunction of conditions, and THEN part or consequent is the prediction. Hence, the particle should include the IF and THEN parts of just one rule (Michigan method; Casillas et al., 2007).

Encoding the rule consequent (THEN part)

There are two approaches for representing the rule consequent. In the first one, the rule consequent is encoded in the particle, and a number of bits are added to the particle for the rule consequent. In the second approach, the rule consequent is not encoded in the particle, and all particles are assumed the same rule consequent. This approach simplifies the PSO design including the fitness calculation and also calculating the position of particles, but the PSO should be run once for each rule consequent. In particular, this approach is appropriate when the number of rule consequents are limited. Our data set consists of two rule consequents, class for patients and healthy peoples. We have used the later approach and did not encode the rule consequent in the particle, and therefore, we executed the PSO once for each rule consequent or class.

Thus, by using Michigan method, each particle represents only one rule, and by not including the consequent of the rule in the encoding, the particle consists of encoding of just the antecedent part of that rule.

Encoding the rule antecedent (IF part)

The structure of each particle, which is the encoding of the rule antecedent (IF part), is as follows:

The particle is divided into n, almost equal size, segments devoted to different conditions of the rule, where n is the number of selected features in the data set. An implicit logical AND operator is also assumed between different segments of each particle. Each segment contains the lower and upper bounds of the corresponding attribute or just the exact value of it, depending on the type of that attribute and a bit that

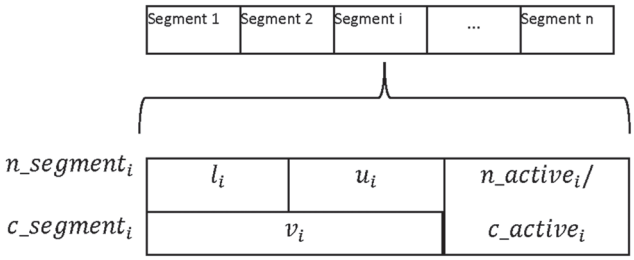

FIGURE 1 Particle structure representing a rule

att1			*att2*			*att3*			*att4*	
10	37	1	32	120	0	5	100	1	3	1

FIGURE 2 A sample encoding for rules

indicates the existence of the attribute in the rule. So each segment of the particle is divided into two or three parts for numerical and categorical attributes, respectively. The segments for numerical attributes are called *n*_segment and segments for categorical attributes are called *c*_segment. The detail of the structure of each segment of the particle, for numerical and categorical attributes, is described below.

Numerical attributes

$n\_segment_i$ $0 \le i \le c$; where n is the number of numerical attributes is further divided into three fields. The first part is l_i and indicates the lower bound of that attribute in the rule. The second part is u_i, and it indicates the upper bound of that attribute in the rule, and the last part is a bit called n_active_i that is "1" if that attribute is presented in the rule, that is, it is an active attribute, and "0" otherwise.

Categorical attributes

$c\_segment_i$ $0 \le i \le k$; where k is the number of categorical attributes is further divided into two fields. The first part is v_i that shows the assumed value of that attribute in the encoding, and the second part is for the field "active" and is called c_active_i.

According to the above discussion, the encoding of the particle is presented in Figure 1.

For example, a rule with three active attributes out of total four attributes in the data set and the encoding of Figure 2 is expressed as *if* $10 < att1 < 37 \wedge 5 < att3 < 100 \wedge att4 = 3$. As it can be seen, the THEN part is not encoded in the rule. In other words, the class label of rules is not included in the encoding.

In this work, all particles belong to the same class of interest. Thus, when the fitness function is computed for particles, they are assumed the same implicit class. Hence, to generate the rules for the other class, a separate run of the algorithm is performed.

Our data set has just two classes, c_1 for healthy people and c_2 for patients with CAD. Therefore, the algorithm is required to be executed once for each class, and run gives the rules for that specific class.

Another issue is the type of attributes in data mining tasks. In this work, we have two types: categorical and numerical types. Therefore, PSO algorithm should be modified for rule discovery to provide a way for encoding these two different types of attributes.

This work combines the categorical and numerical attributes in the encoding of a particle. Excluding the class of the data in the particle encoding, the number of fields in the particle is obtained by Equation 3. In fact, in the proposed approach, the class of the data is not included in the particle encoding.

$$particle_{size} = 3N_{numeric} + 2N_{cat}, \tag{3}$$

where $N_{numeric}$ is the number of numerical attributes in the data set and N_{cat} is the number of categorical attributes in the data set.

Using the above notation, the number of attributes is $N_{att} = N_{numeric} + N_{cat}$. Each rule has an antecedent part that is a conjunction of conditions and a rule consequent that shows the predicted class if the antecedent part is correct, so the rule takes the following form:

if (a conjunction of the conditions or attributes' values) then class $= c_i$

Each instance in our data set is mapped into one of the two classes "CAD" with label c_1 for healthy people and "Normal" with lable c_2 for those with CAD. Therefore, as there are two distinct classes in our data set, which appeared in the rule consequent, two separate running of the algorithm are performed. Each run of the algorithm considers one of the classes as the rule consequent, and the fitness of the rule is measured by assuming class as the predicted class of the rule.

5.3 | Evaluation function

As other evolutionary algorithms, there should be a metric for evaluating the quality of each particle in the swarm. Here, each particle represents a rule. The quality of each rule is related to the number of instances that rule can predict correctly. Four evaluation metrics are measured for that rule in the training set, which are true positive, true negative (TN), false positive, and false negative (Han et al., 2011).

Thus, the evaluation process takes the rule and applies the rule to each data instance in the training set to generate the above evaluation metrics. Then, the values of these metrics are applied to an evaluation function.

Several evaluation functions are used in the literature to measure the effectiveness of rules generated in the classification system. Some of them are accuracy, precision, and sensitivity.

The above-mentioned evaluation functions show the quality of the generated rule and have been used in the literature alone or in combination to evaluate the effectiveness of a rule.

Combination of precision and sensitivity can be used by their multiplication as precision × sensitivity. A very specific rule has high precision and low sensitivity, and so the value of the fitness function is not very high (Freitas, 2013).

Having the accuracy, sensitivity, and precision of a rule, we express the fitness function as a weighted combination of two terms as indicated in Equation 4. The first term is precision × sensitivity, and another by the rule divided by the total data instances in the data set. Including accuracy in the fitness function also leads to the contribution of the last measure (TN) in rule evaluations.

$$Fitnessfunction = w_1(precision \times sensitivity) + w_2(accuracy) \tag{4}$$

5.4 | Position calculation of particles

Particles in our problem are the rules generated for the CAD. The attributes of the corresponding data set are of different types, having both categorical and numerical attributes consisting of integer and real attributes. So according to the velocity equation of PSO, the velocity calculation of each particle includes binary, integer, and real-valued items. In the velocity calculation, addition, multiplication, and subtraction operators act on the particle. In our problem, the particle includes data items with different types. In fact, we should define operators for numerical attributes, equivalent to addition, subtraction, and multiplication for real and integer types.

In this work, a new scheme is proposed to solve the hybrid real–binary particle representation and is given below:

First, in the preprocessing stage, the data set is reordered, and all numerical attributes are placed next to each other. And after obtaining the real-valued attributes, the velocity is calculated according to Equations (5) and (6) for real-integer and binary attributes, respectively. The binary PSO has been suggested in some works like (Kennedy & Eberhart, 1997; Liu et al., 2004; Xue et al., 2012) to solve the problem of binary encoding of particles. According to Equation (6), the addition, subtraction, and multiplication for binary attributes are replaced by logical operators ∨, ⊕, and ∧, respectively.

Real attributes (the same as Equation 2):

$$V_{i(real,integer)}(t+1) = W \times V_i(t) + C_1 \times (pbest_i - X_i(t)) + C_2 \times (gbest - X_i(t)). \tag{5}$$

Binary attributes:

$$V_{i(binary)}(t+1) = [r \wedge V_i(t)]vee(r \wedge (pbest_i \oplus X_i(t))) \vee (r \wedge (gbest_j \oplus X_i(t))), \tag{6}$$

where ∧, ∨, and ⊕ are bitwise logical operations, AND, OR, and XOR, respectively.

The parametre r is a random number between 0 and 1, rounded to the nearest 0 or 1.

Integer attributes:

For integer attributes, the velocity is calculated as Equation (5), and then it is rounded up to the next integer value. As active fields have also binary values, for calculating the velocity, they are treated like binary attributes as Equation (7):

$$act_i(t+1) = r \wedge act_i(t) \vee r \wedge (act_{pbest_i} \oplus act_i(t)) + r \wedge (gbest_j \oplus act_i(t)), \tag{7}$$

where r is a random number between 0 and 1, rounded to the nearest 0 or 1.

Using the velocity of each particle, the position is updated according to Equation (1), with a small change for binary types.

Real and integer attributes are same as Equation (2).

Binary attributes are indicated in Equation (8).

$$X_i(t+1) = X_i(t) \vee V_{i(binary)}(t), \tag{8}$$

where ∨ is bitwise logical operation, OR.

In the final step, to update the position of particles, they are checked against the minimum and maximum boundaries of attributes as follows:

Real and integer attributes:

```
for all particles i do
  for all attributes (j) of particle (i) do
    if attribute (j) is active then
      lower bound attribute (j)= max (X(i),min_attribute)
      upper bound attribute (j) = min (X(i),max_attribute)
    end if
  end for
end for
```

where min_attribute and max_attribute are the minimum and maximum possible values of attribute *i*, respectively.

Binary attributes:

As the position of the particle at binary-valued indices is zero or one, there is no need to adjust these values to the boundaries.

5.5 | Outline of the algorithm

Based on the particle encoding and fitness function of Equation (4), the PSO algorithm is applied to the data set.

First, the data set is separated into two training and test sets. In our tests, the 10-fold cross-validation approach is used. The outline of the Algorithm 1 and Algorithm 2 for rule discovery is as follows:

Algorithm 1 Hybrid PSO rule discovery algorithm

```
Input: Data set D = D_Train ∪ D_Test, PSO parametres, Problem parametres
Output: classification rules for class c_i (i = 1, 2)
Initialize_particles ()
for k = 1 to TrainNumber do
 | Train Array(k)=read one record in data set
end
globalbest = 0
for i = 1 to (pop_size) do
   particle (i) = randomly_generated
   particle(i).fitness=fitness(particle (i), Train_Array, Class_label)
   set particle(i)'s best position to the the particle current position and
   its best fitness to its current fitness
   if particle(i).pbest.fitness >globalbest.fintess then
    | Set globalbest to the particle(i).pbest
   end
   Set metrics TP, TN, FP, FN to zero
   i = i+1
end
***main loop of the algorithm***

for iteration = 1 to (iteration_no) do
   for i = 1 to (pop_size) do
      calculate particle(i).velocity according to Equations (2), (5), and (6)
      calculate particle(i).position according to Equations (1) and (8)
      particle(i).fitness = Fitness (particle (i), Train_Array, Class_label)
      if particle(i).pbest.fitness >particle(i).fitness then
         Set particle(i).pbest.position to the particle(i).position
         and the fitness of the particle(i).pbest to the fitness of particle(i)
         %% Updating the best of the whole swarm or the global best
         if globalbest.fitness >particle(i).pbest.fitness then
          | Set globalbest to particle(i).pbest
         end
      end
      Set metrics TP, TN, FP, FN to zero
      i++
   end
   iteration++
end
Selected_Rule = globalbest
**************************Test of Algorithm*********************

TP=0
TN=0
for i = 1 to TestNumber do
 | R
end
(i)=Read one record from Test set
if Selected_Rule satisfies all attribute in R(i) && R(i).Class==Class_label then
 | TP++
end

if Selected_Rule doesn' t satisfy at least one of attribute in R(i) && R(i).Class ≠ Class_label then
 | TN++
end

AccuracyRate=(TP+TN)/TestNumber
%%Other metrics are obtained similarly.
```

Algorithm 2 Fitness function

```
Function Fitness= Fitness (particle, Train_Array,Class)
Set metrics TP, TN, FP, FN to zero
for i = 1 to Train Number do
    if particle satisfies Train_Array (i) then
        if Train_Array (i).Class==Class_label then
            TP++
        else
            FP++
    else
        if Train_Array (i).Class==Class_label then
            FN++
        else
            TN++
Set Precision to TP/(TP+FP)
Sensitivity to TP/(TP+FN)
and Accuracy to (TP+TN)/(TP+FP+TN+FN)
Fitness= w1*Precision*Sensivity+w2*Accuracy
End Function
```

TABLE 1 Selected features for each feature selection algoritm

Feature selection approach	Number of attributes	Selected attributes
FS1	11	Age, K, EF-TTE, airway disease, typical chest pain, Q wave, Tinversion, ESR, DM, HTN, Region RWMA
FS2	13	Age, ESR, K, EF-TTE, BP, DM, HTN, typical chest pain, atypical, Q wave, St depression, Tinversion, region RWMA

Abbreviations: BP, blood pressure; DM, diabetes milletus; EF-TTE, ejection fraction transthoracic echocardiography; HTN, hypertension; RWMA, regional wall motion abnormalities.

1) The initial random swarm is generated where the size of each particle is obtained by Equation (1). The swarm size is given in Table 1. Basically, this swarm is an array of random rules. Considering the proposed encoding for the particles, each real or integer attribute in the encoding has an upper bound and a lower bound for its value. These bounds are based on the minimum and maximum values of these attributes in the data set. So the randomly generated rules are bound to some minimum and maximum values.
2) The fitness for each particle in the swarm is calculated. The fitness function requires the calculation of evaluation metrics namely true positive, TN, false positive, and false negative. So before calling the fitness function, the algorithm goes through the training set with one training item at a time and evaluates if the rule matches that specific training item in the data set or not.
3) The initial swarm goes through the main loop of the algorithm, and the initial rules are improved interatively according to the PSO equations. At the end of each run of the iteration loop, the best rules in that swarm and the best global rule are selected.
4) At the last iteration, the overall best rule in the swarm is selected.
5) The accuracy of the best discovered rule is tested using the test data.

According to the outline of the PSO-based algorithm, our proposed system provides the best rules from a search space of all possible rules. As the search space is very huge, the evolutionary algorithms are a good candidate for producing the discovery rules. PSO, especially, is a very efficient global search evolutionary algorithm, with few algorithm parametres and light computations. The model is built at the end of the PSO loop, where the last generation contains the near optimal rules in terms of their accuracies on the train set. Then, the accuracy of the model is evaluated on the test set. So by applying a new data to this model, it classifies the specific data into the class patient or healthy, using the final discovered rules of the model. The flowchart of the proposed system is provided in Figure 3.

6 | EXPERIMENTS

There are several steps that need to be performed before starting the experiment, like selecting the appropriate system configuration and data set.

6.1 | System configuration

The implementation of the proposed algorithm has been conducted in MATLAB on a 2.7 GHz Intel Core i7 processor with 8 GB of RAM.

6.2 | Data set

The data set for extracting rules were taken from UCI machine learning repository for CAD (Asuncion & Newman, 2007). The CAD data set consists of 303 records and 56 attributes classified into two different categories, CAD and normal. The patients with diameter narrowing greater than or equal to 50% are categorized as CAD, and the rest are categorized as normal.

6.3 | Feature selection approach

To decrease the number of irrelevant features in the data set, two different feature selection algorithms were used. The first one was multi-objective evolutionary search with 11 features, and the second one is PSO search with 13 features. We named these feature selection approaches as feature selection 1 (FS1) and feature selection 2 (FS2), respectively. Table 1 shows the resulted features for both feature selection techniques.

6.4 | Parametre tuning

The chosen PSO parametres are indicated in Table 2. The number of iterations was selected using trial and error in 10 runs of the algorithm.

The weights of the fitness function were selected as $w_1 = 0.8$ and $w_2 = 0.2$. To evaluate the accuracy of the proposed algorithm, the generated rule set is applied to the test set, and the percent of instances that the rule set predicts accurately is considered as the accuracy rate of that rule set. Equation 4 was applied to the rule set for testing.

The discovered rules for 10 runs of the algorithm from the first experimental setup and for feature selection FS1 and FS2 are presented in Tables 3 to 6. Tables 3 and 4 show the rules obtained using FS1 and for class c_1 and c_2 of the data set, respectively. Similarly, Tables 5 and 6 show the rules obtained using FS2 and for classes c_1 and c_2 of the data set, respectively.

In this work, 10-fold cross-validation technique was used to train and test the accuracy of the rules. The accuracy of rules for class c_1 and for different training and test sets is illustrated in Figure 4, and the accuracy of rules for class c_2 and for different training and test sets is illustrated in Figure 5.

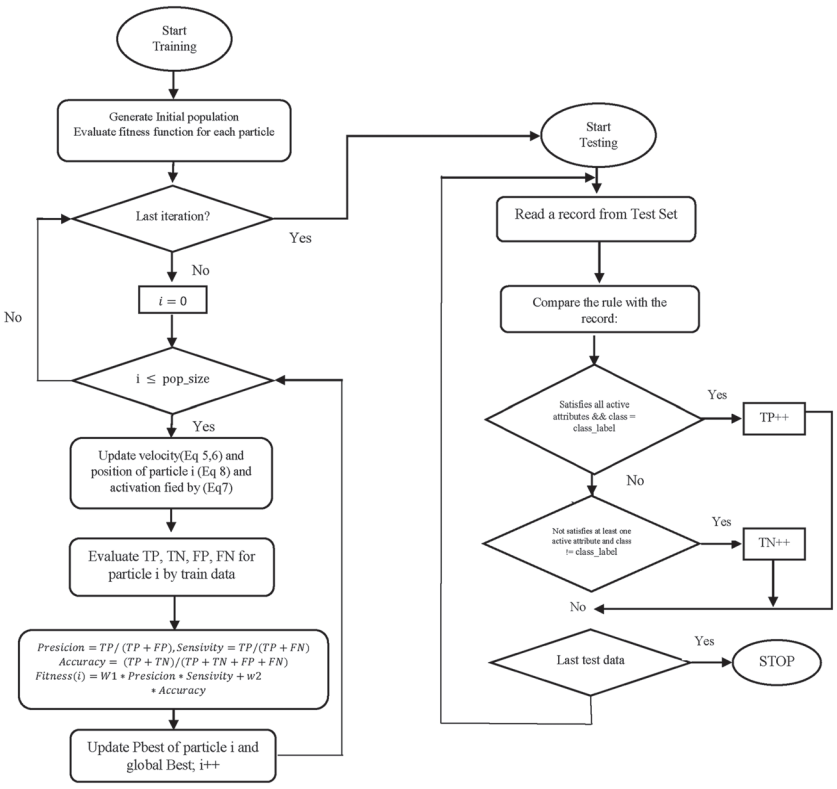

FIGURE 3 Flowchart of the proposed system

TABLE 2 The particle swarm optimization parametres in the hybride binary–real particle swarm optimization algorithm for rule discovery

Parametre	Value
Maximum number of iterations	300
Swarm size	200
W	0.75
C1	0.5
C2	0.5

TABLE 3 Rules extracted by the hybrid binary–real particle swarm optimization algorithm using feature selection 1 for class c_1

Number	Rule
1	$3 < Age < 67 \wedge Airwaydisease = 0$ $\wedge QWave = 0 \wedge TypicalChestPain = 0$ $52 < Age < 80 \wedge Airwaydisease = 0\wedge$ $DM = 0 \wedge TypicalChestPain = 0$
2	$54 < Age < 85 \wedge HTN = 0 \wedge TypicalChestPain = 0$ $37 < Age < 81 \wedge TypicalChestPain = 1 \wedge RegionRWMA = 0$ $\wedge Qwave = 0 \wedge TypicalChestPain = 1$
3	$48 < Age < 68 \wedge 34 < EF - TTE < 58\wedge$ $QWave = 0 \wedge TypicalChestPain = 0$ $3 < k < 6 \wedge 21 < EF - TTE < 58 \wedge Tinversion = 0$ $\wedge HTN = 0 \wedge TypicalChestPain = 0$
4	$31 < Age < 85 \wedge DM = 0 \wedge HTM = 0$ $\wedge TypicalChestPain = 0$ $34 < Age < 84 \wedge DM = 0\wedge$ $\wedge TypicalChestPain = 0 \wedge RegionRWMA = 0$
5	$37 < Age < 62 \wedge Airwaydisease = 0\wedge$ $Tinversion = 0 \wedge HTN = 0 \wedge TypicalChestPain = 0$ $HTN = 0 \wedge TypicalChestPain = 0$
6	$36 < Age < 63 \wedge 3 < k < 5 \wedge DM = 0$ $\wedge HTN = 0 \wedge TypicalChestPain = 0 \wedge RegionRWMA = 0$ $35 < Age < 68 \wedge 3 < K < 5 \wedge Airwaydisease = 0 \wedge Qwave = 0$ $Tinversion = 0 \wedge TypicalChestPain = 0$
7	$35 < Age < 65 \wedge TypicalChestPain = 0$ $6 < ESR < 67 \wedge 4 < K < 6 \wedge Airwaydisease = 0$ $\wedge Qwave = 0 \wedge Tinversion = 0 \wedge TypicalChestPain = 0$
8	$31 < Age < 76 \wedge DM = 0 \wedge TypicalChestPain = 0$ $45 < Age < 65 \wedge Airwaydisease = 0 \wedge QWave = 0$ $\wedge DM = 0 \wedge HTN = 0 \wedge TypicalChest = 0$
9	$31 < Age < 76 \wedge DM = 0TypicalChestPain = 0\wedge$ $32 < Age < 78 \wedge 4 < K < 6 \wedge Tinversion = 0 \wedge TypicalChestPain = 0$
10	$38 < Age < 60 \wedge TypicalChestPain = 0 \wedge RegionRWMA = 0\wedge$ $1 < ESR < 76 \wedge TypicalChestPain = 0$

TABLE 4 Rules extracted by the hybrid binary–real PSO algorith using feature selection 1 for class c_2

Number	Rule
1	$Airwaydisease = 0 \wedge Qwave = 0\wedge$ $Qwave = 0 \wedge Typicalchestpain = 1$
2	$Airwaydisease = 0 \wedge typicalchestpain = 1$ $3 < K < 5 \wedge Qwave = 1 \wedge Typicalchestpain = 1$
3	$Qwave = 0 \wedge Typicalchestpain = 1$ $2 < ESR < 79 \wedge Typicalchestpain = 1$
4	$40 < Age < 83 \wedge Airwaydisease = 0$ $47 < Age < 75 \wedge HTN = 0$
5	$4 < K < 6 \wedge BP = 0 \wedge Qwave = 1$ $HTN = 1 \wedge Typicalchestpain = 1$
6	$Airwaydisease = 0 \wedge HTN = 1\wedge$ $31 < EF - TTE < 55 \wedge Typicalchestpain = 1$
7	$34 < Age < 76 \wedge 3 < K < 6\wedge$ $26 < EF - TTE < 59 \wedge TypicalChestPain = 1$ $Qwave = 0 \wedge TypicalChestPain = 1$
8	$4 < k < 6 \wedge Airwaydisease = 0\wedge$ $3 < K < 6 \wedge HTN = 1$
9	$Qwave = 0 \wedge Typicalchestpain = 1\wedge$ $3 < K < 5 \wedge 29 < EF - TTE < 52 \wedge Airwaydisease = 0\wedge$ $TypicalChestPain = 1$
10	$31 < Age < 83 \wedge 3 < k < 5 \wedge Airwaydisease = 0 \wedge Qwave = 0$ $\wedge Qwave = 0 \wedge Typicalchestpain = 1$

Abbreviations: BP, blood pressure; EF-TTE, ejection fraction transthoracic echocardiography; HTN, hypertension.

TABLE 5 Rules extracted by the hybrid binary–real particle swarm optimization algorithm using feature selection 2 for class c_1

Number	Rule
1	$4 < K < 6 \wedge 118 < BP < 181 \wedge Atypical = 1$ $\wedge Qwave = 0 \wedge Stdepression = 0$ $37 < Age < 71 \wedge 93 < BP < 174 \wedge HTN = 0 \wedge$ $Typicalchestpain = 0 \wedge Qwave = 0 \wedge Stdepression = 0$
2	$41 < Age < 79 \wedge 16 < ESR < 85 \wedge$ $3 < K < 5 \wedge DM = 0 \wedge Atypical = 1$ $36 < Age < 77 \wedge 4 < K < 6 \wedge Atypical = 1$ $\wedge Stdepression = 0$
3	$4 < K < 6 \wedge 97 < BP < 189 \wedge$ $TypicalChestPain = 0$ $Typicalchestpain = 0 \wedge Atypical = 0 \wedge QWave = 0$ $\wedge Stdepression = 0$
4	$DM = 0 \wedge Atypical = 1 \wedge QWave = 0$ $\wedge Tinversion = 0 \wedge RegionRWMA = 0$ $48 < Age < 67 \wedge HTN = 0 \wedge Atypical = 1$ $\wedge Tinversion = 0$
5	$48 < Age < 73 \wedge DM = 0 \wedge HTN = 0 \wedge$ $QWave = 0 \wedge Tinversion = 0$ $37 < Age < 71 \wedge 93 < BP < 174 \wedge HTN = 0 \wedge$ $Typicalchestpain = 0 \wedge QWave = 0 \wedge Tinversion = 0$
6	$42 < Age < 64 \wedge 4 < K < 5 \wedge 130 < BP < 155 \wedge$ $Typicalchestpain = 0 \wedge Stdepression = 0 \wedge Tinversion = 0$ $36 < Age < 77 \wedge 4 < K < 6 \wedge Atypical = 1 \wedge Stdepression = 0$
7	$Typicalchestpain = 0 \wedge QWave = 0$ $2 < ESR < 60 \wedge 3 < K < 6 \wedge 35 < EF - TTE < 56 \wedge Atypical = 1$
8	$35 < Age < 73 \wedge DM = 0 \wedge Atypical = 1$ $\wedge Stdepression = 0$ $3 < ESR < 87 \wedge 3 < K < 6 \wedge DM = 0 \wedge$ $Typicalchestpain = 0 \wedge Atypical = 1$
9	$3 < ESR < 87 \wedge 3 < K < 6 \wedge$ $Typicalchestpain = 1 \wedge HTN = 0$ $Typicalchestpain = 0 \wedge QWave = 0$
10	$32 < EF - TTE < 60 \wedge StDepression = 0 \wedge$ $Typicalchestpain = 0 \wedge Tinversion = 0$ $Typicalchestpain = 0 \wedge Qwave = 0 \wedge DM = 0$

Abbreviations: BP, blood pressure; DM, diabetes milletus; EF-TTE, ejection fraction transthoracic echocardiography; HTN, hypertension; RWMA, regional wall motion abnormalities.

The execution time of the algorithm is influenced by both the algorithm parametres and data set size. The relevant factors are data set size and PSO parametres like iteration number and population size. Actually, by using the PSO algorithm for rule discovery, we first produced few random rules and learned them by PSO using the whole training data. Then, the algorithm improved the rule set, and the best rules were extracted to discover the most relevant factors in classification of CAD data set.

Using PSO algorithm for rule discovery has the advantage of searching the fittest area without needing to examine the whole solution space exhaustively.

Although rule induction algorithms are used for rule discovery tasks, they search the space of candidate solutions locally and hence may not reach a globally acceptable solution like PSO and other evolutionary-based algorithms. Compared with other evolutionary algorithms, PSO has often been used for continuous optimization problems. Here, by changing PSO operators for discrete parts of the particle, it was appropriately adopted for nominal values of attributes. In addition, PSO is more computationally efficient than evolutionary algorithms like genetic algorithm, while yielding the same performance. On the other hand, whereas most approaches for rule induction benefit from local search to find best rules in the problem space, evolutionary algorithms do global search, and so they can discover some interesting rules that other methods cannot discover them (Freitas, 2013). Moreover, in this work, by considering the exact nature of the rules, we made few specific changes at the PSO operations to better fit it for discovering rules.

At the end of each run of the algorithm, 10 rules are kept. Tables 3 and 4 show the rules for class c_1 and c_2 of the algorithm and with FS1 feature set, respectively. Tables 5 and 6 represent the rules for class c_1 and c_2 with FS2 feature set. The discovered rules have many meaningful information about the risk factors of CAD. For example, it can be observed that typical chest pain feature, which is one of the main risk factors of CAD disease, has zero value for almost all cases in Table 3, and in contrast, it has the value one in Table 4, which is the class of patients. So

TABLE 6 Rules extracted by the hybrid binary–real paticle swarm optimization algorithm using feature selection 2 for class c_2

Number	Rule
1	$33 < Age < 81 \wedge 24 < EF - TTE < 52$ $3 < ESR < 51 \wedge 4 < K < 6$
2	$49 < Age < 79 \wedge 3 < K < 5 \wedge Typicalchestpain = 1$ $\wedge Atypical = 0 \wedge Tinversion = 0$ $114 < BP < 181 \wedge DM = 0 \wedge HTN = 1$
3	$41 < Age < 77 \wedge 4 < ESR < 80 \wedge 92 < BP < 187$ $\wedge Typicalchestpain = 1$ $39 < Age < 79 \wedge HTN = 1 \wedge Atypical = 0$
4	$5 < ESR < 59 \wedge 3 < K < 6$ $36 < Age < 66 \wedge Typicalchestpain = 1 \wedge Tinversion = 0$
5	$3 < k < 6 \wedge 120 < BP < 149 \wedge Atypical = 0$ $HTN = 1 \wedge Typicalchestpain = 1$
6	$18 < ESR < 64 \wedge 3 < K < 6 \wedge Typicalchestpain = 1$ $51 < Age < 81 \wedge 3 < K < 5$
7	$4 < K < 6 \wedge Atypical = 0 \wedge Tinversion = 1$ $109 < BP < 186 \wedge Atypical = 0 \wedge Stdepression = 0$
8	$9 < ESR < 77 \wedge Atypical = 0 \wedge Tinversion = 1$ $3 < K < 6 \wedge DM = 1 \wedge Atypical = 0$
9	$5 < ESR < 86 \wedge 4 < K < 6 \wedge 98 < BP < 154$ $\wedge Atypical = 0 \wedge Stdepression = 0$ $3 < K < 5 \wedge 4 < EF - TTE < 35 \wedge Tinversion = 0$
10	$34 < Age < 76 \wedge 3 < K < 6 \wedge 26 < EF - TTE < 59 \wedge TypicalChestPain = 1$ $33 < Age < 59 \wedge Thyroiddisease = 1$

Abbreviations: BP, blood pressure; DM, diabetes milletus; EF-TTE, ejection fraction transthoracic echocardiography; HTN, hypertension; RWMA, regional wall motion abnormalities.

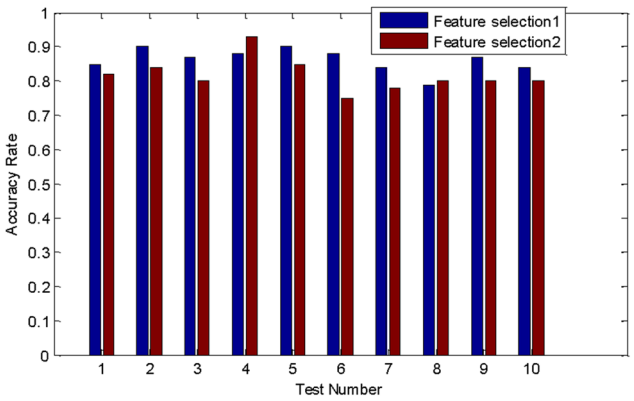

FIGURE 4 The rules' accuracy of class c_1 for different sample sets of training and test sets

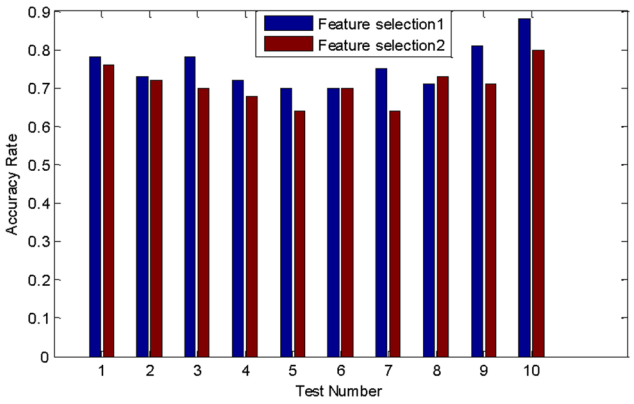

FIGURE 5 The rules' accuracy of class c_2 for different sample sets of training and test sets

our result emphasizes the importance of this rule as a risk factor for CAD disease. According to the common features between FS1 and FS2, some other effective factors include Qwave, diabetes mellitus, hypertension, RegionRWMA, Tinversion, and erythrocyte sedimentation rate. This result is approved by the discovered rules in Tables 3 to 6. For example, people with diabetes mellitus or thyroid disease have the probability of having CAD according to the Rules 8 and 10 in Table 6. Hypertension is another prevalent risk factors of CAD among the most important ones.

TABLE 7 Algorithm accuracy for two classes and two feature selections

Method	Class	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Feature selection 1	class c_1	70	70	91	84
	Class c_2	98	63	98	81
Feature selection 2	class c_1	88	50	92	90
	Class c_2	99	60	99	82

TABLE 8 Comparison of the accuracy provided by the proposed particle swarm optimization for the Z-Alizadeh Sani data with the accuracy results available in the literature

Study	Model	Accuracy (%)
Alizadehsani et al. (2012)	SMO	82.16
Alizadehsani et al. (2013)	Bagging-C4.5	79.54 (LAD)
		61.46 (LCX)
		68.96 (RCA)
Alizadehsani et al. (2016)	Combined information gain for all arteries-SVM	86.14 (LAD)
		83.17 (LCX)
		83.50 (RCA)
Babi c et al. (2017)	SVM	86.67
Hu et al. (2019)	Var- IBLMM	
Zhang et al. (2018)		88.95 ± 3.84
	Exp-CRBM	
Proposed method	Hybrid PSO	84.25

Abbreviations: Exp-CRBM,extend correlation restricted Boltzmann machine; LAD, left anterior descending; LCX, left circumflex; PSO, particle swarm optimization; RCA, right coronary artery; SMO, sequential minimal optimization; SVM, support vector machine; Var- IBLMM, variational finite inverted beta-Liouville mixture model.

This feature has been repeated in many rules in Tables 4 and 6 of class c_2 or patient class, with the value one, and just one rule case has zero value in the classpatients.

Our results show that the PSO with hybrid updating of particles is capable of producing acceptable rules for the given data set. Precision, sensitivity, and specificity of two different classes and for two feature selections are reported in Table 7. This table also shows the average accuracy of the algorithm for two different feature selections and for two classes separately. The comparison between the proposed PSO algorithm and other classification algorithms for the Z-Alizadeh Sani CAD data set is provided in Table 8.

7 | CONCLUSIONS AND FUTURE WORK

Rule discovery is a critical task for data mining. Producing the most effective and accurate rules in a given data set helps to obtain more accurate automated detection. As medical data set classification problems are very important classification problems, this paper focuses on applying novel hybrid PSO algorithm to the CAD data set in order to find the best rules, which are able to detect the given unknown data with high precision. We used a hybrid PSO algorithm to cover both real-valued and binary attributes, which are common in existing medical data sets. The results indicate that hybrid PSO can produce rules with average accuracy of about 85% for both classes. The accuracy of the second feature set (with 13 features) was better than the first one (with 11 features). The current study shows that our proposed methodology provides simple and understandable rules for CAD patients with significant performance. This work proposes an evolutionary algorithm (PSO), which generates rules in CAD domain with high accuracy. Our findings (high performance) can help the large population with CAD and also clinicians to arrive at accurate diagnosis.

Also, it can be observed from the literature that deep learning methods play a significant role in the classification and prediction in various domains (Faust et al., 2018; Jaganathan et al., 2019; LeCun et al., 2015; Yildirim et al., 2019; Yildirim et al., 2018; Zhang et al., 2019). The application of deep learning methods for the detection of various diseases has yielded high performance. For example, Faust et al. (2018) reviewed several studies, which applied deep learning methods on physiological signals. Yildirim et al. ((Yildirim et al., 2019)) investigated the performance of deep learning methods on polysomnogram signals to detect the sleep stages automatically. Based on this evidence, it can be argued that deep learning methods can be used for the diagnosis, prediction, and classification of various diseases and cancers. Thus, in our future works, we will collect more CAD data and then will apply different deep learning algorithms such as convolutional neural network (Lu et al., 2019; Song et al., 2019), long short-term memory (Miao et al., 2019), and recurrent neural network (Wang & Zhang, 2018).

Furthermore, we intend to optimize our proposed PSO algorithm and try other operators to update the velocity. We will also explore the possibility of using other evolutionary algorithms in rule discovery and compare their performance against PSO. Another possibility for future work is to use the whole rule set for particle encoding and to consider their interaction in the fitness.

CONFLICT OF INTEREST

None.

ORCID

Mariam Zomorodi-moghadam https://orcid.org/0000-0002-1308-3453

Pawel Pławiak https://orcid.org/0000-0002-4317-2801

REFERENCES

Abdar, M. (2015). Using decision trees in data mining for predicting factors influencing of heart disease. *Carpathian Journal of Electronic & Computer Engineering*, *8*(2).

Abdar, M., Kalhori, S.R.N., Sutikno, T., Subroto, I.M.I., & Arji, G. (2015). Comparing performance of data mining algorithms in prediction heart diseases. *International Journal of Electrical & Computer Engineering*, *5*(6), 2088–8708.

Abdar, M., & Zomorodi-Moghadam, M. (2018). Impact of patients'gender on parkinson's disease using classification algorithms. *Journal of AI and Data Mining*, *6*(2), 277–285.

Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2018). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*.

Acharya, U. R., Bhat, P. S., Iyengar, S. S., Rao, A., & Dua, S. (2003). Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *Pattern recognition*, *36*(1), 61–68.

Acharya, U. R., Sree, S. V., Ribeiro, R., Krishnamurthi, G., Marinho, R. T., Sanches, J., & Suri, J. S. (2012). Data mining framework for fatty liver disease classification in ultrasound: A hybrid feature extraction paradigm. *Medical physics*, *39*(7Part1), 4255–4264.

Alizadehsani, R., Habibi, J., Sani, Z. A., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., ... Alizadeh-Sani, F. (2013). Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. *Research in cardiovascular medicine*, *2*(3), 133.

Alizadehsani, R., Habibi, J., Sani, Z. A., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., & Bahadorian, B. (2012). Diagnosis of coronary artery disease using data mining based on lab data and echo features. *Journal of Medical and Bioengineering*, *1*(1).

Alizadehsani, R., Zangooei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., ... Nahavandi, S. (2016). Coronary artery disease detection using computational intelligence methods. *Knowledge-Based Systems*, *109*, 187–197.

Alkeshuosh, A. H., Moghadam, M. Z., Al Mansoori, I., & Abdar, M. (2017). Using PSO algorithm for producing best rules in diagnosis of heart disease. In *2017 international conference on computer and applications (icca)*, IEEE, pp. 306–311.

Apté, C., & Weiss, S. (1997). Data mining with decision trees and decision rules. *Future generation computer systems*, *13*(2-3), 197–210.

Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. *Computer methods and programs in biomedicine*, *141*, 19–26.

Asuncion, A., & Newman, D. (2007). Uci machine learning repository.

Babi c, F., Olejár, J., Vantová, Z., & Parali c, J. (2017). Predictive and descriptive analysis for heart disease diagnosis. In *2017 federated conference on computer science and information systems (fedcsis)*, IEEE, pp. 155–163.

Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and information science*, *3*(1), 180.

Baxt, W. G. (1990). Use of an artificial neural network for data analysis in clinical decision-making: The diagnosis of acute coronary occlusion. *Neural computation*, *2*(4), 480–489.

Bazan, J. G., Nguyen, H. S., Nguyen, S. H., Synak, P., & Wróblewski, J. (2000). Rough set algorithms in classification problem, *Rough set methods and applications*: Springer, pp. 49–88.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*: John Wiley & Sons.

Bigus, J. P. (1996). *Data mining with neural networks: Solving business problems from application development to decision support*: McGraw-Hill, Inc.

Casillas, J., Carse, B., & Bull, L. (2007). Fuzzy-xcs: A michigan genetic fuzzy system. *IEEE Transactions on Fuzzy Systems*, *15*(4), 536–550.

Chang, C.-D., Wang, C.-C., & Jiang, B. C. (2011). Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. *Expert systems with applications*, *38*(5), 5507–5513.

Chye, K. H., & Gerry, C. K. L. (2002). Data mining and customer relationship marketing in the banking industry. *Singapore Management Review*, *24*(2), 1–28.

Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. *Artificial intelligence in medicine*, *26*(1-2), 1–24.

Eberhart, R. C., & Kennedy, J. (1995). A new optimizer using particle swarm theory, paper presented at sixth international symposium on micromachine and human science, inst. of electr. and electron. *English Nagoya Japan*.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine*, *161*, 1–13.

Fidelis, M. V., Lopes, H. S, & Freitas, A. A. (2000). Discovering comprehensible classification rules with a genetic algorithm, pp. 805–810.

Freitas, A. A. (2003). A survey of evolutionary algorithms for data mining and knowledge discovery. In *Advances in evolutionary computing*, Springer, pp. 819–845.

Freitas, A. A (2013). *Data mining and knowledge discovery with evolutionary algorithms*: Springer Science & Business Media.

Friedl, M. A, & Brodley, C. E (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, *61*(3), 399–409.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, *19*(1), 1–67.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*: Elsevier.

Hassoon, M., Kouhi, M. S., Zomorodi-Moghadam, M., & Abdar, M. (2017). Rule optimization of boosted c5. 0 classification using genetic algorithm for liver disease prediction. In *2017 international conference on computer and applications (icca)*, IEEE, pp. 299–305.

Hemanth, D.J., & Anitha, J. (2019). Modified genetic algorithm approaches for classification of abnormal magnetic resonance brain tumour images. *Applied Soft Computing*, *75*, 21–28.

Hu, C., Fan, W., Du, J.-X., & Bouguila, N. (2019). A novel statistical approach for clustering positive data based on finite inverted beta-Liouville mixture models. *Neurocomputing*, *333*, 110–123.

Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., ... Chow, E. D. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, *176*(3), 535–548.

Kavitha, R., & Kannan, E. (2016). An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. In *2016 international conference on emerging trends in engineering, technology and science (icetets)*, IEEE, pp. 1–5.

Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *1997 ieee international conference on systems, man, and cybernetics. computational cybernetics and simulation*, *5*, IEEE, pp. 4104–4108.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in medicine*, *23*(1), 89–109.

Ksiażek, W., Abdar, M., Acharya, U. R., & Pławiak, P. (2019). A novel machine learning approach for early detection of hepatocellular carcinoma patients. *Cognitive Systems Research*, *54*, 116–127.

Lavra c, N. (1999). Selected techniques for data mining in medicine. *Artificial intelligence in medicine*, *16*(1), 3–23.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

Liu, Yu, Qin, Zheng, Shi, Zhewen, & Chen, Junying (2004). Rule discovery with particle swarm optimization, *Content computing*: Springer, pp. 291–296.

Liu, J., Zhong, W.-C., Liu, F., & Jiao, L. C. (2003). Classification based on organizational coevolutionary algorithm. *CHINESE JOURNAL OF COMPUTERS-CHINESE EDITION-*, *26*(4), 446–453.

Lu, N., Wu, Y., Feng, L., & Song, J. (2019). Deep learning for fall detection: Three-dimensional cnn combined with lstm on video kinematic data. *IEEE journal of biomedical and health informatics*, *23*(1), 314–323.

Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, *21*(2-3), 427–436.

Miao, H., Li, B., Sun, C., & Liu, J. (2019). Joint learning of degradation assessment and rule prediction for aero-engines via dual-task deep lstm networks. *IEEE Transactions on Industrial Informatics*.

Mishra, A. K. (2016). PSO- based swarm intelligence technique for multi-objective classification rule mining. *International Journal of Computer Applications*, *137*(2).

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2014). A survey of multiobjective evolutionary algorithms for data mining: Part i. *IEEE Transactions on Evolutionary Computation*, *18*(1), 4–19.

Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. In *2008 ieee/acs international conference on computer systems and applications*, IEEE, pp. 108–115.

Pławiak, P. (2018a). Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ecg signals. *Swarm and evolutionary computation*, *39*, 192–208.

Pławiak, P. (2018b). Novel methodology of cardiac health recognition based on ecg signals and evolutionary-neural system. *Expert Systems with Applications*, *92*, 334–349.

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660–674.

Shi, Y. (2001). Particle swarm optimization: Developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation (ieee cat. no. 01th8546)*, *1*, IEEE, pp. 81–86.

Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. In *1998 ieee international conference on evolutionary computation proceedings. ieee world congress on computational intelligence (cat. no. 98th8360)*, IEEE, pp. 69–73.

Song, S., Huang, H., & Ruan, T. (2019). Abstractive text summarization using lstm-cnn based deep learning. *Multimedia Tools and Applications*, *78*(1), 857–875.

Sousa, T., Neves, A., & Silva, A. (2003). Swarm optimisation as a new tool for data mining. In *Proceedings international parallel and distributed processing symposium*, IEEE, pp. 6–pp.

Sousa, T., Silva, A., & Neves, A. (2004). Particle swarm based data mining algorithms for classification tasks. *Parallel Computing*, *30*(5-6), 767–783.

Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)*, *2*(02), 250–255.

Tahyudin, I., & Nambo, H. (2017). The combination of evolutionary algorithm method for numerical association rule mining optimization. In *Proceedings of the tenth international conference on management science and engineering management*, Springer, pp. 13–23.

Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A., Esmaeily, H., ... Ghayour-Mobarhan, M. (2017). Hs-crp is strongly associated with coronary heart disease (chd): A data mining approach using decision tree algorithm. *Computer methods and programs in biomedicine*, *141*, 105–109.

Van der Merwe, DW, & Engelbrecht, A. P. (2003). Data clustering using particle swarm optimization. In *The 2003 congress on evolutionary computation, 2003. cec'03.*, *1*, IEEE, pp. 215–220.

Wang, Z., Sun, X., & Zhang, D. (2007). A PSO-based classification rule mining algorithm. In *International conference on intelligent computing*, Springer, pp. 377–384.

Wang, J., & Zhang, C. (2018). Software reliability prediction using a deep learning model based on the rnn encoder–decoder. *Reliability Engineering & System Safety*, *170*, 73–82.

Wosiak, A., & Zakrzewska, D. (2018). Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis. *Complexity*, *2018*.

Xue, B., Zhang, M., & Browne, W. N. (2012). New fitness functions in binary particle swarm optimisation for feature selection. In *2012 ieee congress on evolutionary computation*, IEEE.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm, *Feature extraction, construction and selection*: Springer, pp. 117–136.

Yildirim, O., Baloglu, U. B., & Acharya, U. R. (2019). A deep learning model for automated sleep stages classification using PSG signals. *International journal of environmental research and public health*, *16*(4), 599.

Yildirim, O., San Tan, R., & Acharya, U. R. (2018). An efficient compression of ecg signals using deep convolutional autoencoders. *Cognitive Systems Research*, *52*, 198–211.

Zhang, N., Ding, S., Liao, H., & Jia, W. (2018). Multimodal correlation deep belief networks for multi-view classification. *Applied Intelligence*, 1–12.

Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys & Tutorials*.

Zhang, J., Zhan, Z.-h., Lin, Y., Chen, N., Gong, Y.-j., Zhong, J.-h., ... Shi, Y.-h. (2011). Evolutionary computation meets machine learning: A survey. *IEEE Computational Intelligence Magazine*, *6*(4), 68–75.

AUTHOR BIOGRAPHIES

Mariam Zomorodi-moghadam, received her BSc. degree in Computer Engineering from Iran university of science and technology, in 2002 and her MSc. degree in Computer Engineering from Amirkabir university of technology (Tehran Polytechnic), in 2007. She received her PhD degree in Computer Engineering from Shahid Beheshti university, in 2014. She is currently an Assistant Professor in the Department of Computer Engineering at Ferdowsi university of Mashhad. She has authored several papers in international conferences and refereed journals. She also serves as a reviewer for several international journals and conferences. Her current research interests include data evolutionary computing, machine learning, and quantum computing.

Moloud Abdar received the bachelor's degree in computer engineering from Damghan University, Iran, in 2015. He also received the master's degree in computer science and engineering from the University of Aizu, Aizu, Japan, in 2018. Currently, he is working toward his Ph.D. degree at the University of Quebec in Montreal, Montreal (QC), Canada. He has written several papers in the fields of data mining, machine learning, and user modeling in some refereed international journals and conferences. He is also very active in several international conferences (e. g., TPC in IEEE AINA 2018-2020) and some referred international journals, such as the IEEE Access, the Applied Soft Computing, the Future Generation Computer Systems (outstanding reviewer in October 2017), the Neurocomputing (outstanding reviewer in January 2017), the Neural Computing and Applications, etc., as a Reviewer. His research interests include data mining, machine learning, ensemble learning, evolutionary algorithms and user modeling. He is a recipient of the Fonds de Recherche du Quebec-Nature et TechnologiesAward (ranked 5th among 20 candidates in the second round of selection process) in 2019.

Zohreh Davarzani, is a PHD student in Ferdowsi University of Mashhad, Iran. Her research interests include: Artificial intelligence, Soft Computing, Fuzzy System and Quantum computing. She is (i) Faculty member of Payame noor University of Tehran, Iran (ii) Iranian Fuzzy system Society. She received her MS and BS degrees from Ferdowsi University of Mashhad, Iran. She published more than 30 papers, including papers in International Journal on Artificial intelligence tools, journal of hybrid information technology and many international conferences.

Xujuan Zhou is Senior lecturer in University of Southern Queensland, Australia. She received her PhD from Queensland University of Technology, Australia in 2009. Her research interests include: Data mining and Machine learning, Health informatics, Data science including big data analysis and visualization, Information systems and management, Information retrieval/filtering, Recommender systems, Opinion mining and sentiment analysis. She received Commonwealth Innovation Connections Grant for the project titled ‘‘Prediction of Cancer Recurrence Using Innovative Machine Learning Approaches’’ in 2017 and successfully completed the project in 2019. She published papers in prestige Journals (e.g., Pattern Recognition Letter, Decision Support Systems, Artificial Intelligence Review, Journal of Clinical Epidemiology and Journal of Medical Intent Research) and conference proceedings (e.g., CIKM, PAKDD, HICSS, WI). She can be contacted via email: xujuan.zhou@usq.edu.au.

Pawel Pławiak, B.Eng., M.Sc., Ph.D., was born in Ostrowiec, Poland, in 1984. He obtained his B.Eng. and M.Sc. degree in Electronics and Telecommunications and his Ph.D. degree with honors in Biocybernetics and Biomedical Engineering at the AGH University of Science and Technology, Cracow, Poland, in 2012 and 2016, respectively. He is an Assistant Professor of the Institute of Teleinformatics, Cracow University of Technology, Cracow, Poland. He has published 15 papers inrefereed international SCI-IF journals. He is a reviewer of many prestigious and reputed journals. His researchinterests include machine learning and computational intelligence (e.g., artificialneural networks, genetic algorithms, fuzzy systems, support vector machines,k-nearest neighbors, and hybrid systems), ensemble learning, deep learning,evolutionary computation, classification, pattern recognition, signal processing and analysis, data analysis and data mining, sensor techniques, medicine, biocy-bernetics, and biomedical engineering. Detail and most up-to date information is available at ResearchGate and Google Scholar.

U. Rajendra Acharya received the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India, and the D.Eng. degree from Chiba University, Japan. He is currently a Senior Faculty Member with Ngee Ann Polytechnic, Singapore. He is also an Adjunct Professor with Taylor's University, Malaysia, an Adjunct Faculty with the Singapore Institute of Technology-University of Glasgow, Singapore, and an Associate Faculty with the Singapore University of Social Sciences, Singapore. He has published more than 400 papers, in refereed international SCI-IF journals (345), international conference proceedings (42), books (17) with more than 25,000 citations in

Google Scholar (with h-index of 83), and ResearchGate RG Score of 47.66. He is ranked in the top 1% of the Highly Cited Researchers for the last three consecutive years (2016, 2017, and 2018) in computer science according to the Essential Science Indicators of Thomson. He has worked on various funded projects, with grants worth more than 2 million SGD. He has three patents and an editorial board member of many journals. His research interests include biomedical signal processing, biomedical imaging, data mining, visualization and biophysics for better healthcare design, delivery, and therapy. He has served as a Guest Editor for many journals.

How to cite this article: Zomorodi-moghadam M, Abdar M, Davarzani Z, Zhou X, Pławiak P, Acharya UR. Hybrid Particle Swarm Optimization for Rule Discovery in the Diagnosis of Coronary Artery Disease. *Expert Systems*. 2021;38:e12485. https://doi.org/10.1111/exsy.12485