computer methods and programs in biomedicine 112 (2013) 92–103

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm

*Fayssal Beloufa**, *M.A. Chikh*

Biomedical Engineering Laboratory, Tlemcen University, Algeria

ARTICLE INFO

Article history:
Received 1 April 2012
Received in revised form
15 July 2013
Accepted 17 July 2013

Keywords:
Artificial Bee Colony
Interpretable classification
Fuzzy rules
Diabetes disease

ABSTRACT

In this study, diagnosis of diabetes disease, which is one of the most important diseases, is conducted with artificial intelligence techniques. We have proposed a novel Artificial Bee Colony (ABC) algorithm in which a mutation operator is added to an Artificial Bee Colony for improving its performance. When the current best solution cannot be updated, a blended crossover operator (BLX-α) of genetic algorithm is applied, in order to enhance the diversity of ABC, without compromising with the solution quality. This modified version of ABC is used as a new tool to create and optimize automatically the membership functions and rules base directly from data. We take the diabetes dataset used in our work from the UCI machine learning repository. The performances of the proposed method are evaluated through classification rate, sensitivity and specificity values using 10-fold cross-validation method. The obtained classification rate of our method is 84.21% and it is very promising when compared with the previous research in the literature for the same problem.

1. Introduction

Diabetes is a complex and complicated disease characterized by either lack of insulin or a resistance to insulin, a hormone which is crucial for metabolism of blood sugar. In a healthy person, the pancreas produces insulin to help metabolize sugar in the blood and maintain blood glucose (sugar) levels within their normal range. Diabetics are unable to produce insulin or are resistant to insulin, and consequently cannot remove glucose from the bloodstream. Whether there is inadequate insulin or insulin resistance, glucose levels in the blood increase and cause severe health problems. The classification of diabetes includes two main clinical classes: Type 1 diabetes (previously called “juvenile diabetes”) is an autoimmune disorder in which the insulin producing beta cells are destroyed by the body’s immune system. As a result the body is unable to produce insulin, Type 2 diabetes (previously called “adult onset diabetes”); in this type insulin is produced in insufficient amounts and/or cannot be used by the body to control blood sugar levels [1]. Therefore the need to detect and treat diabetes becomes obvious to reduce its incidence and costly associated metabolic disease. For this reason, in recent times, many machine learning techniques have been considered to design automatic diagnosis system for diabetes. This paper specifically focuses on the use of fuzzy modeling method to detect medical problems which relies on discovering human comprehensible knowledge. Fuzzy logic is originally proposed by Zadeh [2], who aimed to improve a classification and decision support systems by using fuzzy sets to define the overlapping class definitions. The application of fuzzy “if-then” rules also improves the interpretability of the results and provides more insight cognition into the classifier structure and the decision making process [3]. The performance of fuzzy classifier system depends on the “if-then” rules and their numbers that are generated from numerical data or human experiences. More

* *Corresponding author*. Tel.: +213 670160191.
E-mail addresses: beloufa.fayssal@gmail.com (F. Beloufa), mea_chikh@mail.univ-tlemcen.dz (M.A. Chikh).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.07.009

rules might enhance the classification result but increase the computation load. Many intelligent methods, such as heuristic approaches [4], neural networks [5–7] and clustering method [8–10] have been proposed for building optimal fuzzy classifiers but in recent years Evolutionary Algorithms (EAs) have been widely used to optimize fuzzy classifiers. In the literature several EAs like Genetic Algorithm (GA), particle swarm optimization (PSO) and ant colony optimization (ACO) have been proposed to produce fuzzy classification system. Ganji et al. [11] used ACO algorithm to generate fuzzy classification rules called FSC-ANTMINER. They have used artificial ants in order to explore the search space and gradually make candidate fuzzy rules. Other works have adopted GA to optimize fuzzy classifiers [12–14]. In [15] Sanie et al. used a hybrid genetic algorithm to produce fuzzy rules and boosted it with an Ant Colony Optimization (ACO) heuristic based on local search to improve the quality of their final classification system. In [16] the authors proposed a fuzzy modeling framework able to generate automatically a rule base through a two stage genetic based search. On the other hand, PSO based approaches have proved their efficiency when optimizing fuzzy classification system. In [17] Sousa et al. proposed a first application of PSO as a new tool for classification rule discovery. Holden and Freitas [18] used a hybrid Particle Swarm Optimization and Ant Colony Optimization algorithm for discovering classification rules in data mining. Rani and Deepa [19] proposed a particle swarm optimization approach for optimal design of fuzzy classifier called PSOFLC. Experiments are performed on IRIS dataset, where the proposed technique is compared to two well-known classification techniques, including Genetic Fuzzy Classifier and Gaussian Fuzzy Classifier. The framework used for designing the fuzzy model from the available input–output data through PSO algorithm was proposed in [20] and the detailed encoding method was also provided. Khosla et al. [21] compared the computational efforts of PSO and GA using the similar method in [20]. Recently, a new Artificial Bee Colony algorithm (ABC) [22] was proposed as an alternative of the traditional Evolutionary Algorithms. The ABC algorithm is very simple and flexible when compared to other swarm based algorithms such as Particle swarm optimization. It does not require external parameters like mutation and crossover rates, which are hard to be defined in prior. The algorithm combines local search method with global one and tries to reach a certain balance between exploration and exploitation [22]. The ABC algorithm has been successfully applied to a wide range of applications such as clustering analysis [23], neural network training [24], function optimization [25,26], Wireless Sensor Network [36,37] and Symbolic Regression [38].

In this paper, some modifications of a standard Artificial Bee Colony are introduced. The modified ABC differs from the original one in the sense that if there is no improvement in fitness function, the positions are mutated using blended crossover operator of genetic algorithm for more exploitation and exploration. This modified version of Artificial Bee Colony is proposed as a new tool for building a compact fuzzy rule based classifier without any a priori knowledge. According to our experiments, our model produces compact fuzzy rules based classifier and can work efficiently for diabetes diseases. In addition, the results obtained in this paper show that the proposed technique outperforms other methods in term of classification rate. The rest of this paper is organized as follows: Section 2 introduces the related works on diabetes disease diagnosis. In Section 3 the Fuzzy rule based classification model, Artificial Bee Colony algorithm are described. We discuss about our proposed method in detail in Section 4. The results are presented and discussed in Section 5, and finally, in Section 6, we conclude the paper.

2. Related work

Much research efforts have been performed till today to examine and classify diabetes usually based on artificial intelligence techniques in literature. In [27] a generalized discriminant analysis and least square support vector machine was used for diagnosing Pima Indian diabetes. In this work the authors have reported 79.16% classification accuracy. Polat and Gunes [28] used attribute weighted artificial immune system with 10-fold cross validation method. They obtained a classification accuracy of 75.87%. Tand and Tseng [29] developed GA-based methods to estimate a weight vector of the feature vector applied in the fuzzy k-NN estimation. The accuracy was 81.6% using Binary-coded Genetic Algorithms (BGA) and 82% using Real-coded Genetic Algorithms (RGA) with 3 features. Polat and Günes [30] obtained 89.47% classification accuracies using a principal component analysis and an adaptive neuro-fuzzy inference in diabetes disease diagnostic. In Temurtas et al. [31] a multi-layer neural network and a Probabilistic Neural Network (PNN) were used for diagnosing Pima Indian diabetes. They have reported respectively 79.62% and 78.05% in terms of correct classification rate.

3. Theory

3.1. Fuzzy rules based classification model

A fuzzy classifier consists of linguistic rules which are easy to interpret by the user. The classifier is not a black box; it can be checked for plausibility. This is very important for decision support systems, where users do not accept a computer's evaluation, unless they understand why and how a certain recommendation was given. Therefore, the fuzzy classifier can be a convenient tool in the diagnostic process. The classification problem consists of m training patterns $X_p(x_{p1}, \ldots, x_{pn})(p = 1, 2, \ldots, m)$ from M classes where x_{pi} is the ith attribute value $(i = 1, 2, \ldots, n)$ of the pth training pattern. In this study the fuzzy "if-then" rules can be expressed as follows:

$$\begin{aligned} &R_j : \text{if} x_{p1} \text{ is} A_{j1} \text{ and} x_{p2} \text{ is} A_{j2} \text{ and} \ldots \text{and} x_{pn} \text{ is} A_{jn} \\ &\text{THEN} x_p(x_{p1}, \ldots, x_{pn}) \text{ belongs to class C} \quad j = 1, \ldots, N \end{aligned} \tag{1}$$

Here R_j is the label of the jth rule, $X = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set and C ($c = 1, 2, \ldots, M$) is a class label. We use the fuzzy reasoning method of the winning rule (classical approach) [32] for classifying new patterns by the rules base. The single winner rule R_w is determined for a new pattern $x_p(x_{p1}, \ldots, x_{pn})$ as:

$$w = \text{argmax}\{\mu_j(x_p)\}|j = 1, 2, \ldots, N| \tag{2}$$

where $\mu_j(X_p)$ is the degree of compatibility grade of the pattern X_p with the antecedent fuzzy set A_j ($A_{j1}, A_{j2}, \ldots, A_{ni}$) of the rule R_j. The new pattern X_p is classified as Class C_w, which is the consequent class of the winner rule R_w. In this paper Artificial Bee Colony algorithm is proposed to develop the fuzzy classifier.

3.2. *Artificial Bee Colony algorithm*

Artificial Bee Colony algorithm (ABC), proposed by Karaboga in 2005 for real-parameter optimization, is a recent optimization algorithm which simulates the foraging behavior of a bee colony [22]. The honey bees in this algorithm are categorized into three groups: employed bees, onlooker bees and scout bees. All bees that are exploiting a food source are classified as "employed". The employed bees bring loads of nectar from the food source and share information with onlooker bees which are waiting in the hive for information to be shared by dance of employed bees about the food sources. The duration of a dance is proportional to the fruitfulness of the food source currently being exploited by the dancing bee. Onlooker bees tend to choose a food source according to the probability proportional to the quality of that food source. Therefore, good food sources attract more bees than the bad ones. Scout bees search for new food sources in the vicinity of the hive. Whenever a scout or onlooker bee finds a food source, it becomes employed. In the case when a food source is exploited fully, all the employed bees associated with the food source abandon it, and may again become scouts or onlookers. Scout bees perform the job of exploration, whereas employed and onlooker bees perform the job of exploitation. The colony is equally separated into employed bees and onlooker bees. Each solution in the search space consists of a set of optimization parameters which represent a food source position. The number of employed bees is equal to the number of food sources position and the quality of this food is called "fitness value". The process of bees seeking for good food sources is applied to find the optimal solution. Let $u_i = [u_{i1}, u_{i2}, \ldots, u_{iD}]$ represent the ith position of a food source (ith *solution*) in the population, where $i = 1, 2, \ldots$, SN, SN is the population size and $j = 1, 2, \ldots, D$, D is dimension of the solution u_i; then each position is generated randomly according to Eq. (3):

$$u_{ij} = u_{\min,j} + \mathrm{rand}[0, 1] * (u_{\max,j} - u_{\min,j}) \tag{3}$$

In order to produce a new solution $S_i = [S_{i1}, S_{i2}, \ldots, S_{iD}]$ from a neighborhood of an existing solution u_i, the ABC uses the following expression:

$$s_{ij} = u_{ij} + r(u_{ij} - u_{hj}) \tag{4}$$

where $h = 1, 2, \ldots$, SN and $j = 1, 2, \ldots, D$ are randomly chosen indexes. h is different from i, r is a random number between [−1,1]. An onlooker bee chooses a position of a food source u_i with a probability p_i calculated as follows:

$$p_i = \frac{\mathrm{fit}_i}{\sum_{j=1}^{n} \mathrm{fit}_j} \tag{5}$$

Initialization
Repeat
- **(1) Employed bee stage:**
  Perform an update process for each solution in the solution population.
- **(2) Onlooker stage:**
  Randomly select solutions depending on their fitness values then perform the same update process for each selected solution.
- **(3) Scout stage:**
  Select one of the most inactive solutions, and then replace it by a new randomly generated solution.

Until (terminated condition are satisfied)

Fig. 1 – ABC algorithm.

where fit represent the quality (fitness value) of a food source u_i, when the food source cannot be improved for a predetermined number of cycles given by the parameter "limit", the bees are abandoned and they are replaced with a new food source by the scouts. After each solution S_i is produced and evaluated by the artificial bee, a greedy selection process is employed as the selection operation between the old and the candidate solution. The new solution replaces the older one if it has an equal or better fitness value than the old solution. Otherwise, the older one is retained. The basic algorithm of ABC is summarized in Fig. 1.

4. Modified ABC fuzzy classifier learning

The main objective of this paper is to develop an efficient approach that can improve medical performance for diagnosis of diabetes disease. The novelty of our approach is using Artificial Bee Colony algorithm with some modifications, in order to create an optimal fuzzy classifier with small number of rules that are simple and comprehensible by the physicians. In this work we used two fuzzy sets (Low (L) and High (H)) for each attribute with trapezoidal membership functions. These fuzzy sets are defined by six real-valued P0, P1, P2, P3, P4, and P5 as shown in Fig. 2. Where P0 and P5 are fixed and represent the minimum and the maximum value of the input variable. The other values are used for the creation of the membership functions and each one of them has its own limits such that P1 has {P0, P5}, P2 has {P0, P5}, P3 has {P1, P5}, and P4 has {P2, P5}.

4.1. *Proposed modified ABC*

In the ABC algorithm, each employed bee (onlooker bee) generates a new food source in the neighborhood of its present

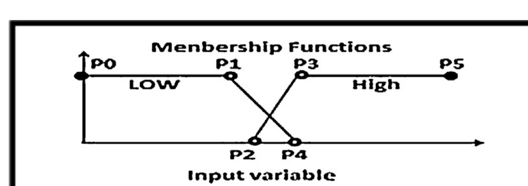


Fig. 2 – Schematic of membership functions for each input variable.

solution, so the new candidate solution can be good or bad depending on the fitness function. This means that if there is an improvement in fitness function, the new candidate solution is good and it replaces the old one in next iteration. If not, the new candidate solution is bad and converged fitness is not the optimal one. This is due to the mechanism of updating solution which is limited in providing a variety of solutions. In such a case it is necessary to introduce a new search mechanism which may help to produce a wide solutions variety that lead to mitigate the weakness of updating mechanism. This can be done by adopting a blended crossover (BLX-α) operator of genetic algorithm in ABC in order to improve the exploitation and exploration capability of the employed bees and onlooker bees. The pseudo code for generating a new solution by blended crossover (BLX-α) operator is summarized in Algorithm 1.

Algorithm 1 *(The Blended crossover operator).*

1: Choose two solutions x_1, x_2 randomly from the population
2: A value of new solution Y is generated by:

$$Y = \begin{cases} a + r * (b - a) : \text{if} x_{min} \le Y \le x_{max} \\ \text{repeat generation : otherwise} \end{cases}$$

Where:
$a = x_1 - \alpha(x_2 - x_1)$
$b = x_2 - \alpha(x_2 - x_1)$
r: random number in [0,1]
x_{min}, x_{max} represent the variable's lower and upper bounds respectively and α is a positive parameter

The proposed version of ABC differs from other versions in the sense that if there is no improvement in fitness function in the employed or onlooker stage, the positions are mutated using blended crossover operator.

4.2. Individual representation and population initialization

Each position of a food source in the ABC based method is represented in order to determine a fuzzy classification system. The codification of a fuzzy classification system designed to be evolved by modified ABC algorithm is given in Fig. 3. This figure shows the food source's representing the solutions Si to the problem. This solution comprises two types of parameter vectors $S_i = [R_i, f_i]$. One is parameter $f_i = [f^i{}_1, \ldots, f^i{}_k, \ldots, f^i{}_n]$ represented by n binary bits for the selection of features. If $f^i{}_j = 0$, the k-th input variable is excluded from the rules $R_i = R^i_1, \ldots, R^i_j, \ldots, R^i_{Mx}$; otherwise f^i_j is included ($k = 1, \ldots, n$). The other parameter vector $R_i = R^i_1, \ldots, R^i_j, \ldots, R^i_{Mx}$ consists of control vector $mf^i_{jk} \in \{1, 2\}$ for specifying the type of membership function (low, high) defined for A_{ji}, and parametric vector. The parametric vector encode four real-valued parameters $p^i_{jk,1}$, $p^i_{jk,2}$, $p^i_{jk,3}$, $p^i_{jk,4}$ that determine the points of a membership functions. After that the antecedents have been created, we have selected a suitable consequent for each antecedent from the given training patterns by the following procedure [4]:

Step 1: Calculate the compatibility of each training pattern $x_p(x_{p1}, \ldots, x_{pn})$ for the jth fuzzy rule as:

$$\mu_j(x_p) = \mu_{j1}(x_p1) * \mu_{j2}(x_p2) * \ldots * \mu_{jn}(x_{pn}) \tag{6}$$

Step 2: For each class C (C = 1...M) calculate the relative sum of the compatibility grades for the jth rule

$$\alpha_{\text{class } c}(R_j) = \sum_{x_p \in \text{class c}} \mu_j(x_p) \tag{7}$$

Step 3: Find class C for the jth rule with the maximum value of α class c

$$\alpha_{\text{class c}}(R_j) = \text{argmax}\{\alpha_{\text{class i}}(R_j) | i = 1, 2, \ldots, M\} \tag{8}$$

The label l^i_j is represented by one bit set to 0/1 means that the corresponding rule is used/not used in the rule-base encoded in the corresponding position. The control parameters f^i_{jk} and l^i_j are randomly assigned a value in {0, 1} and mf^i_{jk} is randomly assigned a value in {1, 2}. The parametric vector is randomly initialized by using the opposition based initialization method [33] in order to guarantee an initial population with certain quality and diversity.

4.3. Fitness function

In the same way as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In this paper evaluation function for fuzzy classification system is defined as follows:

$$\text{fit}(S_i) = C_R(S_i)^2 * (L - \text{Num}_R(S_i) + 1) \tag{9}$$

where S_i is the ith solution, $C_R(S_i)$ is the classification rate calculated by:

$$\text{Classification Rate} = 100 * \frac{TP + TN}{TN + TP + FN + FP} \tag{10}$$

where TP, TN, FP and FN denote respectively: true positives, true negatives, false positives and false negatives. L is the number of rules generated in the previous stage, and $\text{Num}_R(S_i)$ is the number of active rules. In this way we make sure that

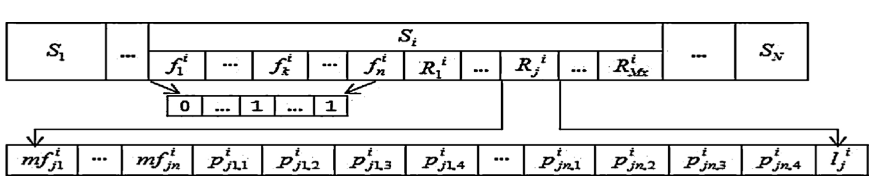


Fig. 3 – The coding structure of solution.

we keep a best fuzzy classifier with small number of rules and high Classification Rate.

4.4. *Realized algorithm*

Based on the above explanation, the pseudo-code of proposed method is given below and the flowchart is shown in Fig. 4.

Algorithm 2 *(Pseudo code of proposed method.)*.

1: Parameter initialization: Number of Solutions: SN, number of rules: M_x, number of Cycle: C. The parameters (limit, α).
2: Initialize the population of solutions u_i; $i = 1, \ldots,$ SN randomly.
3: Evaluate the fitness function fit (u_i) of the population by using Eq. (9).
4: C = 1.
5: Repeat
6: Produce new solutions s_i for the employed bees by using Eq. (4) and evaluate them (fit(s_i)) by Eq. (9).
***7: If there is no improvement in fitness function, perform** Algorithm 1 **to produce new candidate solutions s_i and evaluate them (fit(s_i)) by Eq.** (9).*
8: Apply the greedy selection process for the employed bees.
9: Calculate the probability values *Pi* for the solutions u_i by Eq. (5).
10: Produce the new solutions s_i for the onlookers from the solutions u_i selected depending on *Pi* and evaluate them (fit(s_i)) by Eq. (9).
***11: If there is no improvement in fitness function, perform** Algorithm 1 **to produce new candidate solutions s_i and evaluate them (fit(s_i)) by Eq.** (9).*
12: Apply the greedy selection process for the onlookers.
13: Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution u_i by (3).
14: Memorize the best solution achieved so far.
15: C = C + 1.
16: Until (C = maximum cycle number).

5. Experimental results

5.1. *Diabetes disease database*

The Pima Indians Diabetes database that we have used is provided from the UCI (University of California Irvine) repository of machine learning [34]. This database contains 768 patterns with 8 features that belong to two classes (Not Diabetic, Diabetic). The 8 features are: Number of times pregnant, Plasma glucose concentration, Diastolic blood pressure, Triceps skin fold thickness, Insulin, Body mass index, Diabetes pedigree function, Age. There are 500 patterns from patients who are not diabetic and 268 patterns from patients who are known to have diabetes. The reason for using this dataset is that because it is very commonly used among the other classification systems that we have used to compare this study with for Pima Indian diabetes diagnosis problem.

Table 1 – Representation of confusion matrix.

Actual	Predicted	
	Negative	Positive
Negative	TN	FN
Positive	FP	TP

5.2. *Performance measures*

The performances of the implemented classifier was evaluated by computing the percentages of sensitivity (SE), specificity (SP) and classification rate calculated by Eq. (10), the respective definitions of SE and SP are as follows:

(a) Sensitivity %:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

(b) Specificity %:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12}$$

where TP, TN, FP and FN denote respectively:

- True positives: classifies Diabetic as Diabetic.
- True negatives: classifies No Diabetic as No Diabetic.
- False positives: classifies No Diabetic as Diabetic.
- False negatives: classifies Diabetic as No Diabetic.

(c) Confusion matrix:

A confusion matrix contains information about actual and predicted classifications done by a classification system (Table 1).

5.3. *Results and discussion*

We know that the number of fuzzy if-then rules exponentially increases as the number of attributes or the number of member functions MFs increases, to cope with this problem, we propose a novel Artificial Bee Colony (ABC) algorithm that will reduce the number of rules in order to facilitate the diagnosis; in this work, each input attribute has two membership functions, so 256 fuzzy rules have been extracted logically, the overall complexity of the knowledge base of the fuzzy classifier increases substantially. To reduce this complexity, the application of ABC algorithm is proposed. In the design of fuzzy classifier, the interpretability of the rule set has been considered an important factor. This interpretability is measured by calculating the number of rules; fuzzy classifiers containing fewer fuzzy rules are always more interpretable than those with more fuzzy rules. The performance of the proposed method was tested with regard to the Classification Rate, Sensitivity, Specificity and confusion matrix. Two experiments were performed in this study. In the first experiments, the features selection has not been applied and in the second experiments the features selection was activated and features specific for each rule base were selected through the modified ABC search. The optimum parameters values of modified ABC involved in the evolutionary process are shown in Table 2.

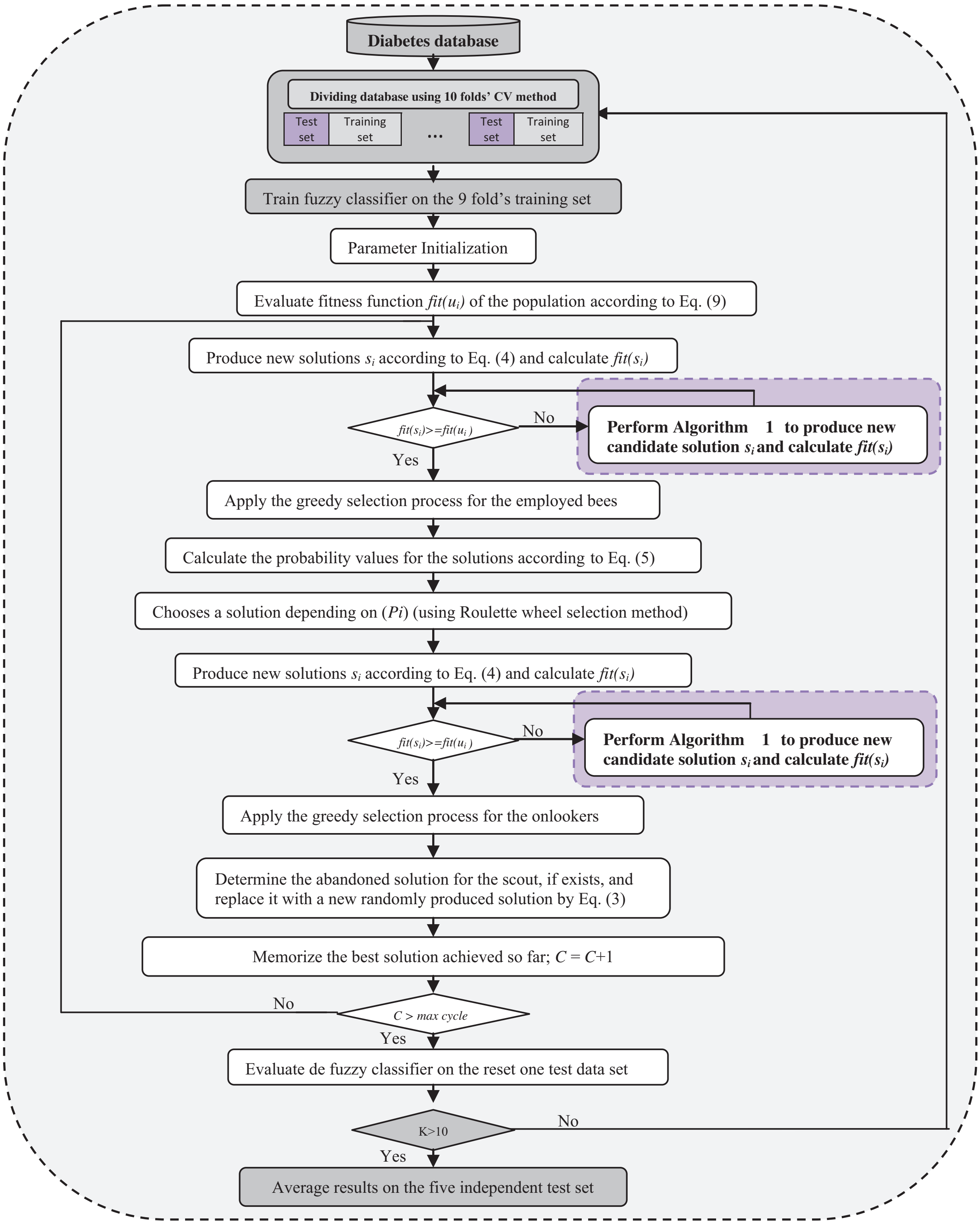


Fig. 4 – The flow chart of the proposed algorithm.

These parameters values were determined experimentally by monitoring change of the fitness function value, in order to give the best classification performance. According to these parameters values, the breakdown accuracy results with all features and optimal features found through modified ABC search in terms of Classification Rate, number of rules and number of features for all folds are shown in Table 3.

As can be seen from Table 3, fuzzy rules selected have relatively small number with an average of 12.2 rules (without feature selection). The number of features and rules used by different folds when performing feature selection varies

Table 2 – Parameters specification in proposed algorithm.

Parameters specification	Value
Number of rules in each solution "M_x"	20
Number of food source, employed and onlooker bees	100
Maximum cycle number "C"	200
Value of the parameter "*limit*"	10
Value of the parameter "α"	0.5

significantly with an average of 3.7 features and 7.1 rules. The difference in the number of rules and features between folds is due to the changing of training data from a fold to another. Table 4 gives the confusion matrix showing the classification results of this classifier on fold 2 from the diabetes diseases dataset. In this case, the classification rate obtained with modified ABC algorithm was 83.33% using all features. The modified ABC with feature selection obtains 84.61% classification rate.

In addition, as it can be seen in Table 5 which depicts the fuzzy rules of fold 2 with feature selection and all features, the fuzzy rules use understandable description of the antecedent conditions and their class labels in the form of "if-then" rules. These fuzzy rules help the physician to understand the relations between the factors or symptoms and the diagnosis outcome. To compare the performance of the proposed modified ABC, the basic ABC algorithm is developed and the performance comparison is given in Table 6. In our experiments, we used the same population number and the maximum evaluation number for each test problem. ABC algorithm produces the candidate solution from its parent by a simple operation based on taking the difference of randomly determined parts of the parent and a randomly chosen solution from the population. This mechanism of updating solution is limited in providing a variety of solutions. In the modified ABC we have introduced a new search mechanism by adopting a blended crossover (BLX-α) operator of genetic algorithm which may help to produce a wide solutions variety that lead to mitigate the weakness of updating mechanism of standard version of ABC. This means that the candidate solution is produced by Eq. (4) and if there is no improvement in fitness function we perform the Algorithm 1 (blended crossover operator) in order to improve the exploitation and exploration capability of the algorithm. As in ABC, the best solution discovered so far in modified ABC is not always held in the population since it might be replaced with a randomly produced solution by a scout. Therefore, it might not contribute to the production of trial solutions. The ABC and the modified ABC algorithm have only one control parameter (limit) apart from Colony Size and Maximum Cycle Number. As can be seen from the Table 6, the best results were obtained with the modified version of ABC with respect to the mean and maximum classification results. The best result was 85.89% and the mean result was 82.68% when using all attribute and the best result when using feature selection was 85.89% and the mean result was 84.21%. Compared with the basic ABC the modified ABC has enhanced further the mean classification rate by 3.07% (using all attribute) and 2.81% (using feature selection), on the other hand the maximum classification rate has enhanced by 3.84% (using all attribute) and 4.0% (using feature selection). The convergence performance for the second fold of the modified ABC and the Basic ABC is also studied. This is shown in Fig. 5 using feature selection and all features. From this figure, it is observed that ABC algorithm takes more number of generations to converge than Modified ABC, the ABC is unable to attain convergence before the 122th generations using all features and 99th generations using feature

Table 3 – Classification rate and number of rules and features using 5 fold cross validation method.

Folds	All features				Feature selection				
	Training data (%)	Testing data (%)	Number of rules	CPU time (s)	Training data (%)	Testing data (%)	Number of rules	Number of features	CPU time (s)
1	79.71	79.48	11	130.2	83.18	82.05	7	4	92.2
2	84.05	83.33	12	122.4	83.76	84.61	5	3	80.2
3	82.60	80.76	13	132.6	83.33	85.89	9	4	99.2
4	83.33	85.89	14	131.9	84.04	83.33	9	5	104.3
5	82.60	83.33	12	125.6	84.63	85.89	7	3	81.4
6	83.76	82.05	10	126.0	83.91	84.61	8	4	95.4
7	81.15	82.05	11	127.8	84.34	83.33	10	5	110.6
8	83.33	84.61	14	124.2	84.20	83.33	3	2	60.3
9	80.00	83.33	13	128.5	84.78	83.33	6	4	89.3
10	82.83	82.05	12	129.0	84.49	85.89	7	3	84.1
Average	82.32	82.68	12.2	127.8	84.20	84.21	7.1	3.7	89.7

Table 4 – Confusion matrices of the modified ABC algorithm.

Actual	Predicted		
	Diabetics	Non-diabetics	Method
Diabetics	22	4	Modified ABC with all features
Non-diabetics	9	43	
Diabetics	21	5	Modified ABC with feature selection
Non-diabetics	8	44	

Table 5 – Linguistic fuzzy rules of fold 1 for diabetes disease.

The rules with all features
R1: If (Npreg is H) and (Glu is L) and (BP is H) and (Skin is L) and (Insulin is L) and (BMI is H) and (PED is H) and (Age is L) then (Class is Non-diabetic)
R2: If (Npreg is L) and (Glu is L) and (BP is L) and (Skin is L) and (Insulin is L) and (BMI is L) and (PED is H) and (Age is L) then (Class is Non-diabetic)
R3: If (Npreg is H) and (Glu is L) and (BP is H) and (Skin is L) and (Insulin is L) and (BMI is H) and (PED is L) and (Age is L) then (Class is Non-diabetic)
R4: If (Npreg is L) and (Glu is L) and (BP is L) and (Skin is H) and (Insulin is L) and (BMI is L) and (PED is L) and (Age is L) then (Class is Non-diabetic)
R5: If (Npreg is H) and (Glu is L) and (BP is H) and (Skin is L) and (Insulin is H) and (BMI is L) and (PED is L) and (Age is L) then (Class is Non-diabetic)
R6: If (Npreg is L) and (Glu is L) and (BP is H) and (Skin is L) and (Insulin is L) and (BMI is L) and (PED is L) and (Age is L) then (Class is Non-diabetic)
R7: If (Npreg is L) and (Glu is L) and (BP is L) and (Skin is L) and (Insulin is L) and (BMI is L) and (PED is L) and (Age is L) then (Class is Non-diabetic)
R8: If (Npreg is H) and (Glu is H) and (BP is H) and (Skin is L) and (Insulin is H) and (BMI is L) and (PED is H) and (Age is L) then (Class is diabetic)
R9: If (Npreg is L) and (Glu is H) and (BP is L) and (Skin is L) and (Insulin is H) and (BMI is H) and (PED is H) and (Age is H) then (Class is diabetic)
R10: If (Npreg is L) and (Glu is H) and (BP is L) and (Skin is H) and (Insulin is L) and (BMI is L) and (PED is L) and (Age is L) then (Class is diabetic)
R11: If (Npreg is H) and (Glu is H) and (BP is H) and (Skin is H) and (Insulin is H) and (BMI is H) and (PED is H) and (Age is H) then (Class is diabetic)
R12: If (Npreg is H) and (Glu is H) and (BP is H) and (Skin is H) and (Insulin is L) and (BMI is H) and (PED is L) and (Age is H) then (Class is diabetic)
The rules with feature selection
R1: If (Glu is L) and (BMI is L) and (Age is L) then **(Class is Non-diabetic)**
R2: If (Glu is L) and (BMI is H) and (Age is L) then **(Class is Non-diabetic)**
R3: If (Glu is H) and (BMI is L) and (Age is L) then **(Class is diabetic)**
R4: If (Glu is H) and (BMI is H) and (Age is L) then **(Class is diabetic)**
R5: If (Glu is H) and (BMI is H) and (Age is H) then **(Class is diabetic)**

We note: L = LOW, H = HIGH, Npreg = number of times pregnant, Glu = plasma glucose concentration, BP = diastolic blood pressure, Skin = triceps skin fold thickness, BMI = body mass index, PED = diabetes pedigree function.

Table 6 – The results of comparison.

	MABC			ABC		
	Best	Worst	Average	Best	Worst	Average
All features						
Classification rate (%)	85.89	78.48	82.68	82.05	76.98	79.61
Sensitivity (%)	84.61	76.92	81.91	80.76	73.92	78.54
Specificity (%)	86.53	78.84	83.07	82.69	76.92	80.18
Feature selection						
Classification rate (%)	85.89	82.05	84.21	81.88	80.43	81.40
Sensitivity (%)	84.61	80.76	83.45	84.61	76.92	80.76
Specificity (%)	86.53	80.76	84.60	84.61	80.76	82.88

selection. Even though the proposed modified ABC takes less generations (at 88th generations when using all features and at 75th generations when performing feature selection) the blended crossover operator incorporated into them performs well in tuning the solution and classification rate is very good than the basic ABC. Some examples of the parameters used in each test problem are listed in Table 8.

The classification rate (CR) mean solutions and the standard deviation of the ABC and the MABC for 20 independent runs using these parameters are listed in Table 9.

From the Table 9 we can see that the best means are obtained by MABC. When the number of rules in each solution is 20, the number of food source, the employed and onlooker bees is 100, the parameters limit and α have a value of 10,

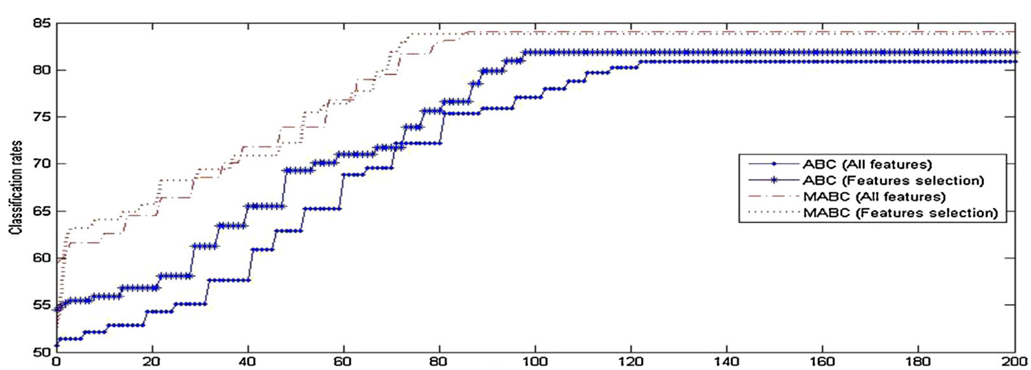
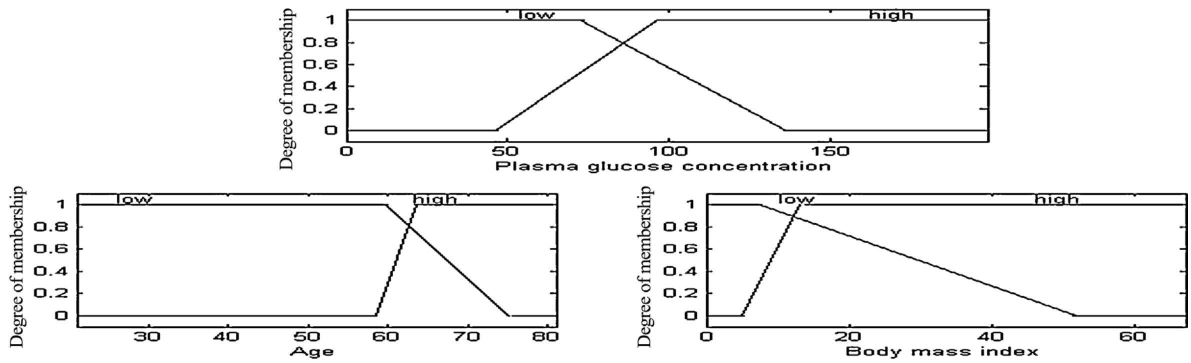
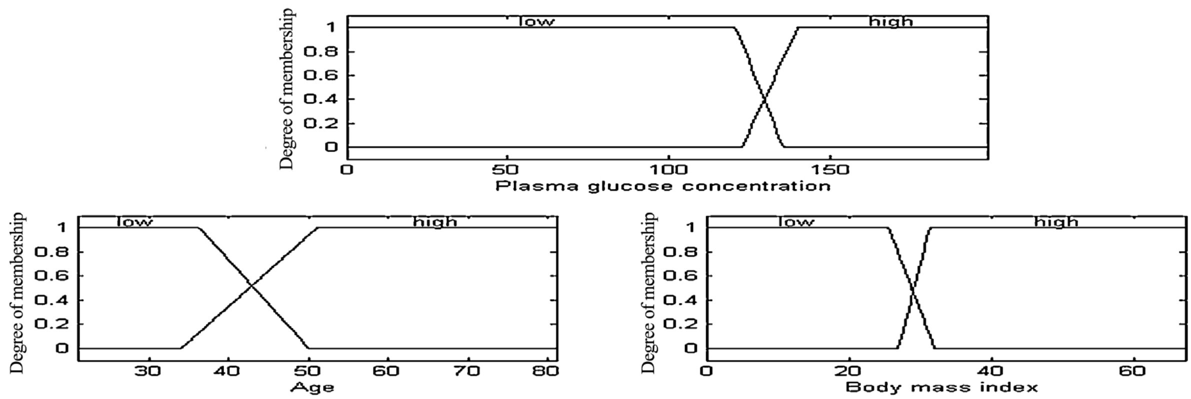


Fig. 5 – Convergence of modified ABC and ABC during learning for diabetes database (fold2).



The initialization of the modal points in the first iteration



Optimal modal points obtained after 200 iterations

Fig. 6 – Optimal membership functions.

0.5 respectively, the MABC algorithm achieved respectively a classification rate mean of 78.28% with standard deviation of 4.66 (using feature selection), and a classification rate mean of 77.05% with standard deviation of 4.59 (using all features). In the same condition ABC algorithm achieved respectively just a classification rate mean of 73.59% with standard deviation of 5.68 (using feature selection), and a mean of 73.22% with standard deviation of 6.32 (using all features). Therefore, it can be noticed form the results that MABC algorithm is more efficient than ABC, and the obtained final solution of MABC algorithm is better than that of ABC in most cases.

The adjustment of the location of membership functions is an important issue. If the modal points of the membership functions are not properly adjusted, this can affect the classification results, for this reason the objective of our work is the adjustment of the modal points. For membership functions, standard position updating mechanism of ABC is applied. For each point of membership functions (bee's position) in the population an adaptable position is randomly initialized. During each iteration, new membership functions points are created according to the Eqs. (3) and (4). The selected features in the second fold are (Plasma glucose concentration, Age and Body mass index), the initialization of the modal points in the first iteration and the optimal membership functions (optimal modal points) obtained after 200 iterations for these features are shown in Fig. 6.

At the same time, the results obtained for other classifiers from literature are shown in Table 7. In [28] the authors used fuzzy classifier for solving the problem of the resources allocation mechanism, they obtained 84.42% but this classifier is not interpretable (black box type) knowing that interpretability is often required in medical diagnosis. The same authors have used adaptive neuro-fuzzy inference system to diagnosis of diabetes disease in [30], they obtained 89.47% but there is no explanation and discussion about the fuzzy rules, their

Table 7 – Classification rate obtained with our proposed system and other classifiers from literature.

Methods	Accuracy (%)	Study
Fuzzy-AIRS [28] (10xCV)	84.42	[28]
Neuro-fuzzy inference system [30] (10xCV)	89.47	[30]
GDA_LSSVM [27] (10xCV)	79.16	[27]
Fuzzy modeling [35] (10xCV)	77.65	[35]
BGA-fuzzy-KNN [29] (5xCV)	81.6	[29]
RGA-fuzzy-KNN [29] (5xCV)	82	[29]
ML-NN [31] (10xCV)	79.62	[31]
PNN [31] (10xCV)	78.05	[31]
GA-fuzzy classifier [16] (10xCV)	71.49	[16]
ACO-fuzzy classifier [11] (10xCV)	84.24	[11]
MABC (all features)	82.68	Our study
MABC (feature selection)	84.21	
ABC (all features)	79.61	
ABC (feature selection)	81.40	

Table 8 – Parameters specification for each test problem in proposed algorithm.

Test	Number of rules in each solution "Mx"	Number of food source, employed and onlooker bees	Maximum cycle number "C"	Value of the parameter "limit"	Value of the parameter "α"
1	20	200	200	20	0.1
2	20	50	200	5	0.4
3	20	100	200	10	0.7
4	20	100	200	10	0.5
5	20	100	200	20	0.9

Table 9 – The classification rate mean and standard deviation of 20 independent runs for each test problem.

Test	ABC				MABC			
	Feature selection		All features		Feature selection		All features	
	Mean (%)	SD	Mean (%)	SD	Mean (%)	SD	Mean (%)	SD
01	72.72	2.98	71.44	6.43	76.75	4.92	74.88	3.81
02	**73.47**	**5.54**	**72.77**	**7.24**	**77.90**	**5.40**	**76.04**	**4.35**
03	70.01	5.20	71.65	6.72	76.54	6.81	73.84	5.36
04	**73.59**	**5.68**	**73.22**	**6.32**	**78.28**	**4.66**	**77.05**	**4.59**
05	70.05	5.13	69.21	6.53	75,99	7.87	76.22	4.02

number, and the type of membership functions used. Our paper focuses on the interpretability of obtained results which means that each obtained class is justified by an activated fuzzy rule. Ganji's et al. in [11] have used Ant colony optimization (ACO) algorithm to create an optimal fuzzy classifier, this algorithm has multiple control parameters which are difficult to adjust them. It is known that these control parameters affect the performance of algorithms significantly. In our study we have used ABC algorithm that is simple, flexible and easy to implement, it has only one control parameter. By applying this technique the number of rules has been reduced to an average of 7.1 rules instead of 15 rules that are obtained in [11] by using ACO algorithm. With an average of 7.1 rules the doctor will not find any difficulties to read knowledge base. We can see also from this Table that the classification rates obtained by modified ABC outperform the basic ABC and other classifiers reported in the literature. The obtained fuzzy rules by our proposed approach are closer to the medical expert reasoning. The ABC-fuzzy classifier justifies its results by the fuzzy rules. So the expert can easily check the fuzzy model classifier for plausibility, and can verify why a certain classification result was obtained for a certain patient, by checking the degree of fulfillment of the individual fuzzy rules. Instead of a simple classification the physician also gets a description of the patients in terms of the fuzzy rules that are active for this case (Tables 8 and 9).

To illustrate more the effectiveness of the proposed approach two other data sets are tested. We have used breast cancer and IRIS database taken from UCI (University of California Irvine) repository of machine learning as medical database. The breakdown accuracy results of proposed method on breast cancer and IRIS databases in terms of Classification Rate, number of rules, number of features for all folds are shown in Table 10. The breast cancer database [39] originally contained 699 cases and 9 attributes, but 16 of these are missing class values, so they were discarded leaving only 683 case. The used dataset consists of 2 classes where 444 cases were benign and 239 cases were malignant. The IRIS database

Table 10 – Classification rate and number of rules and features using 10 fold cross validation method.

Fold	Breast cancer database							Iris database		
	All features			Feature selection				All features		
	Training data (%)	Testing data (%)	Num of rules	Training data (%)	Testing data (%)	Num of rules	Num of feature	Training data (%)	Testing data (%)	Num of rules
1	99.18	98.55	8	98.69	98.55	7	5	97.77	100	6
2	98.69	97.10	9	98.04	98.55	6	3	98.51	100	7
3	98.37	98.55	11	98.37	97.10	5	4	97.03	93.33	4
4	98.04	98.55	13	99.18	98.55	8	7	97.77	100	6
5	97.71	97.10	12	98.37	98.55	5	3	97.77	100	6
6	98.20	98.55	11	99.18	97.10	6	4	97.03	93.33	5
7	98.37	97.10	9	98.85	98.55	5	3	98.51	100	7
8	98.85	97.10	14	99.02	98.55	6	4	99.25	100	6
9	99.02	97.10	10	97.71	97.10	9	6	97.77	100	7
10	97.71	98.55	11	98.04	98.55	4	3	98.51	100	6
Average	98.41	97.82	10.8	98.54	98.11	6.1	4.2	97.99	98.66	6.0

[40] contains 150 patterns with four features that belong to three classes (Iris Setosa, Iris Versicolour, and Iris Virginica), each class has 50 patterns. From Table 10 we can see that the obtained classification accuracy of our proposed method for breast cancer database is 97.82% with an average of 10.8 rules (without features selection), and 98.11% with an average of 6.1 rules (and 4.2 features when performing features selection). For IRIS database, our proposed method is tested using all features because the feature selection is not necessary when the database have a few input variables (just four features). Basing on the results presented in Table 10, it can be summarized that the proposed method generates automatically a knowledge base with an average of 6 rules to justify the classification and an accuracy of 98.66%. It is shown from these results that our proposed method can also support the decision making process of the physician and provide further information using a few active fuzzy rules.

Experimental results have shown that the proposed approach with Fuzzy-ABC is simple and effective in explaining the interpretability of fuzzy classifier by reducing the 256 rules to an average of 12.2 efficient rules when using all attribute and an average of 7.1 efficient rules using features selection while preserving the model accuracy at a satisfactory level. The results obtained in this paper are very interesting and can be used confidently to help clinician for decision making in their diagnosis problems. Therefore, we can conclude that the modified Artificial Bee Colony could be a very effective in building fuzzy rule base classifier for medical decision support system.

6. Conclusion

This paper presents an efficient and reliable modified ABC algorithm for diabetes disease. The proposed modified ABC has been used as an evolutionary algorithm to create an optimal fuzzy classifier. We explain in this study how tuning both the optimal rules base and the membership functions simultaneously with high readability and accuracy. The fuzzy classifier built with modified ABC approach can also support the decision-making process of the physician and provides further information using a few active fuzzy rules. The effectiveness of the proposed algorithm has been demonstrated through Pima Indian diabetes dataset. According to obtained results, fuzzy method with modified ABC algorithm could be a powerful tool for diagnosis of diabetes disease.

REFERENCES

[1] National Diabetes Information Clearinghouse (NDIC), http://diabetes.niddk.nih.gov/dm/pubs/d-iagnosis
[2] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.
[3] J. Valente, Semantic constraints for membership function optimization, IEEE Transactions on Fuzzy Systems 19 (1999) 128–138.
[4] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and System 52 (1) (1992) 21–32.
[5] D. Chakraborty, N.R. Pal, A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification, IEEE Transactions on Neural Networks 15 (1) (2004) 110–123.
[6] C.T. Lin, C.S.G. Lee, Neural network-based fuzzy logic control and decision system, IEEE Transactions on Computers 40 (12) (1991) 1320–1336.
[7] D. Nauck, R. Kruse, A neuro-fuzzy method to learn fuzzy classification rules from data, Fuzzy Sets and Systems 89 (3) (1997) 277–288.
[8] S. Abe, R. Thawonmas, A fuzzy classifier with ellipsoidal regions, IEEE Transactions on Fuzzy Systems 5 (3) (1997) 358–368.
[9] C.Y. Lee, C.J. Lin, H.J. Chen, A self-constructing fuzzy CMAC model and its applications, Information Sciences 177 (1) (2007) 264–280.
[10] J.A. Roubos, M. Setnes, J. Abonyi, Learning fuzzy classification rules from labelled data, Information Sciences 150 (1/2) (2003) 77–93.
[11] M.F. Ganji, M.S. Abadeh, A fuzzy classification system based on ant colony optimization for diabetes disease diagnosis, Expert System with Application 38 (2011) 14650–14659.
[12] R. Alcal, J. Alcala-Fdez, F. Herrera, J. Otero, Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation, International Journal of Approximate Reasoning 44 (1) (2007) 45–64.
[13] Y.C. Hu, Finding useful fuzzy concepts for pattern classification using genetic algorithm, Information Sciences 175 (1/2) (2005) 1–19.
[14] E. Zhou, A. Khotanzad, Fuzzy classifier design using genetic algorithms, Pattern Recognition 40 (12) (2007) 3401–3414.
[15] M.S. Abadeh, J. Habibi, E. Soroush, Induction of fuzzy classification systems using evolutionary ACO-based algorithms, in: Proceedings of the First Asia International Conference on Modeling and Simulation (AMS 07) (IEEE), 2007.
[16] J. Guzaitis, A. Verikas, A. Gelzinis, M. Bacauskiene, A framework for designing a fuzzy rule-based classifier, in: F. Rossi, A. Tsoukias (Eds.), Proceedings of the 1st International Conference on Algorithmic Decision Theory (ADT 09) Venice, Italy, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 434–445.
[17] T. Sousa, A. Silva, A. Neves, Particle swarm based data mining algorithms for classification tasks, Parallel Computing 30 (5/6) (2004) 767–783.
[18] N. Holden, A. Freitas, A hybrid particle swarm/ant colony algorithm for the classification of hierarchical biological data, in: Proceedings of the conference companion on Genetic and evolutionary computation (GECCO 07), ACM, New York, NY, USA, 2007, pp. 2745–2750.
[19] C. Rani, S.N. Deepa, Design of optimal fuzzy classifier system using particle swarm optimization, in: International Conference on Innovative Computing Technologies (ICICT 10) (IEEE), 2010, pp. 1–6.
[20] A. Khosla, S. Kumar, K.K. Aggarwal, A framework for identification of fuzzy models through particle swarm optimization, in: IEEE Indicon Conference, Chennai, India, 2005, pp. 388–391.
[21] A. Khosla, S. Kumar, K.R. Ghosh, A comparison of computational efforts between particle swarm optimization and genetic algorithm for identification of fuzzy models, in: Annual Conference of the North American Fuzzy Information Processing Society (IEEE), 2007, pp. 245–250.
[22] D. Karaboga, An idea based on honey bee swarm for numerical optimization, Technical Report TR06, Computer Engineering Department, Erciyes University, Turkey (2005).
[23] D. Karaboga, C. Ozturk, A novel clustering approach: artificial bee colony (ABC) algorithm, Applied Soft Computing 1 (2011) 652–657.

[24] D. Karaboga, C. Ozturk, Neural networks training by artificial bee colony algorithm on pattern classification, Neural Network World 19 (2009) 279–292.
[25] D. Karaboga, B. Basturk, Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems, Lecture Notes in Computer Science 4529 (2007) 789–798.
[26] G. Wei-Feng, L. San-Yang, A modified artificial bee colony algorithm, Computers and Operations Research 39 (2012) 687–697.
[27] K. Polat, S. Gunes, A. Arslan, A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square suport vector machine, Expert Systems with Applications 34 (1) (2008) 482–487.
[28] K. Polat, S. Gunes, An improved approach to medical data sets classification: artificial immune recognition system with fuzzy resource allocation mechanism, Expert Systems with Applications 24 (4) (2007) 252–270.
[29] P.H. Tand, M.H. Tseng, Medical data mining using BGA and RGA for weighting of features in fuzzy K-NN classification, in: Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 2009, pp. 3070–3075.
[30] K. Polat, S. Günes, An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, Digital Signal Processing 17 (4) (2007) 702–710.
[31] H. Temurtas, N. Yumusak, F. Temurtas, A comparative study on diabetes disease diagnosis using neural networks, Expert Systems with Applications 36 (4) (2009) 8610–8615.
[32] O. Cordón, M.J. Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, International Journal of Approximate Reasoning 20 (1999) 21–45.
[33] H. Tizhoosh, Opposition-based learning: a new scheme for machine intelligence, in: Proceedings on International Conference on Computational Intelligence for Modeling, Control and Automation, 2005, pp. 695–701.
[34] A. Frank, A. Asuncion, UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, 2010, http://archive.ics.uci.edu/ml.
[35] S.N. Ghazavi, T.W. Liao, Medical data mining by fuzzy modeling with selected features, Artificial Intelligence in Medicine 43 (3) (2008) 195–206.
[36] C. Ozturk, D. Karaboga, B. Gorkemli, Probabilistic dynamic deployment of wireless sensor networks by artificial bee colony algorithm, Sensors 11 (6) (2011) 6056–6065.
[37] D. Karaboga, S. Okdem, C. Ozturk, Cluster Based Wireless Sensor Network Routing using Artificial Bee Colony Algorithm, Wireless Networks, Springer, The Netherlands 18 (7) (2012) 847–860.
[38] D. Karaboga, C. Ozturk, N. Karaboga, B. Gorkemli, Artificial bee colony programming for symbolic regression, Information Sciences 209 (2012) 1–15.
[39] J. Merz, P.M. Murphy, UCI Repository of Machine Learning Databases, 1996, http://www.ics.uci.edu/leam/MLRepository.htrnl.
[40] D.J. Newman, S. Hettich, C.L. Blake, P.M.C.J. Merz, UCI Repository of Machine Learning Databases, 1998, http://www.ics.uci.edu/leam/MLRepository.htrnl.