Development of a Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System for diabetes diagnosis

Fatemeh Mansourypoor, Shahrokh Asadi*

Faculty of Engineering, Farabi Campus, University of Tehran, Iran

ARTICLE INFO

Keywords:
Reinforcement Learning
Evolutionary
Diabetes diagnosis
Fuzzy Rule-Based
Genetic Algorithm

ABSTRACT

The early diagnosis of disease is critical to preventing the occurrence of severe complications. Diabetes is a serious health problem. A variety of methods have been developed for diagnosing diabetes. The majority of these methods have been developed in a black-box manner, which cannot be used to explain the inference and diagnosis procedure. Therefore, it is essential to develop methods with high accuracy and interpretability. In this study, a Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System (RLEFRBS) is developed for diabetes diagnosis. The proposed model involves the building of a Rule Base (RB) and rule optimization. The initial RB is constructed using numerical data without initial rules; after learning the rules, redundant rules are eliminated based on the confidence measure. Next, redundant conditions in the antecedent parts are pruned to yield simpler rules with higher interpretability. Finally, an appropriate subset of the rules is selected using a Genetic Algorithm (GA), and the RB is constructed. Evolutionary tuning of the membership functions and weight adjusting using Reinforcement Learning (RL) are used to improve the performance of RLEFRBS. Moreover, to deal with uncovered instances, it makes use of an efficient rule stretching method. The performance of RLEFRBS was examined using two common datasets: Pima Indian Diabetes (PID) and BioSat Diabetes Dataset (BDD). The experimental results show that the proposed model provides a more compact, interpretable and accurate RB that can be considered to be a promising alternative for diagnosis of diabetes.

1. Introduction

Diabetes mellitus is one of the most common diseases in the world. According to the International Diabetes Federation, over 415 million individuals suffered from this disease globally in 2015 [1]. This amount is estimated to reach 642 million in 2040. Diabetes is a metabolic disorder where the body is unable to properly produce or respond to insulin which is required for regulating glucose (sugar). There are two types of diabetes: type 1, which is primarily caused by autoimmune pancreatic β-cell destruction and characterized by absolute insulin deficiency, and type 2, which is characterized by insulin resistance and relative insulin deficiency.

Besides contributing to heart disease, diabetes also increases the risk of developing kidney disease, blindness, nerve damage and blood vessel damage. The exact cause of the disease remains unknown [2], however, a combination of factors including genetics, environment and lifestyle have been shown to influence the likelihood of developing this disease [3]. The symptoms of the disease include frequent urination, thirst, weight loss, extreme fatigue, blurred vision, and prolonged wound healing. Common methods of diagnosing diabetes are the A1C or HbA1C test, the Fasting Blood Plasma (FBP) test, the Fasting Plasma Glucose (FPG) test and the Oral Glucose Tolerance Test (OGTT). In these mentioned tests, a diagnosis of diabetes is made based on the patient's blood sample. For instance, in the OGTT, after nearly 8 h of fasting, the patient has a high-glucose drink and 2 h later, a blood sample is drawn to determine its glucose level.

Although the occurrence of diabetes is rapidly increasing, existing methods are not capable to eliminate the problem; nevertheless, if it is detected in its early stage, the disease can be controlled. The main problem with this disease lies in its incorrect or late diagnosis which complicates treatment efforts. Therefore, considerable amount of efforts should be made so as to develop a more effective diagnosis method such that diabetes can be controlled in its early stage.

As a technique for diagnosing diseases and classifying patients, data mining aids in discovering patterns that are extremely useful for faster detection of diabetes and its negative effects. Medical data mining refers to the process of discovering hidden patterns and information in large medical datasets, analyzing the data and using the results to diagnose the diseases [4]. The process plays a vital role in reducing the unnecessary
costs associated with various medical tests [5]. Numerous data mining techniques have been used to diagnose several diseases so far including diabetes [6], stroke [7], cardiovascular problems [8] and cancer [9].

With respect to diabetes, as shown in Table 1, a variety of data mining and machine learning methods have been proposed. However, the majority of these models are black-box, the detection and inference mechanisms which cannot be explained. Thus, it is essential to use those models that have interpretability. Fuzzy Rule-Based Systems (FRBSs) have high interpretability [10,11], which makes them greatly suitable for medical diagnosis.

Fuzzy logic was first introduced by Dr. Lofti Zadeh in 1969 [12] who predicted that “…fuzzy sets are likely to be applied in the field of medical diagnostics …” [13]. In recent years, the growing popularity of fuzzy set theory in medical applications has proven the correctness of this prediction. FRBSs are including a series of interpretable if-then rules. For example, the following rule indicates that if the plasma glucose and Body Mass Index (BMI) are very high, the patient has diabetes:

\[
\text{If Plasma glucose is very high and Body mass index is very high Then Diabetes.}
\]

As shown in Table 1, fuzzy logic has been previously used in diagnosing diabetes. Ganji et al. [14] proposed a fuzzy ant colony optimization technique for diagnosing diabetes by extracting a set of fuzzy if-then rules. An adaptive neuro fuzzy inference system was designed by KRhoca et al. [15] to estimate the risk of developing diabetes using several input parameters including age, blood cholesterol, gender and body shape. Beloufa et al. [16] developed a new algorithm by adding a mutation operator to the artificial bee colony algorithm. The modified algorithm offers the new performance measure that is superior compared to the original and used to automatically create and optimize membership functions and rules that can be directly obtained from the data. The classification method proposed by Gardaras et al. [17] processes datasets in order to determine fuzzy borders for the classes of the problem. Chang et al. [18] also proposed an evolutionary approach for directly creating a fuzzy classifier system from the data without any assumptions regarding prior knowledge about the distribution of the data.

As mentioned earlier, accurate and early diagnosis of diabetes is critical in preventing the deadly side effects of this disease. This requires accurate data as well as identification of the suitable features for analysis. Physicians are often faced with an overwhelming amount of information, which hinders rapid and precise decision-making. Under such circumstances, the prerequisites of decision support systems (DSSs) are intensified. In addition to the precision, ease of use must be considered in Medical Decision Support Systems (MDSSs). Several problems exist in current systems including inaccurate diagnosis, lack of interpretability and rule inconsistencies.

In this paper, a Reinforcement Learning-based Evolutionary Fuzzy

<table>
<thead>
<tr>
<th>Author et al.</th>
<th>Year</th>
<th>Ref</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanker et al.</td>
<td>1996</td>
<td>[19]</td>
<td>✓</td>
</tr>
<tr>
<td>Park et al.</td>
<td>2001</td>
<td>[20]</td>
<td>✓</td>
</tr>
<tr>
<td>Chang et al.</td>
<td>2004</td>
<td>[18]</td>
<td>✓</td>
</tr>
<tr>
<td>Goncalves et al.</td>
<td>2006</td>
<td>[21]</td>
<td>✓</td>
</tr>
<tr>
<td>Polat et al.</td>
<td>2007</td>
<td>[22]</td>
<td>✓</td>
</tr>
<tr>
<td>Luukka et al.</td>
<td>2007</td>
<td>[23]</td>
<td>✓</td>
</tr>
<tr>
<td>Polat et al.</td>
<td>2008</td>
<td>[24]</td>
<td>✓</td>
</tr>
<tr>
<td>Kahramanli et al.</td>
<td>2008</td>
<td>[25]</td>
<td>✓</td>
</tr>
<tr>
<td>Ghazavi et al.</td>
<td>2008</td>
<td>[26]</td>
<td>✓</td>
</tr>
<tr>
<td>Temurtas et al.</td>
<td>2009</td>
<td>[27]</td>
<td>✓</td>
</tr>
<tr>
<td>Gadaras et al.</td>
<td>2009</td>
<td>[17]</td>
<td>✓</td>
</tr>
<tr>
<td>Obeyli et al.</td>
<td>2010</td>
<td>[28]</td>
<td>✓</td>
</tr>
<tr>
<td>Ganji et al.</td>
<td>2010</td>
<td>[29]</td>
<td>✓</td>
</tr>
<tr>
<td>Luukka et al.</td>
<td>2011</td>
<td>[30]</td>
<td>✓</td>
</tr>
<tr>
<td>Isa et al.</td>
<td>2011</td>
<td>[31]</td>
<td>✓</td>
</tr>
<tr>
<td>Ozabaki et al.</td>
<td>2011</td>
<td>[32]</td>
<td>✓</td>
</tr>
<tr>
<td>Rodriguez et al.</td>
<td>2011</td>
<td>[33]</td>
<td>✓</td>
</tr>
<tr>
<td>Orkcu et al.</td>
<td>2011</td>
<td>[34]</td>
<td>✓</td>
</tr>
<tr>
<td>Beloufa et al.</td>
<td>2013</td>
<td>[35]</td>
<td>✓</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2013</td>
<td>[36]</td>
<td>✓</td>
</tr>
<tr>
<td>Yilmaz et al.</td>
<td>2014</td>
<td>[37]</td>
<td>✓</td>
</tr>
<tr>
<td>Gurbuz et al.</td>
<td>2014</td>
<td>[38]</td>
<td>✓</td>
</tr>
<tr>
<td>Van et al.</td>
<td>2014</td>
<td>[39]</td>
<td>✓</td>
</tr>
<tr>
<td>Feng et al.</td>
<td>2015</td>
<td>[40]</td>
<td>✓</td>
</tr>
<tr>
<td>Christopherson et al.</td>
<td>2015</td>
<td>[41]</td>
<td>✓</td>
</tr>
<tr>
<td>Erkaymaz et al.</td>
<td>2017</td>
<td>[42]</td>
<td>✓</td>
</tr>
<tr>
<td>Ramezani et al.</td>
<td>2017</td>
<td>[43]</td>
<td>✓</td>
</tr>
</tbody>
</table>
Rule-Based System (RLEFRBS) for diabetes diagnosis is proposed to handle the mentioned challenges with the goal of designing a diagnosis system with high accuracy and interpretability. Our proposed model comprises a two-step process: (1) reducing the number of rules and conditions and (2) using Genetic Algorithm (GA) and Reinforcement Learning (RL) to increase the consistency among the rules. The procedure of the proposed model is summarized as follows:

- The initial rules are generated using numerical data. The resulting rules are required to be pruned using a confidence measure.
- In order to avoid overfitting the training data and to improve interpretability, the conditions in the rules are pruned for creating simpler and more interpretable rules.
- The GA is used to select an appropriate subset of the candidate rules in order to create the initial Rule Base (RB).
- To improve the system’s performance using an evolutionary approach, the membership functions of the rules should be tuned. In this regard, the GAs are known proficient to find the optimal solutions in complex search spaces [44].
- The weights of the rules are adjusted with the goal of improving the system’s performance. The proposed model takes advantage of RL to adjust the rule weights in a dynamic online manner.
- Some instances may not be covered by any rules. In this study, the rule stretching is employed based on the idea of generalizing existing rules in order to cover all instances.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model and provides further details about the contributions. The experiments and results are presented in Section 3 followed by a discussion in Section 4. Finally, the concluding remarks and future streams are presented in Section 5.

2. RLEFRBS: Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System for diabetes diagnosis

An advantage of FRBSs is that the patterns in the patient data can be extracted in the form of interpretable rules using computer processing power. Fig. 1 depicts a generic FRBS architecture.

The Knowledge Base (KB) lies at the core of the FRBSs and consists of fuzzy if-then rules.

2.1. RB building

The building of the RB involves four steps: (1) rule learning, (2) rule pruning, (3) pruning rule antecedents, and (4) evolutionary rule selection. In the following, each step is described in detail.

2.1.1. Learning rules

In RLEFRBS, the first step in building the RB involves generating rules using numerical data.

Suppose there are \( m \) labeled patterns \( x_p = (x_{p1}, \ldots, x_{pn}) \) from \( M \) classes to classify \( n \)-dimensional patterns where \( x_p \) represents the value of the \( i \)-th attribute in pattern \( x_p \) and \( p = 1, 2, \ldots, m \). In this paper, \( n \)-dimensional patterns are classified using fuzzy rules in the following manner:

\[
\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and } \ldots \text{ and } x_i \text{ is } A_{qi} \text{ then Class } C_q \text{ with } CF_q
\]  

Where \( R_q \) is the \( q \)-th fuzzy rule; \( x_p = (x_{p1}, \ldots, x_{pn}) \) is the \( n \)-dimensional pattern vector; \( A_{qi} \) is an antecedent fuzzy set for the \( i \)-th attribute, \( C_q \) is the consequent class; and \( CF_q \) is the rule weight.

For each rule, a unique consequent class is determined based on patterns that is compatible with the antecedent part of the rule [46-48]. The compatibility grade of the pattern \( x_p = (x_{p1}, \ldots, x_{pn}) \) with the antecedent part of rule \( R_q \) is calculated in two steps. Firstly, the grade is calculated for each attribute value \( x_{pi} \) and its corresponding linguistic term \( A_{qi} \) via the membership function of the linguistic term denoted as \( \mu_{A_{qi}}(x_{pi}) \). Secondly, the product operator is applied to yield the compatibility grade of the pattern \( x_p \) with the antecedent part of \( R_q \).

\[
\mu_{A_{qi}}(x_{pi}) = \mu_{A_{qi}}(x_{pi}) \times \cdots \times \mu_{A_{qi}}(x_{pn})
\]

In Relation 1, the rule \( R_q \) can be viewed as an association rule in data mining (i.e., \( A_i \Rightarrow C_q \)) whose confidence is calculated in the following manner:

\[
\text{Confidence}(A_i \Rightarrow C_q) = \frac{\sum_{p:\text{Class } C_q} \mu_{A_i}(x_p)}{\sum_{p} \mu_{A_i}(x_p)}
\]

The confidence of a rule \( (A_i \Rightarrow C_q) \) refers to “the ratio of compatible patterns with both the antecedent part \( A_i \) and the consequent class \( C_q \) to compatible patterns with the antecedent part \( A_i \)”. Among the classes in the problem, the one having maximum confidence for the antecedent part of \( A_i \) is chosen as the consequent class. In other words, \( C_q \) is the class which satisfies the following constraint:

\[
c(A_i \Rightarrow C_q) = \max \{c(A_i \Rightarrow C_q) | q = 1, 2, \ldots, M\}
\]

There are two situations in which the antecedent part \( A_i \) cannot be used to generate any linguistic rules: (1) when more than one class has maximum confidence (i.e., a unique \( C_q \) cannot be specified) and (2) when no class has maximum confidence.

In the proposed model, the initial weights of the rules are calculated based on the rule accuracy. The weights are then adjusted using RL.

2.1.2. Rule pruning

During the process of extracting rules using the available instances, the number of generated rules tends to increase exponentially in proportion to the number of dimensions in the problem space.

Assuming there are \( m \) labeled patterns \( x_p = (x_{p1}, \ldots, x_{pn}) \), where \( p = 1, 2, \ldots, m \) from \( M \) classes in an \( n \)-dimensional continuous pattern space wherein each attribute domain is discretized into \( K_i \) linguistic values, each antecedent fuzzy set can be one of the discretized fuzzy sets. Therefore, in total, there are \( K_1 \times \cdots \times K_n \) possible combinations of the antecedent fuzzy set, and hence, the same number of rules. A pruning step is required to reduce the number of generated rules by eliminating redundant rules. The rules may be assessed using a number of criteria; in this study, the generated rules are pruned based on confidence, which is an appropriate measure adopted from data mining. Therefore, candidate rules having the same consequent classes are sorted in the descending order according to (5). The initial RB is then generated by selecting a predefined number of rules from each class. Using (5), the candidate
rules of each class are ranked based on various confidence levels.

\[ e(R_i) = \sum_{s_r \in \text{Class}_{C_i}} \mu_s(x_r) - \sum_{s_r \notin \text{Class}_{C_i}} \mu_s(x_r) \]  

(5)

### 2.1.3. Pruning rule antecedents

The rules generated using the aforementioned approach typically overfit the training data. In order to overcome this problem and improve interpretability, one can identify and eliminate redundant conditions in the antecedent part to simplify the rules as much as possible and maximize the performance of the pruning data.

In this pruning method, for each rule, the conditions of the antecedent part are first arranged in the descending order of importance, and the criterion is defined as follows:

\[ IV = \frac{P_r}{P_r + P_n} \]  

(6)

where \( p_r \) and \( p_n \) denote the number of positive (belonging to the consequent class of the rule) and negative (belonging to another class) pruning data points covered by the rule, respectively. Essentially, pruning refers to the task of finding a position for cutting down the list of antecedents. The rule-value metric is often used as the criterion for identifying the position of the cut-point. The metric is defined as follows:

\[ V(r) = \frac{P_c - P_p}{P_r + P_n} \]  

(7)

The pruning process begins with the last condition in the antecedent part of each rule, and the conditions are eliminated one at a time followed by the calculation of \( V(r) \). The conditions preceded by the one maximizing \( V(r) \) are eliminated. If a tie occurs, shorter rules are preferred. Subsequent to pruning the conditions, the initial weights of the rules are calculated based on the training data using (7).

### 2.1.4. Evolutionary rule selection

The rules resulting from pruning the antecedents are used as candidate rules in the rule-selection process. In their studies, Ishibuchi et al. [49] used a GA for the rule selection to choose a subset \( (S) \) of the initial rules \( (S_{all}) \) with high classification accuracy. Therefore, the process is formulated as follows:

Maximize \( f_1(S) \) and Minimize \( f_2(S) \)

(8)

Where \( S \subseteq S_{all} \), \( f_1(S) \) represents the number of correctly classified training patterns by a RB \( S \); \( f_2(S) \) is the number of fuzzy if-then rules in \( S \); and \( S_{all} \) denotes the total number of candidate rules.

In the RLEFRBS, a smaller RB (i.e., one with fewer rules) is preferred; thus, a constraint is imposed on the number of rules in the RB. If the objective is to maximize the number of correctly classified training patterns using at most \( N \) Rules, then the problem is formulated as follows:

Maximize \( f_1(S) \)

Subject to \( f_2(S) \leq N_{rule} \)

(9)

The important steps are detailed in the following subsections.

#### 2.1.4.1. A. Chromosome encoding

The rule selection process in this study uses classic binary encoding, and each gene is assigned to one rule. In order to enhance the population diversity, the chromosomes are randomly initialized using the values zero and one. This means that a chromosome of length \( N \) can be represented as shown below:

\[ S = s_1s_2...s_{q-1}s_q \]  

(10)

Where, \( s_q = 0 \) indicates that the \( q \)th candidate rule \( (R_q) \) is not included in the RB \( S \) whereas \( s_q = 1 \) shows that \( R_q \) is included in \( S \). Thus, the initial population is constructed by randomly generating a predetermined number of chromosomes.

### 2.1.4.2. B. Fitness Function

The fitness of each individual in the population (i.e., each RB \( S \)) according to the two objectives is defined as follows:

\[ \text{fitness}(S) = w_1f_1(S) - w_2 \max \{0, f_2(S) - N_{rule} \} \]  

(11)

Where \( w_1 \) and \( w_2 \) are specified such that \( w_1 < < w_2 \) in order to attach a large penalty when the constraint in relation (9) is not satisfied.

For each individual, the first objective function \( f_1(S) \) is calculated by classifying all the given training patterns using \( S \). The single winner rule is used as the reasoning method. Therefore, classification is performed by identifying a single winner rule for each pattern. As a result, some rules may remain unchosen as winner rules for all the patterns. Such rules can be eliminated without the loss of classification accuracy of the RB \( S \). The elimination of unnecessary rules simultaneously improves the second objective function \( f_2(S) \). The elimination is performed by changing each string of the current population to zero.

#### 2.1.4.3. C. Mutation

In order to effectively decrease the number of fuzzy if-then rules in each chromosome, biased mutation is employed in this step. This means that the probabilities for the mutations from 1 to 0 and 0 to 1 are unequal. The mutation operation is as follows:

\[ s_r = 1 \rightarrow s_r = 0 \text{ with probability } p_m(1 \rightarrow 0) \]  

(12)

\[ s_r = 0 \rightarrow s_r = 1 \text{ with probability } p_m(0 \rightarrow 1) \]  

(13)

### 2.2. Improving the RB

Here, the objective is to improve the performance of the RLEFRBS using three steps: (1) rule tuning, (2) adjusting rule weights using RL, (3) and rule stretching. In the following, each step is described in detail.

#### 2.2.1. Evolutionary rule tuning

Tuning aims to induce superior cooperation among the rules [50,51]. In doing so, variations are introduced into the shapes of the membership functions to enhance their global interaction. This is often performed as a post-processing activity following the extraction of fuzzy rules. A widely-used approach to tuning is one that focuses on Data Base (DB) definition. It is often referred to as tuning of membership functions or DB tuning [51-55]. Tuning methods typically refine the parameters that specify the membership functions of the global linguistic labels. Since the membership function parameters are interdependent, in high-dimensional (large-scale) problems, the tuning problem becomes one of optimization on a very complex search space. This impacts the performance of the optimization methods. A good solution is to use lateral tuning membership functions [52]. The approach takes advantage of the linguistic 2-tuples representation [55] and simplifies the search space by considering a single parameter for each membership function. Thus, it is easier to derive optimal models, especially in complex or high-dimensional problems and convergence speed is improved. The symbolic translation of a label involves the lateral displacement of its membership function, which is a number in the interval \((-0.5, 0.5)\) and expresses the extent of displacement between two adjacent lateral labels. Suppose a linguistic fuzzy partition \( S = \{s_0, ..., s_{L-1}\} \) where \( L \) is the number of labels. The symbolic translation of a label \( s_i \) in \( S \) using the 2-tuple notation is as follows:

\[ (s_i, \alpha_i), s_i \in S, \alpha_i \in [-0.5, 0.5] \]  

(14)

The symbolic translation of a label represented by the 2-tuple \((s_i, -0.3)\) and the associated lateral displacement are depicted in Fig. 2. The tuning process in the RLEFRBS makes use of an evolutionary approach. The important steps are detailed in the following subsections.
2.2.1.1. A. Chromosome encoding. Let us suppose there are \( n \) variables and \( L \) labels for each variable. Each chromosome is then encoded in the following form wherein the genes in the chromosomes are real numbers.

\[
(a_1^1, \ldots, a_1^L, \ldots, a_n^1, \ldots, a_n^L)
\]  

(15)

In order to take advantage of the available information, the current FRBS is included in the population as an initial solution. The remaining individuals in the population are randomly generated in the interval \((-0.5 \text{ and } 0.5)\).

2.2.1.2. B. Fitness function. A given chromosome is evaluated based on the classification accuracy of the available FRBS. This is done by taking into account the new membership functions defined by the corresponding chromosome.

2.2.1.3. C. Crossover operator. Here the BLX-\( \alpha \) operator is applied where \( \alpha \) is set to 0.5. Using the operator, the exploration and exploitation degrees of the crossover can be readily tuned. To perform the crossover operation, the operator is applied twice. Suppose two real-coded chromosomes are going to be crossed: \( X = (x_1, \ldots, x_g) \) and \( Y = (y_1, \ldots, y_g) \) where \( x_i, y_i \in (a_i, b_i) = [-0.5, 0.5] \subset \mathbb{R} \). Using the BLX-\( \alpha \) operator, the offspring are \( Z = (z_1, \ldots, z_g) \) where \( z_i \) is (uniformly) generated at random in \([l_i, u_i]\).

\[
i_i = \max\{a_i, c_{\min} - A\} \\
u_i = \min\{b_i, c_{\max} + A\} \\
c_{\min} = \min\{x_i, y_i\} \\
c_{\max} = \max\{x_i, y_i\} \\
A = (c_{\max} - c_{\min})/\alpha (16)
\]

2.2.2. Adjusting weights using RL

The rules in the RB include a weight parameter, which is a number within the interval \((0,1)\) appended to each rule and is interpreted as its strength. Weight adjusting is an approach that maintains the interpretability of the FRBS while enhancing the classification accuracy. In the RLEFRBS, we leverage RL to adjust the weights associated with the rules in the RB. RL is an online unsupervised learning method that acts by only interacting with the environment without using a model of the system. The learning agent is rewarded (or punished) for its actions such that it becomes more likely to take appropriate action in the future. In the proposed approach, in addition to weight adjusting, the parameters of the antecedent membership functions are tuned to some extent. The overall structure of the RL algorithm used to adjust the rule weights is as follows:

```
RL (RuleSet, Data, Target, FiringThreshold, ModifyThreshold)
1 // RuleSet = \{r_1, r_2, \ldots, r_n\} // n is the number of rules
2 // r_i = (A_1 \wedge A_2 \wedge \ldots \wedge A_n) \rightarrow \text{Class } C \text{ with } CF
3 // Data : Training Data
4 // Target : Classes of each Training Data
5 For i=1 to |Data| // each sample in Training Data
6   R = Winner Rule (Rule Set, Data [i])
7   C = Classify (Data [i], R) // Class of ith Training Data
8   // if (C = Target [i])
9   \[ CF^\text{new}_R = CF^\text{old}_R + \eta_i (1 - CF^\text{old}_R) \] // Reward (R)
10 // Identify members of group B
11 for j=1 to |Rule| // Rule j
12   if (Bring weight (Rule j) > FiringThreshold)
13     Add (Rule [j]) to B
14 End
15 End
16 if (B ≠ null)
17 for k=1 to |B|
18   Update (Counter of B [k])
19   if Counter of (B[k]) < FiringThreshold
20     \[ CF_{B[k]} = CF_{B[k]} - \eta_i CF^\text{old}_{B[k]} \] // Punish (B[k])
21 Else
22 Modify antecedent Membership Function of (B[k])(eq 21)
23 End
24 End
25 Else
26 Update (Counter of (R))
27 if (Counter of (R) < ThresholdModify)
28 \[ CF^\text{new}_R = CF^\text{old}_R - \eta_i CF^\text{old}_R \] // Punish (R)
29 Else
30 Modify antecedent Membership Function of (R) (eq 21)
31 End
32 End
33 Else
34 // Identify A
35 for i=1 to |Rule| // Rule i
36 if (Classify (Data [i], Rule [i]) = Target [i])
37   Add (Rule [i]) to A
38 End
39 End
40 if (A ≠ null)
41 for m=1 to |A|
42 \[ CF_{A[m]} = CF_{A[m]} + \eta_i (1 - CF_{A[m]}^\text{old}) \] // Reward (A[m])
43 End
44 End
45 End
46 End
47 End
```
Fig. 2. Symbolic translation of a label and lateral displacement of the associated membership function.

Fig. 3. Flow diagram of RL algorithm.
Fig. 3 depicts the algorithm’s flowchart.

2.2.2.1. Identifying members of groups A and B. In this study, we use the single winner rule reasoning method. Thus, for each input instance, a single winner rule is identified. If the classification result is incorrect, not only the winner rule but also a group of rules (denoted as group A) that is capable of correctly classifying the instance must be punished. In fact, the wrong rule is identified as the winner as the rules in group A fail to determine the true class. As a result of such fluctuations, it is necessary to modify the antecedent part.

2.2.2.2. Encouraging and punishing the rules. Equations (17) and (18) are used to encourage (increase weights) and to punish (decrease weights) the rules, respectively.

\[
CF_{\text{new}} = CF_{\text{old}} + \eta_1 \cdot (1 - CF_{\text{old}}) \quad (17)
\]

\[
CF_{\text{new}} = CF_{\text{old}} - \eta_2 CF_{\text{old}} \quad (18)
\]

where \(\eta_1\) and \(\eta_2\) are the learning rate.

2.2.2.3. C. Counters associated with the rules. In general, merely adjusting rule weights does not lead to a sufficient improvement in the classification accuracy. In cases wherein the rules are unable to cover the training data to an acceptable degree, it is recommended that the antecedent part of the rules be modified to achieve high accuracy. Therefore, two metrics are defined for each rule: (1) the number of punishments and (2) the number of classification result fluctuations.

2.2.2.3.1. C.1. Counting rule punishments. This counter is incremented if the associated rule becomes a member of group B or makes an incorrect classification decision as a winner rule. If the value of the counter exceeds a predefined threshold, the learning process switches to the antecedent modification phase.

This counter is updated as follows:

\[
p_{CR_{ki}} = \begin{cases} 
 p_{CR_{ki}} + 1 & \text{if } R_k \text{ should be punish} \\
p_{CR_{ki}} & \text{otherwise} 
\end{cases} \quad (19)
\]

where \(p_{CR}\) represents the number of times \(R_k\) is punished, \(k = 1, 2, \ldots, N\).

2.2.2.3.2. C.2. Counting rule fluctuations. Occasionally, classification decisions tend to fluctuate between correct and incorrect results; in other words, at times, a rule results in the correct classification while incorrectly classifying the input instance at other times. As a result of such frequent fluctuations, it is necessary to modify the antecedent part. For each rule, a separate counter is used to determine the frequency of its fluctuations. If the counter exceeds a predefined threshold, the learning process switches to the antecedent modification phase. The counter is updated as follows:

\[
SC_{CR_{ki}} = \begin{cases} 
 SC_{CR_{ki}} + 1 & \text{if } R_k \text{ swings} \\
SC_{CR_{ki}} & \text{otherwise} 
\end{cases} \quad (20)
\]

where \(SC_{CR}\) is the counter maintaining the number of fluctuations for rule \(R_k\), \(k = 1, \ldots, N\).

2.2.2.4. D. Modifying the antecedent part. In the antecedent modification phase, all the membership functions of the fuzzy rules are shifted into an allowed interval in order to increase or decrease their firing strength. If the training input instance is included in the rule’s class, the shift results in higher firing strength for the rule; in contrast, if the training instance is not included in the rule’s class, the shift decreases the firing strength of the rule. The extent of the shift depends on the distance between the value of the sample and the membership function center in each dimension.

\[
\begin{align*}
\text{mfc}_i^{\text{new}} &= \begin{cases} 
 \text{mfc}_i^{\text{old}} - \alpha \times (\text{mfc}_i^{\text{old}} - x_{pi}) & \text{if } \text{mfc}_i^{\text{old}} \geq x_{pi}, \mu_{R_i} \text{ is increased} \\
\text{mfc}_i^{\text{old}} + \alpha \times (\text{mfc}_i^{\text{old}} - x_{pi}) & \text{if } \text{mfc}_i^{\text{old}} \leq x_{pi}, \mu_{R_i} \text{ is decreased} 
\end{cases} \\
\end{align*} \quad (21)
\]

In relation (21), \(mfc_{R_i}^{\text{old}}\) represents the membership function center in dimension \(i\) of the rule \(R_i\), \(k = 1, \ldots, N\), the changes in value of which shift the respective membership function along its own feature axis, and \(0 \leq \alpha \leq 1\) is the shift rate.

2.2.3. Rule stretching

In FRBSs, some instances may not be covered by any rules. This issue becomes challenging as the number of such instances increases. To overcome the non-covering problem, Eibeorg et al. [56] used rule stretching or generalization which is achieved by removing one or more conditions of the antecedent part. The generalization of a rule is minimal if it does not delete more conditions of the antecedent part than necessary to cover the query instance. Having all minimal generalizations, the authors re-evaluate each rule using its Laplace accuracy on the training data. Then, the rule with the highest evaluation is used to classify the instance. Given that every rule is re-evaluated, EB-Stretching is computationally complex. The worst case occurs when all training instances suffer from the non-covering problem and has a complexity of \(O(|R_k| \cdot |D_m|)\) where \(|R_k|\) and \(|D_m|\) represent the number of rules and the size of the training set, respectively.

Fig. 4 demonstrates an instance of rule stretching. Two classes (a and b) as well as two rules (R1 and R2) are shown. In Fig. 4a, two rules along with their covered instances are shown. Moreover, Fig. 4b presents the instances (?) that are not covered by these rules. In Fig. 4a, the rules are generalized such that the instances are covered by R2.

In this study, in order to address the problem of instances not being covered and rule stretching, the conditions of the antecedent part are considered in the order of pruning. Thereafter, for each instance that is not covered, the procedure begins by successively removing the last condition in the antecedent part of the rule while the compatibility between the instance and the rule is checked. This continues until the instance is covered by the rule or the antecedent part of the rule contains no more conditions. Once all the rules compatible with the instance are obtained, they are evaluated using the following criterion, and the instance is classified by the rule having the maximum evaluation.

\[
RS = \frac{p + 1}{p + n + 1} \times \frac{k + 1}{m + 2} \quad (22)
\]

where \(p\) and \(n\) represent the number of positive and negative instances.

- a. before rule stretching
- b. after rule stretching

Fig. 4. Rule stretching.
covered by the rule, respectively, \( k \) is the number of remaining conditions in the antecedent part, and \( m \) denotes the initial number of conditions in the antecedent.

In this study, the system accuracy is calculated as shown below:

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}
\]

where \( TP, FP, TN \) and \( FN \) denote the diagnosis of patients correctly classified as diabetic (True-Positive), diagnosis of nondiabetic patients classified as diabetic (False-Positive), diagnosis of patients correctly classified as nondiabetic (True-Negative) and diagnosis of diabetic patients classified as nondiabetic (False-Negative), respectively.

The overall design of the RLEFRBS is shown in Fig. 5 while Fig. 6 illustrates the architecture of the system based on the foregoing details.

3. Experimental results

In this section, the proposed model (i.e., RLEFRBS) is evaluated using two datasets. In the following, the classification accuracy of the RLEFRBS is compared to that of several other methods in extant literature. Furthermore, the discussion section focuses on the general interpretation of the rules in the RB of the datasets. The steps of the proposed model were coded in MATLAB 2016. The computer used was a PC with AMD A10-7400P.

3.1. Datasets

The proposed model (i.e., RLEFRBS) is evaluated using two datasets namely Pima Indian Diabetes (PID) [57] and BioSat Diabetes Dataset (BDD) [58]. PID is used in the majority of data mining techniques for analyzing results. Table 2 summarizes the characteristics of the datasets including the number of instances, attributes, and classes. Among the 19 attributes in BDD, a total of 8 are selected for use in the problem space.

3.2. Implementation

As explained in Section 2, the implementation of the proposed model involves two steps: (1) constructing the RB and (2) optimizing the rules.

3.2.1. RB building

In order to construct the RB, we follow the steps in Subsection 2.1: learning rules, rule pruning, pruning rule antecedents, and evolutionary rule selection. In the following, the details of each step and the resulting RB for the two datasets are briefly discussed.

3.2.1.1. Learning rules. Two-thirds of the instances in each dataset are randomly selected as training data while the remaining instances are used for testing purposes. The training data are further divided into training and pruning sets at random at a ratio of 2:1. In this study, for each variable, five triangular fuzzy numbers are considered: very low, low, medium, high, and very high. For example, the numbers assigned with the blood glucose variable in the PID are shown in Fig. 7. Based on the possible combinations of the fuzzy numbers, there are 7593 and 10530 possible rules in the rule-learning phase for the datasets PID and BDD, respectively.

3.2.1.2. Rule pruning. The previous step (learning rules) results in a large number of rules for both datasets, which is expected given the number of problem dimensions (i.e., eight variable). However, some rules are redundant and are required to be eliminated before constructing the RB. Previous works have used a variety of methods to prescreen the generated rules. In RLEFRBS, this is done using the confidence criterion in data mining. The initial RB is constructed by selecting 100 rules from each class—a total of 200 rules. The properties of the initial RB are shown in Table 3.

3.2.1.3. Pruning rule antecedents. As shown in Table 3, each rule antecedent in the databases contains eight conditions on average. In general, rules with fewer conditions in the antecedent are easier to interpret and understand. In this step, the redundant conditions in the antecedents are identified and eliminated. The results thus obtained are shown in Table 4. On comparing the average number of conditions in the antecedent parts of the rules before and after pruning (Tables 3 and 4), improved interpretability is observed. In addition to improved interpretability, pruning leads to lower specificity, which in turn allows a greater number of instances to be covered. A comparison of the number of unlabeled training and test instances in Tables 3 and 4 clearly demonstrates this. Furthermore, under certain conditions, the pruning process may create duplicate rules, which are eliminated from the RB. As shown in Table 4, subsequent to pruning, the number of rules in the PID and BDD can be reduced to 140 and 136, respectively. Therefore, overall, the results in Table 4 suggest that the resulting RB surpasses the initial RB in terms of the number of rules and average number of conditions in the antecedents.

3.2.1.4. Evolutionary rule selection. The design objective of the FRBS is to construct a compact RB with a high classification accuracy. Based on extant literature, one classical approach is to directly search for a set of rules with sufficient classification accuracy. GAs are a subclass of evolutionary algorithms that are widely used for determining optimal solutions in complex search spaces. In the RLEFRBS, a GA is used in the evolutionary rule selection in order to choose a subset of rules with sufficient accuracy. The parameters are tuned using trial and error and are shown in Table 5. In this algorithm, roulette-wheel selection is applied for selecting the rules while the population members are combined using a uniform function. Moreover, during the elite selection stage, the five top chromosomes in the current population are transferred to the next generation. Table 6 summarizes the specifications of the final RB, which contains 19 and 6 rules for the PID and BDD datasets, respectively. Based on the number of rules in the RB and the average number of conditions in rule antecedents (Table 6), the resulting RB provides adequate interpretability for both the datasets.

3.2.2. Improving the RB

As explained in Subsection 2.2, the aim of this step is to improve the

![Fig. 5. Overall RLEFRBS.](image-url)
performance of the RB. The results are summarized as follows.

3.2.2.1. **Evolutionary rule tuning.** The objective of tuning is to enhance the FRBS performance (Subsection 2.2.1). Given the desirable properties of GAs, the evolutionary approach is applied for tuning the membership functions of the rules. The parameters of the algorithm are tuned using the trial and error method. The GA parameter values are listed in Table 7. In the GA, the roulette-wheel selection is applied to the members using a random mutation operator. The evaluation results are shown in Table 8. On comparing the results in Tables 6 and 8, an improved system accuracy

<table>
<thead>
<tr>
<th>Dataset</th>
<th>the number of instances</th>
<th>the number of attributes</th>
<th>the number of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pima Indian Diabetes dataset (PID)</td>
<td>768</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Biostat Diabetes Dataset (BDD)</td>
<td>403</td>
<td>19</td>
<td>2</td>
</tr>
</tbody>
</table>
and better coverage of the instances by the rules in the RB can be observed. This confirms the enhancement of the overall system performance.

3.2.2.2. Adjusting weights. The rules in the resulting RB are weighted. As mentioned in Subsection 2.2.2, using weight adjustment, the classification accuracy can be improved while maintaining the FRBS interpretability. In this study, in order to obtain further improvement, we leverage RL, which has a number of beneficial properties including interactivity. Thus, in order to adjust the rule weights, the algorithm parameters are tuned using the trial and error method (Table 9).

The RLEFRBS evaluation results for both datasets subsequent to the weight adjustments are presented in Table 10. On comparing the results in Tables 8 and 10, a substantial improvement in the system accuracy can be observed for both training and test instances after adjusting the rule weights.

3.2.2.3. Rule stretching. As discussed in Subsection 3.2.2, some instances are occasionally not covered by the rules in the RB. As is evident from Table 10, certain instances in the PID remain uncovered by the rules in the RB; thus, no labels are assigned to these instances. In order to

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of rules in the RB</th>
<th>Total number of conditions in rule antecedents</th>
<th>Average number of conditions in rule antecedents</th>
<th>System accuracy on training data</th>
<th>Number of unlabeled training instances</th>
<th>System accuracy on test instances</th>
<th>Number of unlabeled test instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>200</td>
<td>1600</td>
<td>8</td>
<td>0.62</td>
<td>176</td>
<td>0.76</td>
<td>23</td>
</tr>
<tr>
<td>BDD</td>
<td>200</td>
<td>1600</td>
<td>8</td>
<td>0.88</td>
<td>112</td>
<td>0.91</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of rules in the RB after pruning</th>
<th>Total number of conditions in rule antecedents after pruning</th>
<th>Average number of conditions in rule antecedents after pruning</th>
<th>System accuracy on training data</th>
<th>Number of unlabeled training instances</th>
<th>System accuracy on test instances</th>
<th>Number of unlabeled test instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>140</td>
<td>800</td>
<td>5/7</td>
<td>0.72</td>
<td>44</td>
<td>0.82</td>
<td>4</td>
</tr>
<tr>
<td>BDD</td>
<td>136</td>
<td>848</td>
<td>6/2</td>
<td>0.94</td>
<td>0</td>
<td>0.92</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of generations (T)</th>
<th>Population size (n₀)</th>
<th>Crossover Rate (pₖ)</th>
<th>Mutation Rate (pₑ)</th>
<th>W₁</th>
<th>W₂</th>
<th>Nₑₑₑ</th>
<th>elite selection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>500</td>
<td>100</td>
<td>0.9</td>
<td>0.001</td>
<td>0.1</td>
<td>0.25</td>
<td>0.75</td>
<td>20</td>
</tr>
<tr>
<td>BDD</td>
<td>500</td>
<td>100</td>
<td>0.9</td>
<td>0.001</td>
<td>0.1</td>
<td>0.25</td>
<td>0.75</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of rules in the final RB</th>
<th>Total number of conditions in rule antecedents</th>
<th>Average number of conditions in rule antecedents</th>
<th>System accuracy on training data</th>
<th>Number of unlabeled training instances</th>
<th>System accuracy on test instances</th>
<th>Number of unlabeled test instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>19</td>
<td>91</td>
<td>4.7</td>
<td>0.73</td>
<td>57</td>
<td>0.83</td>
<td>8</td>
</tr>
<tr>
<td>BDD</td>
<td>6</td>
<td>17</td>
<td>2.8</td>
<td>0.93</td>
<td>18</td>
<td>0.95</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of generations (T)</th>
<th>Population size (n₀)</th>
<th>Crossover Rate (pₖ)</th>
<th>Mutation Rate (pₑ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>500</td>
<td>100</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>BDD</td>
<td>500</td>
<td>100</td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Fig. 7. Instance partitioning of the blood glucose attribute in PID.
overcome this problem in the RLEFRBS, we use the rule stretching procedure detailed in Subsection 3.2.2. Table 11 summarizes the results.

The system accuracy for the newly covered instances as well as overall accuracy subsequent to the rule stretching are calculated for the PID dataset, and the results are shown in Table 12. As can be observed, the unlabeled instances are appropriately covered by the rules, and system accuracy for the instances is adequate.

3.3. Evaluating the proposed model

Fig. 8 through 10 illustrate the final RB associated with the PID and BDD datasets. The rules in the RB are assessed based on the instances covered by the rules. The results are summarized in Table 13 through 16. For each rule, the number of instances covered by the rule as well as the number of instances that are correctly (or incorrectly) classified by the rules, is shown. Based on the tables, the rule performance can be evaluated in terms of accuracy using real data. For instance, with respect to the third rule of the PID dataset, 44 and 2 classified instances are correct and incorrect respectively. It is clear that the rule performance is superior. As the rule weights are adjusted using RL, which is based on the interactions between the environment and the agent, a rule with superior performance is expected to have a larger weight, i.e., the rule performance should be reflected in the adjusted weights. This is confirmed on examining the performance of the rules in Table 13 through 16 and the rule weights in Fig. 8 through 10.

For instance, the third rule for the PID dataset was shown to provide a superior performance, which is consistent with the weight of the rule of 0.98, as shown in Fig. 8. Indeed, the superior performance is reflected in the adjusted weight of the rule.

In the figures depicting the resulting RB (Figs. 8–10), several recurrent attributes can be identified. These attributes are more important in diagnosing diabetes. The attribute frequencies in the final rules RB are listed in Tables 17 and 18. Based on these tables, in the PID dataset, age, Plasma Glucose Tolerance (PGT), and Number of Pregnancies (NOP), are found to have the highest significance in the diagnosis of diabetes whereas Stabilized Glucose (stab.glu) is the most relevant in the BDD dataset. These are confirmed by looking at the performance of the rules containing these attributes (Tables 13–16).

### Tables

<table>
<thead>
<tr>
<th>Table 8</th>
<th>RB tuning results using a GA.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>Number of rules in the RB</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>PID</td>
<td>19</td>
</tr>
<tr>
<td>BDD</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9</th>
<th>Weight adjustment algorithm parameter values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>S.T</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>PID</td>
<td>200</td>
</tr>
<tr>
<td>BDD</td>
<td>200</td>
</tr>
</tbody>
</table>

S.T: Switch Threshold of learning process. β: Firing strength to create groups A and B.

<table>
<thead>
<tr>
<th>Table 10</th>
<th>RLEFRBS performance evaluation results after adjusting rule weights.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>Number of rules in the RB</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>PID</td>
<td>19</td>
</tr>
<tr>
<td>BDD</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 11</th>
<th>Rule stretching results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>Number of unlabeled training instances before rule stretching</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>PID</td>
<td>51</td>
</tr>
<tr>
<td>BDD</td>
<td>0</td>
</tr>
</tbody>
</table>

3.4. Comparison

Tables 19 and 20 compare the accuracy of the proposed model with those of previous works. The tables indicate that our method offers high accuracy for the datasets.

Next, the performance of the RLEFRBS was examined using real patient data collected from a total of 50 patients with the help of physicians specializing in diabetes. Table 21 presents the accuracy of the RLEFRBS for the data. As can be observed, sufficient accuracy is realized.

4. Discussion

In this section, in order to demonstrate the interpretability of the proposed model, we elaborate on several rules in the final RBs. This is followed by a discussion on the interpretability of the entire RB.

4.1. A. PID rules

1. If BMI is Very High and blood glucose is High then Yes Diabetes.
2. If DBF is Medium and PGT is Very High and DBP is Very High Then Yes diabetes.
3. If PGT is High and NOP is Very Low and Age is Very Low and BMI is Medium Then Yes Diabetes.

4.2. B. BDD rules

1. If Stab. Glu is Very High and blood glucose is High then Yes Diabetes.
2. If Stab. Glu is Very High Then No Diabetes.
3. If Stab. Glu is Medium and Waist is Very High and Weight is Medium Then Yes Diabetes.

4.3. Overall interpretability of the RB

Based on a closer examination of the dataset resulting from the rules, it is possible to arrive at more general conclusions.
### Table 12
RLEFRBS accuracy after covering instances through rule stretching.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of correctly covered training instances</th>
<th>Number of wrong covered training instances</th>
<th>Number of correctly covered test instances</th>
<th>Number of wrong covered test instances</th>
<th>System accuracy on training data</th>
<th>System accuracy on test instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>33</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td>0.7388</td>
<td>0.8397</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule</th>
<th>NoP</th>
<th>Pgt</th>
<th>DBP</th>
<th>TSFT</th>
<th>THBS</th>
<th>BMI</th>
<th>DPI</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R2</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R3</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R4</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R5</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R6</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R7</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R8</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R9</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
<tr>
<td>R10</td>
<td>99.7 124.9 228.1</td>
<td>33.2 41.0 75.1</td>
<td>-0.5 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
<td>0.1 0.4 0.7</td>
</tr>
</tbody>
</table>

**Fig. 8.** The RB corresponding to the PID dataset.

**Fig. 9.** The RB corresponding to the PID dataset.
4.3.1. Overall interpretation of the RB associated with PID

1. Based on R9, R4, R6, R7, R10, R11, R13, and R14, diabetes is less likely to occur in individuals under the age of 30 years. In addition, R15, R16, R17, R18, and R19 indicate that people over the age of 30 years have a greater chance of developing the disease. Thus, being older than 30 years is a risk factor for diabetes.

2. With respect to PGT, R1, R2, R9, R15, R16, R17, R18, and R19 confirm that patients with extremely high blood glucose levels (i.e., exceeding 120) suffer from diabetes regardless of their age.

3. According to R3, R6, R7, R10, R11, R13, and R14, the risk of diabetes is very low if the patients have experienced very few pregnancies (i.e., fewer than two). However, R9, R15, R16, R17, R18, and R19 indicate that diabetes becomes more likely if the number of pregnancies is few, medium, high, or very high (i.e. exceeding two times). Therefore, having been pregnant more than twice is a possible risk factor for diabetes.

4. R9, R15, R16, R17, R18, and R19 indicate that pregnant women over the age of 30 years are more prone to developing diabetes. R16, R18, and R19 indicate a higher risk among overweight pregnant women (whose body mass index (BMI) exceeds 25). Moreover, according to R9, R15, R16, R17, R8, and R19, a higher risk of diabetes is observed among pregnant women whose blood glucose is very high (i.e. greater than 120 mg/dL).

5. Blood pressure is among the most important factors that cause or intensify diabetes. R2, R16, R17, R18, R19, R7, and R14 suggest that high or very high diastolic blood pressure (i.e. greater than 70 mmHg) is a concerning sign for diabetes. If the patient also has high blood glucose, he/she suffers from diabetes (R16, R17, R18, and R19); however, if the patient is under 30 years of age and has had fewer than two pregnancies (the threshold for non-diabetic patients), he/she is not affected by diabetes (R2, R7, and R14).

6. R14, R16, and R18 identify another risk factor, which is BMI. According to R9, R15, R16, R17, R8, and R19, a higher risk of diabetes is observed among pregnant women whose blood glucose is very high (i.e. greater than 120 mg/dL). Fig. 10. The RB corresponding to the BDD dataset.

<table>
<thead>
<tr>
<th>Rules</th>
<th>Variable Name</th>
<th>Class</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td></td>
<td>0</td>
<td>0.95</td>
</tr>
<tr>
<td>R2</td>
<td></td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td>R3</td>
<td></td>
<td>1</td>
<td>0.82</td>
</tr>
<tr>
<td>R4</td>
<td></td>
<td>1</td>
<td>0.93</td>
</tr>
<tr>
<td>R5</td>
<td></td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>R6</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 13** Performance of the RB corresponding to the PID dataset on training instances.

**Table 14** Performance of the RB corresponding to the PID dataset on test instances.

4.3.1. Overall interpretation of the RB associated with PID

1. Based on R9, R4, R6, R7, R8, R10, R11, R13, and R14, diabetes is less likely to occur in individuals under the age of 30 years. In addition, R15, R16, R17, R18, and R19 indicate that people over the age of 30 years have a greater chance of developing the disease. Thus, being older than 30 years is a risk factor for diabetes.

2. With respect to PGT, R1, R2, R9, R15, R16, R17, R18, and R19 confirm that patients with extremely high blood glucose levels (i.e., exceeding 120) suffer from diabetes regardless of their age.

3. According to R3, R6, R7, R10, R11, R13, and R14, the risk of diabetes is very low if the patients have experienced very few pregnancies (i.e., fewer than two). However, R9, R15, R16, R17, R18, and R19 indicate that diabetes becomes more likely if the number of pregnancies is few, medium, high, or very high (i.e. exceeding two times). Therefore, having been pregnant more than twice is a possible risk factor for diabetes.

4. R9, R15, R16, R17, R18, and R19 indicate that pregnant women over the age of 30 years are more prone to developing diabetes. R16, R18, and R19 indicate a higher risk among overweight pregnant women (whose body mass index (BMI) exceeds 25). Moreover, according to R9, R15, R16, R17, R8, and R19, a higher risk of diabetes is observed among pregnant women whose blood glucose is very high (i.e. greater than 120 mg/dL).
### Table 15
Performance of the RB corresponding to the BDD dataset on training instances.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Total number of classifications</th>
<th>Number of correct classifications</th>
<th>Number of incorrect classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>169</td>
<td>161</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 16
Performance of the RB corresponding to the BDD dataset on test instances.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Total number of classifications</th>
<th>Number of correct classifications</th>
<th>Number of incorrect classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 17
Attribute frequencies in the RB corresponding to the PID dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>Age</td>
<td>Pgt</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>DPP</td>
</tr>
<tr>
<td></td>
<td>THSI</td>
<td>TSFT</td>
</tr>
</tbody>
</table>

### Table 18
Attribute frequencies in the RB corresponding to the BDD dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDD</td>
<td>stab.glu</td>
<td>Waist</td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Chol</td>
<td>Hip</td>
</tr>
<tr>
<td></td>
<td>Ratio</td>
<td>Hdi</td>
</tr>
</tbody>
</table>

### Table 19
Comparing the accuracy of the RLEFRBS to other methods in the literature for Pima dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Ref</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incremental SVM-FP</td>
<td>[59]</td>
<td>73.80%</td>
</tr>
<tr>
<td>General Regression Network</td>
<td>[60]</td>
<td>80.21%</td>
</tr>
<tr>
<td>Inverted Hierarchical Neuro-Fuzzy System</td>
<td>[21]</td>
<td>78.0%</td>
</tr>
<tr>
<td>Fuzzy similarity Classifier</td>
<td>[61]</td>
<td>75.29%</td>
</tr>
<tr>
<td>Incremental Random Forests</td>
<td>[62]</td>
<td>76.80%</td>
</tr>
<tr>
<td>Incremental PCA</td>
<td>[63]</td>
<td>68.10%</td>
</tr>
<tr>
<td>CA.45 rules</td>
<td>[25]</td>
<td>67.0 ± 2.9</td>
</tr>
<tr>
<td>MLP + BP</td>
<td>[24]</td>
<td>75.8 ± 6.2</td>
</tr>
<tr>
<td>CART</td>
<td>[25]</td>
<td>72.8</td>
</tr>
<tr>
<td>Semi-Naïve Bayes</td>
<td>[24]</td>
<td>76.0 ± 0.8</td>
</tr>
<tr>
<td>kNN</td>
<td>[24]</td>
<td>71.9</td>
</tr>
<tr>
<td>ID3</td>
<td>[25]</td>
<td>71.7 ± 6.6</td>
</tr>
<tr>
<td>Kohonen</td>
<td>[25]</td>
<td>72.7</td>
</tr>
<tr>
<td>LS-SVM</td>
<td>[24]</td>
<td>78.21</td>
</tr>
<tr>
<td>MLNN with LM(10 × FC)</td>
<td>[27]</td>
<td>79.62</td>
</tr>
<tr>
<td>MLNN with LM</td>
<td>[27]</td>
<td>82.0</td>
</tr>
<tr>
<td>RLEFRBS</td>
<td></td>
<td>84%</td>
</tr>
</tbody>
</table>

### Table 20
Comparing the accuracy of the RLEFRBS to other methods in the literature for BDD dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Ref</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID3, CA.45 &amp; CART Ensembles</td>
<td>[64]</td>
<td>91.56%</td>
</tr>
<tr>
<td>Multi-layer classifiers</td>
<td>[65]</td>
<td>93.05%</td>
</tr>
<tr>
<td>RLEFRBS</td>
<td></td>
<td>99%</td>
</tr>
</tbody>
</table>

7. The rule pertaining to the pedigree function (R2) shows that a medium value increases the likelihood of developing diabetes; in this case, the patient definitely suffers from the disease provided that his/her blood glucose is more than 100 mg/dL. Based on R4, R7, R8, R11, R14, R17, R18, and R19, diabetes can be developed even if the value of the pedigree function is very low. In this case, on using R4, R7, R8, R11, and R14, if the function value is very low and the patient is younger than 30 years, he/she does not have diabetes; however, if the patient is over 30 years of age, has had two pregnancies, and has a very high blood glucose level (exceeding 120 mg/dL), using R17, R18, and R19, the patient is diabetic. Thus, it is obvious that age, number of pregnancies, and blood glucose level are important in diagnosing diabetes.

8. R4, R7, R11, R16, R17, R18, and R19 concern another diabetes risk factor-very low insulin levels in the patient's blood sample (serum). According to R4, R7, and R11, the patient does not have diabetes if the following conditions hold: (1) the insulin level in the patient's blood is very low, (2) the patient is younger than 30 years of age, and (3) the pedigree function value is also very low. Contrarily, very low levels of insulin in addition to over two pregnancies, being over the age of 30 years, and having very high blood glucose levels are indicative of a high diabetes risk as shown by R16, R17, R18, and R19. Thus, the significance of age, number of pregnancies and blood glucose levels is apparent.

9. Another factor that represents a diabetes risk is the triceps skinfold thickness. It can prove useful in combination with other factors. In cases wherein the triceps skinfold thickness is low, the patient is under the age of 30 years, and the glucose is high, based on R8 and R10, the patient is not diabetic; however, if the thickness is low, the age is over 30 years, and glucose is very high, based on R17, the patient is diabetic. Based on R15, R16, R18, and R19, the patient suffers from diabetes on the condition that thickness is very low or medium, his/her age exceeds 30, the patient has had more than two pregnancies, and the blood glucose is very high. Once again, the significance of the age, number of pregnancies and blood glucose levels is apparent.

As indicated above, very high blood glucose is extremely important in diagnosing diabetes. On using the aforementioned rules, in case the symptoms point to the presence of diabetes, the physician can simply measure the blood glucose to make a definite decision. For instance, a patient whose age and BMI are greater than 30 years and 25, respectively, suffers from diabetes provided that his/her blood glucose level falls in the diabetic range (i.e. more than 120 mg/dL), based on R17, R18, and R19, the patient is diabetic. Thus, it is obvious that age, number of pregnancies and blood glucose in diagnosing diabetes is apparent.

### 4.3.2. Overall interpretation of the RB associated with BDD

1. According to the resulting RB, blood glucose is the most important attribute in the dataset. The first rule in the RB obtained from BDD (i.e. R1) specifies a higher bound for this attribute in non-diabetic patients. Accordingly, if the blood glucose is very low (i.e. lower than 100 mg), the patient is not diabetic. In a similar vein, R2 and R3...
establish a lower bound for blood glucose. Thus, if blood glucose is high or very high (i.e. greater than 200 mg), the patient has diabetes. However, if the glucose level is low, the presence of diabetes depends on other factors. In this case, based on R4, the patient is not diabetic if his/her blood cholesterol is low; otherwise, according to R6, he/she has diabetes.

2. R5 and R6 establish a lower bound for waist size in diabetic patients. As indicated, a large waist size (i.e. more than 36 inches) increases the risk of developing diabetes.

3. Moreover, R5 and R6 is concerned with the risk of weight: if the patient’s weight is medium and waist size exceeds 36 inches, the patient has diabetes.

4. The next attribute relates to age. According to the rules, diabetes is more likely to develop over the age of 30 years. As such, if the blood cholesterol and glucose levels are low, based on R4, the patient is not diabetic; otherwise, high blood cholesterol indicates the presence of diabetes (R6).

5. R4 and R6 indicate that the risk of diabetes is higher if the patient’s blood cholesterol is high (i.e., over 240 mg).

As mentioned earlier, FRBSs have two contradicting objectives: accuracy and rule interpretability. This is especially important in medical applications.

Although a precise definition of the FRBS interpretability is lacking, the number of rules and average rule length (i.e., average number of conditions in the antecedent) are often used as two components of interpretability [66]. Based on the obtained RBs (Figs. 8–10), and the corresponding specifications in Table 10 (number of rules), there are 19 rules for PID and 6 for BDD. Considering the average number of rules in the antecedent part (4.7 and 2.6, respectively), the system interpretability can be confirmed.

The important components of the MDSS include the user interface, RB, and decision-making component, i.e., the inference engine. The system inputs include information regarding the patient status. In fact, these are the parameters that are identified as important in diagnosing diabetes: age, number of pregnancies, blood pressure, BMI, and pedigree function among others. The system receives the parameters through the user interface. The RB is the primary component of a DSS. This RB comprises a number of rules that are used to make a diagnosis. The rules in this study were proven to be highly accurate and interpretable. Thus, using the rules in this RB, one can very accurately diagnose diabetes. The user is informed of the results via the user interface. Users of this DSS include patients, nurses, and physicians. Its primary objective is to help physicians and ordinary individuals to make faster and more accurate diagnoses at a lower cost.

5. Conclusion

Nowadays, diabetes has become a serious health problem. Early diagnosis of the disease is critical in preventing its adverse effects. A number of methods have been proposed for early diagnosis of diabetes. In this paper, a novel FRBS is proposed to achieve the maximum accuracy and interpretability. The performance of the proposed system was examined using two common datasets for diabetes: PID and BDD. The performance assessment results of the proposed model indicated that the final RB has high accuracy and interpretability which makes it suitable for medical diagnosis. Furthermore, according to the obtained results, the RB well matches to the real patient data; therefore, it is capable to model diabetes behavior and represent the knowledge in patient data. The RB may be used as the KB in an expert system in order to help both physicians and ordinary individuals. Nevertheless, diagnosing diabetes is a complex process and the algorithms must be tested on a larger number of datasets in future works. In addition, the performance of other optimization methods such as Particle Swarm Optimization (PSO) could be investigated. Future works may take advantage of type-2 fuzzy sets for assessing and improving the performance of the proposed system.

References