A Case Study on Medical Diagnosis of Cardiovascular Diseases Using a Genetic Algorithm for Tuning Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets

J. Sanz*, M. Pagola*, H. Bustince*, A. Brugos†
* Dept. Automática y Computación
† Dept. Ciencias de la Salud
Universidad Pública de Navarra
Campus de Arrosadia, 31006 Pamplona, Spain
E-mail: joseantonio.sanz@unavarra.es

A. Fernández
Dept. Computer Science
University of Jaén
23071 Jaén, Spain

F. Herrera
Dept. Computer Science
and Artificial Intelligence
University of Granada
18071 Granada, Spain

***Abstract*—In this contribution, we use Fuzzy Rule-Based Classification Systems for classifying the patients with respect to the risk of suffering cardiovascular diseases. Specifically, we use a methodology in which the linguistic labels of the classifier are modeled by means of IVFSs. Thereafter, they are genetically post-processed for tuning the amplitude of the support of the upper bound of each membership function. In this manner a good management of the uncertainty, associated with the definition of the fuzzy terms, is provided to the system.**

We show the goodness of our methodology by comparing its performance with respect to the one provided by the initial system in this specific medical case. First, we study the global classification improvement and then, we carry out an exhaustive analysis of the behavior of our approach in which we observe the enhancement achieved in several specific situations.

***Index Terms*—Linguistic Fuzzy Rule-Based Classification Systems, Interval-Valued Fuzzy Sets, Tuning, Ignorance Functions, Genetic Fuzzy Systems, Cardiovascular Disease.**

I. Introduction

Fuzzy Rule-Based Classification Systems (FRBCS) [13] are useful tools to deal with classification problems. They are widely employed because of their capability to built a linguistic model interpretable by the users. Moreover, they offer the possibility of mixing information coming from different sources, i.e. as expert knowledge, mathematical models or empirical measures.

A key problem of representing the knowledge by means of fuzzy sets is to choose the membership function which best represents the information. Sometimes, it is really difficult to select the membership degree of each element to the fuzzy set. This problem led Zadeh to suggest the notion of type-2 fuzzy sets as an extension of fuzzy sets [28]. One of the most used extensions are the Interval-Valued Fuzzy Sets (IVFSs), which are a particular case of Type-2 Fuzzy Sets. IVFSs were introduced by Sambuc in 1975 [22]. IVFSs have been successfully applied in tasks such as classification [23], [26] or image processing [3], among others.

In [23], authors introduced a method to enhance the performance of FRBCSs by extending the Knowledge Base (KB) with IVFSs. That is, it is proposed to characterize the linguistic labels that compound the attributes of the problem by means of IVFSs. Thereafter, they apply a post-processing genetic tuning step in which the amplitude of the support of the IVFSs is tuned allowing the improvement of the classification accuracy.

The aim of this work is to apply this methodology on the classification of patients with respect to their risk of suffering cardiovascular diseases (CVDs). These diseases are the ones affecting to the heart and to the arteries, mainly the arteries of the brain, legs and heart. Most of these disorders are induced because of the decrease of either the caliber or the diameter of the arteries, which is due to the presence of fat concentrations adhered to the artery walls leading to hinder the blood flow.

However, the lack of supply blood is not only manifested in the heart but also in the legs and in the brain, leading to disorders for the patient implying the risk of heart attacks, thrombosis or rupture of blood vessels, among others. CVDs conform the main health problem in adult population in general, being in the first place of the list of death cause of persons older than forty five years in many countries. As an example, about 100,000 persons per year die in Spain due to these diseases, representing a death rate of 75-150 deaths per 100,000 inhabitants depending on the region. This rate is similar in most of the developed countries [10]. Therefore, it is really important to obtain a quick diagnosis and to estimate the patients' risk of developing a CVD in order to allow the quick start of the treatment and the reduction of the risk.

In order to estimate the risk, the doctors look up specific tables called REGICOR [21] in which different features are considered: gender, age, presence or absence of diabetes, systolic and diastolic blood pressure, total cholesterol and the value of the HDL cholesterol. The value obtained with this procedure quantifies the risk of the patient of suffering a CVD in ten years.

978-1-61284-076-5/11/$26.00 ©2011 IEEE

In this contribution, we are going to obtain a FRBCS to predict the risk of suffering a CVD, using as input values only the physical values that can be measured directly by the doctor, i.e. the gender, age, smoking, blood pressure and the body mass index. The objective of the system is to give the doctor a quick and reliable estimation of the risk, in such a way that he/she can use our system as a help to decide the necessity of making more analysis (cholesterol, triglycerides, etc.) to obtain a more accurate value of the risk and to start the treatment. We must point out that the resulting system can be more useful in developing countries since blood tests are not usually performed routinely.

The initial KB used as the base in which apply our methodology is generated by means of two basic and well-known fuzzy rule learning algorithms, namely the Chi et al.'s method [4] and the Fuzzy Hybrid Genetics-Based Machine Learning (FH-GBML) algorithm by Ishibuchi and Yamamoto [15]. We will study the behavior of our methodology with respect to the base FRBCSs both when considering the classification accuracy and when analyzing in detail scenarios which should be fulfilled in this specific problem.

The paper is organized as follows: we present in Section II the basic concepts employed in the paper for FRBCSs. In Section III we describe in detail the IVFSs model, showing the modifications introduced in the Fuzzy Reasoning Method (FRM). The proposal to tune the amplitude of the upper bound of the IVFSs is introduced in Section IV. Section V shows our experimental framework and the experimental analysis carried out. Finally, the conclusions of this work are presented in Section VI.

II. Fuzzy Rule Based Classification Systems And Fuzzy Learning Method

FRBCSs are a very useful tool in Data Mining, since they allow the inclusion of all the available information in system modeling, both the one that comes from expert knowledge and the one from empirical measures and mathematical models, deriving on a very interpretable model and, therefore, allowing the knowledge representation to be understandable for the system users.

Any classification problem consists of m training patterns $x_p = (x_{p1}, \ldots, x_{pn}, y_p)$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern.

In this work we use fuzzy rules of the following form for our FRBCSs:

$$\text{Rule } R_j: \quad \text{If} \quad x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \quad \text{then} \quad \text{Class} = C_j \text{ with } RW_j,$$

where 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 (we use triangular membership functions), C_j is a class label, and RW_j is the rule weight [12]. Specifically, in this work the rule weight is computed using the Penalized Certainty Factor (PCF) defined in [14] as:

$$PCF_j = \frac{\sum_{x_p \in ClassC_j} \mu_{A_j}(x_p) - \sum_{x_p \notin ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} \quad (1)$$

.

Fuzzy learning methods are the basis to build a FRBCS. The algorithms used in this work are: 1) the method proposed in [4], that we have called the Chi et al.'s rule generation; 2) the Fuzzy Hybrid Genetics-Based Machine Learning (FH-GBML) algorithm proposed by Ishibuchi and Yamamoto in [15].

A. Chi et al.'s Rule Generation

To generate the fuzzy rule base, this FRBCSs design method determines the relationship between the variables of the problem and establishes an association between the space of the features and the space of the classes by means of the following steps:

1) *Establishment of the linguistic partitions.* Once the domain of variation of each feature A_i is determined, the fuzzy partitions are computed.
2) *Generation of a fuzzy rule for each example* $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$. To do this it is necessary:
   - 2.1 To compute the matching degree $\mu(x_p)$ of the example to the different fuzzy regions using a conjunction operator (usually modeled with a minimum or product T-norm).
   - 2.2 To assign the example x_p to the fuzzy region with the greatest membership degree.
   - 2.3 To generate a rule for the example, whose antecedent is determined by the selected fuzzy region and whose consequent is the label of class of the example.
   - 2.4 To compute the rule weight.

We must remark that rules with the same antecedent can be generated during the learning process. If they have the same class in the consequent we just remove one of the duplicated rules, but if they have a different class only the rule with the highest weight is kept in the rule base.

B. Fuzzy Hybrid-Genetic Based Machine Learning Rule Generation Algorithm

Different GFSs have been proposed in the specialized literature for designing fuzzy rule-based systems in order to avoid the necessity of linguistic knowledge from domain experts [5], [8], [11], [20].

The basis of the method described here, the FH-GBML algorithm [15], consists of a Pittsburgh approach where each rule set is handled as an individual. It also contains a Genetic Cooperative Competitive Learning (GCCL) approach (an individual represents a unique rule), which is used as a kind of heuristic mutation for partially modifying each rule set, because of its high search ability to efficiently find good fuzzy rules.

This method simultaneously uses four fuzzy set partitions for each attribute, as shown in Figure 1. As a result, each antecedent attribute is initially associated with 14 fuzzy sets generated by these four partitions as well as a special "do not care" set, i.e., 15 in total.

The main steps of this algorithm are described below:

Step 1: Generate N_{pop} rule sets with N_{rule} fuzzy rules.

Step 2: Calculate the fitness value of each rule set in the current population.

Step 3: Generate (N_{pop} -1) rule sets by selection, crossover and mutation in the same manner as the Pittsburgh-style algorithm. Apply a single iteration of the GCCL-style algorithm (i.e., the rule generation and the replacement) to each of the generated rule sets with a pre-specified probability.

Step 4: Add the best rule set in the current population to the newly generated (N_{pop} -1) rule sets to form the next population.

Step 5: Return to Step 2 if the pre-specified stopping condition is not satisfied.

Next, we will describe every step of the algorithm:

- Initialization: N_{rule} training patterns are randomly selected. Then, a fuzzy rule from each of the selected training patterns is generated by choosing probabilistically (as shown in (2)) an antecedent fuzzy set from the 14 candidates $B_k (k = 1, 2, \ldots, 14)$ (see Figure 1) for each attribute. Then each antecedent fuzzy set of the generated fuzzy rule is replaced with *don't care* using a pre-specified probability $P_{don't\ care}$.

$$P(B_k) = \frac{\mu_{B_k}(x_{pi})}{\sum_{j=1}^{14} \mu_{B_j}(x_{pi})} \tag{2}$$

- Fitness computation: The fitness value of each rule set S_i in the current population is calculated as the number of correctly classified training patterns by S_i. For the GCCL approach the computation follows the same scheme, counting the number of correct hits for each single rule.
- Selection: It is based on binary tournament.
- Crossover: The substring-wise and bit-wise uniform crossover are applied in the Pittsburgh part. In the case of the GCCL part only the bit-wise uniform crossover is considered.
- Mutation: Each fuzzy partition of the individuals is randomly replaced with a different fuzzy partition using a pre-specified mutation probability for both approaches.

For more details about this proposal, please refer to [15].

III. Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets

In this section we present the model that employs IVFSs to represent the linguistic labels of FRBCSs. The use of IVFSs allows to handle the uncertainty associated with the ad-hoc construction of fuzzy partitions and, in this way, it is possible to increase the performance of the system.

In the remaining of this section, we briefly introduce the IVFSs and then we describe in detail de FRBCSs with the linguistic labels modeled by means of IVFSs and also the adaptation of the Fuzzy Reasoning Method (FRM) in order to work with this representation.

A. Interval-Valued Fuzzy Sets

The IVFSs [1] are an extension of Fuzzy Sets [27]. In 1975 Sambuc presented the concept of IVFS in his doctoral thesis. He applied IVFSs to medical diagnosis in thyroid pathology [22]. Later, in the eighties, Gorzalczany [9] and Turksen [25] gave relevance to the IVFSs and were definitively established. We must point out that Interval Type-2 Fuzzy Sets (IT2FSs) are a particular case of Type-2 Fuzzy Sets. In [16], [17], [18] is proved that IVFSs are a particular case of IT2FSs.

We denote by $L([0,1])$ the set of all closed subintervals of the closed interval $[0,1]$; that is:

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\} \ .$$

$L([0,1])$ is a partially ordered set with respect to the relation $\leq_L$ defined in the following way; given **x, y** $\in L([0,1])$: **x** $\leq_L$ **y** if and only if $\underline{x} \leq \underline{y}$ and $\overline{x} \leq \overline{y}$.
$(L([0,1]), \leq_L)$ is a complete lattice where the smallest element is $0_L = [0,0]$ and the largest is $1_L = [1,1]$.

The following definition can be found in [2], [18], [19], [24]:

Definition 1: An Interval-valued fuzzy set (IVFS) A on the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$.

Obviously, $A(u) = [\underline{A}(u), \overline{A}(u)] \in L([0,1])$ is the membership degree of u $\in U$.

B. IVFS Model

As we have stated previously, we are going to model the linguistic labels by means of IVFSs. To do so, we generate the initial KB by means of any rule learning algorithm (the two previous ones introduced in Section II in this case). Then, starting from the fuzzy sets which compose the initial KB we construct each IVFS as follows:

- The lower bound corresponds to the initial membership function.
- The upper bound is centered in the maximum of the lower bound (being symmetrical in both sides) and its amplitude of the support is 50% greatter than the one of the intial membership function.

As we construct the IVFSs after the rule generation process, we will only study their influence in the FRM.

Due to the modeling of the linguistic labels by means of IVFSs, the RW will be compounded by a tuple (PCF_{Lj}, PCF_{Uj}) whose computation will be done following the Expression (1), considering the lower and the upper bounds as the terms in each case. That is:

$$PCF_{Lj} = \frac{\sum_{x_p \in ClassC_j} \underline{A_j}(x_p) - \sum_{x_p \notin ClassC_j} \underline{A_j}(x_p)}{\sum_{p=1}^{m} \underline{A_j}(x_p)} \tag{3}$$

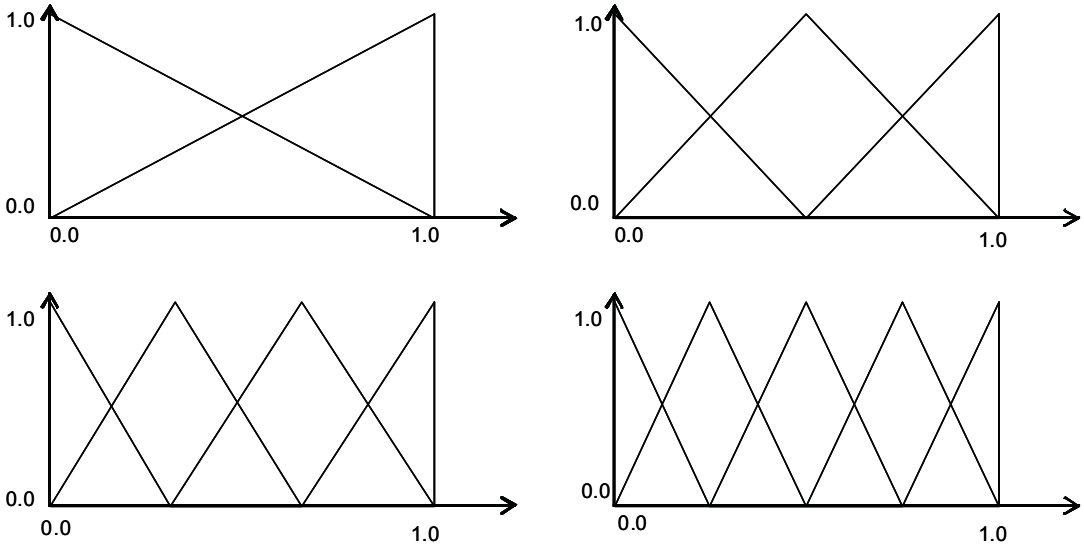

Fig. 1. Four fuzzy partitions for each attribute membership function

$$PCF_{Uj} = \frac{\sum_{x_p \in ClassC_j} \overline{A_j}(x_p) - \sum_{x_p \notin ClassC_j} \overline{A_j}(x_p)}{\sum_{p=1}^{m} \overline{A_j}(x_p)} \quad (4)$$

As the lower bound of each IVFS is the same fuzzy set used by the rule learning algorithm, the rule weight associated with the lower bound (PCF_{Lj}) is equal to RW_j.

The use of IVFSs also implies the following two changes in the FRM:

- *Matching degree between the antecedent of the rule and the example*: We apply a T-norm both to the lower bound and the upper bound as follows:

$$\mu_L A_j(x_p) = T(\underline{A_{j1}}(x_{p1}), \ldots, \underline{A_{jn}}(x_{pn})), \quad j = 1, \ldots, L. \quad (5)$$

$$\mu_U A_j(x_p) = T(\overline{A_{j1}}(x_{p1}), \ldots, \overline{A_{jn}}(x_{pn})), \quad j = 1, \ldots, L. \quad (6)$$

  Therefore, the matching degrees obtained form the following interval:

$$[\mu_L A_j(x_p), \mu_U A_j(x_p)]$$

- As *association degree* we take the mean between the product of the matching degree by the rule weight associated with the lower and the upper bound respectively. That is,

$$b_j^k = \frac{\mu_L A_j(x_p) * PCF_L j^k + \mu_U A_j(x_p) * PCF_U j^k}{2} \quad k = 1, \ldots, M, \quad j = 1, \ldots, L. \quad (7)$$

At this point we already have a single value associated with the class. Therefore, we can apply the rest of the algorithm as in the general FRM.

IV. Genetic Amplitude Tuning of the Upper Bound of the IVFS

The length of the IVFSs can be seen as a representation of the uncertainty associated with the definition of the membership functions. In the initial construction of the IVFSs we have added the upper bound of all the linguistic labels making their amplitude greater than the one of the lower bound in the same proportion in all cases. In this manner, we consider the same uncertainty degree for all the fuzzy labels but, as the amount of available information can be different depending on the variable, the length of each IVFSs can vary.

In order to look for the optimal amount of uncertainty each IVFS represents, we propose a post-processing genetic tuning step in which we perform slight changes to the amplitude of the support of the IVFSs. In this manner, we contextualize the fuzzy partitions for each specif problem leading to an improvement of the classification accuracy.

The modification of the amplitude is given by a number within the interval [0, 1], that is, from the situation in which both bounds are the same (value 0) to the situation in which the amplitude of the upper bound is twice than the amplitude of the lower bound (value 1). The amplitude of the upper bound will be uniformly increased according to intermediate values. The noticeable situations are depicted in Figure 2.

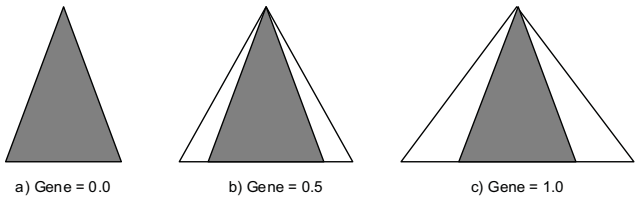

Fig. 2. Gene values representation in the genetic amplitude tuning. a) Upper and lower bounds are the same. b) Initial construction of the IVFSs. c) Upper bound amplitude is twice than the one of the lower bound

In order to apply the genetic tuning, we will consider the use of CHC algorithm [7], which presents a good trade-off between diversity and convergence, being a good choice in complex problems. The components needed to design this process are explained below:

1) *Coding Scheme:* A real coding is considered, where each gene of the chromosome represents the modification of the amplitude of the support as defined above. Thus, there are as many genes as fuzzy partitions in the Data Base.
2) *Chromosome Evaluation:* The fitness function is the classification accuracy.
3) *Initial Gene Pool:* The initial pool is obtained with the first individual having all genes with value 0.5 (the initial FRBCS). The second and the third individuals having all genes with values 0 and 1 respectively, whereas the remaining individuals are generated at random in [0, 1].
4) *Crossover Operator:* We consider the Parent Centric BLX (PCBLX) operator, which is based on the BLX-α. We consider the incest prevention mechanism, checking and modifying an initial threshold, in order to apply the PCBLX operator.
5) *Restarting approach:* When the threshold value is lower than zero, all the chromosomes are regenerated at random within the interval [0, 1]. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm.

V. Experimental Study

The experimental study is aimed to show the improvements achieved by the application of our methodology with respect to the initial FRBCSs employed in this work. We will show the global improvement and the advantages of the application of our approach for both the patient and the health institution.

In the remainder of this section, we will first describe the experimental framework and then we analyze the achieved results.

A. Experimental Framework

In this section we describe the data used in the generation of the FRBCS to predict the risk of suffering a CVD.

In 1948, Framingham Heart Study [6] led to the identification of the major CVD risk factors: high blood pressure, high blood cholesterol, smoking, obesity, diabetes, and physical inactivity, as well as a great deal of valuable information on the effects of related factors such as blood triglyceride and HDL cholesterol levels, age, gender, and psychosocial issues.

Nowadays, the Framingham tables are being adapted to the features of the spanish population by means of a well contrasted calibration process which is under validation. They allow to quantify the risk of a heart problem like angina pectoris or myocardial infarction in ten years.

In the Spanish tables called REGICOR [21], using the values of gender, diabetes, smoking, systolic blood pressure, diastolic blood pressure, total cholesterol and cholesterol HDL we can obtain the risk that can be: Low, Mild, Moderate or High.

The dataset consist of 904 cases obtained from the clinical records of seven health centers of Pamplona (Navarra, Spain) during 2008. We have followed the Spanish Law of personal data protection (LOPD). In each clinical case, a doctor has assigned a REGICOR risk value following the REGICOR tables. Furthermore, the doctor also has taken into account all the data in the medical history of the patient. Therefore, some data can differ from the one recorded in the tables since the doctor takes into account his/her own knowledge.

In the data set there exist 300 cases of class Low, 300 of class Mild, 300 of class Moderate and 4 of class High[1]. All of them are not diabetic but a similar FRBCS using the data from diabetic patients could be obtained.

As input values we are going to use only the following attributes:

- Gender.
- Smoking.
- Blood systolic pressure.
- Blood diastolic pressure.
- Body mass index.

These attributes are collected to provide a fast diagnosis tool to the doctor, since all of them can be obtained in a simple medical encounter. In this manner, the doctor can decide the necessity of making more medical tests, i.e. the total cholesterol or the value of the HDL cholesterol, in order to obtain a more suitable risk degree.

To carry out the experiment we have considered a *5-folder cross-validation model*, i.e., five random partitions with a twenty per cent of the patterns, using the combination of four of them (eighty per cent) for training and the remaining one for testing. For each data-set we consider the average results of the five partitions.

We will apply the same configuration for both FRBCS approaches (Chi and FH-GBML), which consists of product t-norm as conjunction operator, together with PCF for the rule weight and the FRM of the winning rule. Furthermore, we have selected the use of three labels per variable in the case of the Chi FRBCS.

For the FH-GBML algorithm, we consider the following values for the specific parameters of the genetic process:

[1]It is really difficult to find people with such risk level due to they are used to go to health centers and therefore they are treated so their risk is usually low

- Number of fuzzy rules: $5 \cdot d$ rules.
- Number of rule sets: 200 rule sets.
- Crossover probability: 0.9.
- Mutation probability: $1/d$.
- Number of replaced rules: All rules except the best-one (Pittsburgh-part, elitist approach), number of rules / 5 (GCCL-part).
- Total number of generations: 1000 generations.
- Don't care probability: 0.5.
- Probability of the application of the GCCL iteration: 0.5.

where d stands for the dimensionality of the problem (number of variables).

Finally, we have considered the following values for the parameters of the genetic tuning:

- Population Size: 50 individuals.
- Number of evaluations: 5000 · number of variables.
- Bits per gene for the Gray codification (for incest prevention): 30 bits.

B. Analysis of the Usefulness of the Genetic Amplitude Tuning

Table I shows the results achieved by the two FRBCSs used in this work, both when the KB is composed of fuzzy sets (Base) and when the KB is formed of IVFSs which are post-processed with the Genetic Amplitude Tuning (IVFS_GAT). These results are grouped by pairs of columns, training and testing, and the best result achieved in test is stressed in **boldface**.

TABLE I
RESULTS IN TRAIN (TR.) AND TEST (TST) OF BOTH METHODS WITH AND WITHOUT THE GENETIC AMPLITUDE TUNING.

Chi			
Base		IVFS_GAT	
$Tr.$	Tst	$Tr.$	Tst
72.42	70.28	75.32	**72.44**
FH-GBML			
Base		IVFS_GAT	
$Tr.$	Tst	$Tr.$	Tst
73.96	71.85	77.07	**74.00**

Results of Table I shown that our methodology allows the improvement of the classification accuracy of the initial FRBCS in both cases. Specifically, we raise the performance of both base classifiers a 2% which is an important improvement.

From now on, we are going to perform a deep analysis of the behavior of our methodology in two specific scenarios which can provide benefits both for the patients and for the medical institutions. We use confusion matrices as they allow us to show easily the number of correctly classified patterns and the class in which the patterns are classified when they are misclassified.

Table II shows the confusion matrices of the results provided both in training and testing by the Chi et al. rule learning algorithm. In the first part of the table are presented the results when the KB is composed of fuzzy sets and, in the second part, the results achieved when the linguistic labels are modeled by means of IVFSs and are genetically post-processed. Table III presents the confusion matrices for the FH-BML algorithm with and without our methodology and follows the same structure of Table II.

TABLE II
CONFUSION MATRIX IN TRAINING AND TESTING OF THE CHI ET AL. ALGORITHM WITH AND WITHOUT GENETIC AMPLITUDE TUNING.

Base Chi et al. algorithm									
Train					Test				
Class	0	1	2	3	Class	0	1	2	3
0	208	6	7	0	0	50	4	1	0
1	52	93	77	0	1	14	24	21	0
2	11	35	171	0	2	3	10	40	0
3	0	0	2	0	3	0	0	0	0
Chi et al. algorithm with IVFS_GAT									
Train					Test				
Class	0	1	2	3	Class	0	1	2	3
0	213	3	6	0	0	53	1	1	0
1	39	127	56	0	1	12	34	13	0
2	9	51	157	0	2	2	15	36	0
3	0	0	2	0	3	0	0	0	0

The first specific scenario is the one in which the patient would have a low risk degree of suffering a CVD. If the patient is not classified with such risk degree he/she would have to pass more medical tests producing over cost to the health institution. From Tables II and III it is observed that after the application of our methodology diminishes notably the number of misclassification of patients with a low risk degree in both FRBCSs.

Another important situation is the one in which the patient would have a mild or moderate risk of suffering a CVD. The patient should be exhaustive analyzed, since if its diagnosis underestimate the risk degree his/her health would be in danger. From the results shown in Tables II and III it is noticed that the number of patients classified with a low risk diminishes in both cases. However, when applying our methodology to the KB generated by the Chi et al. algorithm most of the patients are classified with a mild risk. Furthermore, when applying our approach to the KB created by the FH-GBML method, most of the patients are classified with a moderate risk.

These findings allow us to assert that our methodology is suitable to face this problem. Specifically, we recommend the use of our methodology with the KB generated by the FH-GML algorithm, since it fulfills the suitable situations. Furthermore, in the case of having doubts between mild or moderate risk, it shows a trend to classify the patients with moderate risk which is good for them, since the subsequent medical tests they have to pass will allow to quantify accurately their risk.

VI. CONCLUSION AND FUTURE LINES

In this contribution we have presented a methodology for improving the classification accuracy of FRBCSs. First, we model the linguistic labels by means of IVFSs in order to take into account the semantic uncertainties related to the definition of the membership functions and then, we post-process each IVFS in such a way that their amplitude is tuned leading to a good management of the uncertainties of the system.

We have applied our methodology to the detection of the risk of the patients of suffering a CVD. The experimental study

TABLE III
CONFUSION MATRIX IN TRAINING OF THE FH-GML ALGORITHM WITH AND WITHOUT GENETIC AMPLITUDE TUNING.

Base FH-GBML algorithm										
Train						Test				
Class	0	1	2	3		Class	0	1	2	3
0	187	32	3	0		0	45	8	2	0
1	17	144	61	0		1	6	36	17	0
2	5	53	159	0		2	0	16	37	0
3	0	0	2	0		3	0	0	0	0
FH-GBML algorithm with IVFS_GAT										
Train						Test				
Class	0	1	2	3		Class	0	1	2	3
0	207	10	5	0		0	51	1	3	0
1	15	127	80	0		1	5	30	24	0
2	2	32	183	0		2	0	10	43	0
3	0	0	2	0		3	0	0	0	0

has shown the suitability of our approach, since it enhances the results of the both base FRBCSs considered in this work. We stress the goodness of our methodology applied to the KB generated by the FH-GBML algorithm, since it improves the behavior of the classifier in all the specific scenarios we have analyzed.

In future works doctors will verify the results and further research will be done in order to increase the performance of the FRBCSs.

ACKNOWLEDGMENTS

This work was supported in part by the Spanish Ministry of Science and Technology under projects TIN2008-06681-C06-01 and TIN2010-15055.

REFERENCES

[1] H. Bustince, J. Montero, M. Pagola, E. Barrenechea and D. Gomez, A survey of Interval-Valued Fuzzy Sets, *In: Handbook of Granular Computing*, Addison Wesley, pp. 491 – 516, (2008)

[2] H. Bustince, E. Barrenechea and M. Pagola, Generation of interval-valued fuzzy and Atanassov's intuitionistic fuzzy connectives from fuzzy connectives and from K_α operators: laws for conjunctions and disjunctions, amplitude, *Int. J Intell. Syst.* vol. 23, no. 6, 2008, pp. 680–714.

[3] H. Bustince, E. Barrenechea, M. Pagola and J. Fernandez, Interval-valued fuzzy sets constructed from matrices: Application to edge detection, *Fuzzy Set Syst.* vol. 160, num. 13, 2009, pp. 1819–1840.

[4] Z. Chi, H. Yan and T. Pham, Fuzzy algorithms with applications to image processing and pattern recognition, *World Scientific* (1996)

[5] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann and L. Magdalena, Ten years of genetic fuzzy systems: Current framework and new trends, *Fuzzy Set Syst.* vol. 141, num. 1, 2004, pp. 5–31.

[6] T. R. Dawber, G. F. Meadors and F. E. Moore Jr., Epidemiological approaches to heart disease: the Framingham Study, *Am. J. Public Health* vol. 41, num. 3, 1951, pp. 279–281.

[7] L. J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, *In: Foundations of Genetic Algorithms*, Morgan Kaufman, pp. 265–283, (1991)

[8] M. J. Gacto, R. Alcalá, F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. *Soft Comput.* vol. 13, num. 3, 2009, pp. 419–436.

[9] M. B. Gorlzakczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, *Fuzzy Set Syst.*, vol. 21, 1987, pp. 1–17.

[10] M. Grau, J. Marrugat, Funciones de riesgo en la prevención primaria de las enfermedades cardiovasculares, *Rev. Esp. Cardiol.* vol. 61, 2008, pp. 404–416.

[11] F. Herrera, Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence* vol. 1, 2008, pp. 27–46.

[12] H. Ishibuchi and T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, *IEEE T. Fuzzy Syst.* vol. 9, num. 4, 2001, 506–515.

[13] H. Ishibuchi, T. Nakashima and M. Nii, Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining, *Springer-Verlag* (2004)

[14] H. Ishibuchi and T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, *IEEE T. Fuzzy Syst.* vol. 13, 2005, pp. 428–435.

[15] H. Ishibuchi, T. Yamamoto and T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, *IEEE T. Syst. Man. Cy. B* vol. 35, num. 2, 2005, pp. 359–365.

[16] E. E. Kerre, A first view on the alternatives of fuzzy sets theory, *in: B. Reusch, K-H Temme (Eds), Computational intelligence in Theory and Practice, Psyshica-Verlag, Heidelberg* pp. 55–72, 2001.

[17] G. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, *Prentice Hall, Upper Saddle River, NJ*, 1995.

[18] J. M. Mendel and R. I. J. Robert, Type-2 fuzzy sets made simple, *IEEE T. Fuzzy Syst.* vol. 10, 2002, pp. 117–127.

[19] J. M. Mendel, Advances in type-2 fuzzy sets and systems, *Inform. Sciences* vol. 177, no. 1, 2007, pp. 84–110.

[20] Y. Nojima, H. Ishibuchi and I. Kuwajima, Parallel distributed genetic fuzzy rule selection, *Soft Comput.* vol. 13, num. 3, 2009, pp. 511–519.

[21] J. Sala, J. Marrugat, R. Masia and M. Porta, Improvement in survival after myocardial infarction between 1978-85 and 1986-88 in the REGICOR study (REgistre GIroni del COR) registry, *Eur Heart J.* vol. 16, 1995, pp. 779–84.

[22] R. Sambuc, Function ϕ-flous, application a l'aide au diagnostic en pathologie thyroidienne, *Ph.D. thesis*, University of Marseille. (1975)

[23] J. Sanz, A. Fernandez, H. Bustince and F. Herrera, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning, *Inform. Sciences* vol. 180, no. 19, 2010, pp. 3674–3685

[24] H. R. Tizhoosh, Image thresholding using type II fuzzy sets, *Pattern Recogn.* vol. 38, no. 12, 2005, pp. 2363–2372.

[25] I. B. Turksen, Interval valued fuzzy sets based on normal forms, *Fuzzy Set Syst.* vol. 20, no. 2, 1986, pp. 191–210

[26] H. Wu and J. M. Mendel, Classification of Battlefield Ground Vehicles Using Acoustic Features and Fuzzy Logic Rule-Based Classifiers, *IEEE T. Fuzzy Syst.* vol. 15, no. 1, 2007, pp. 56–72

[27] L. A. Zadeh, Fuzzy sets, *Information and Control* vol. 8, no. 3, 1965, pp. 338–353.

[28] L. A. Zadeh, Quantitative fuzzy semantics, *Inform. Sciences* vol. 3, no. 2, 1974, pp. 159–176.