A Compact Evolutionary Interval-Valued Fuzzy Rule-Based Classification System for the Modeling and Prediction of Real-World Financial Applications with Imbalanced Data

José Antonio Sanz, Dario Bernardo, Francisco Herrera, *Member, IEEE*, Humberto Bustince, *Member, IEEE* and Hani Hagras, *Fellow, IEEE*

Abstract— The current financial crisis has stressed the need of obtaining more accurate prediction models in order to decrease the risk when investing money on economic opportunities. In addition, the transparency of the process followed to make the decisions in financial applications is becoming an important issue. Furthermore, there is a need to handle the real-world imbalanced financial data sets without using sampling techniques which might introduce noise in the used data. In this paper, we present a compact evolutionary interval-valued fuzzy rule-based classification system, which is based on IVTURS$_{FARC-HD}$ (Interval-Valued fuzzy rule-based classification system with TUning and Rule Selection) [22]), for the modeling and prediction of real-world financial applications. This proposed system allows obtaining good predictions accuracies using a small set of short fuzzy rules implying a high degree of interpretability of the generated linguistic model. Furthermore, the proposed system deals with the financial imbalanced datasets with no need for any preprocessing or sampling method and thus avoiding the accidental introduction of noise in the data used in the learning process. The system is also provided with a mechanism to handle examples that are not covered by any fuzzy rule in the generated rule base. To test the quality of our proposal, we will present an experimental study including eleven real-world financial datasets. We will show that the proposed system outperforms the original C4.5 decision tree, type-1 and interval-valued fuzzy counterparts which use the SMOTE sampling technique to preprocess data and the original FURIA, which is a fuzzy approximative classifier. Furthermore, the proposed method enhances the results achieved by the cost sensitive C4.5 and it gives competitive results when compared with FURIA using SMOTE, while our proposal avoids pre-processing techniques and it provides interpretable models that allow obtaining more accurate results.

Index Terms— Financial applications, Interval-Valued Fuzzy Sets, Interval-Valued Fuzzy Rule-Based Classification Systems, Evolutionary algorithms.

1. INTRODUCTION

The recent financial crisis highlighted fundamental weaknesses in the long term global approach to financial modeling and prediction. Hence, there is a need for new more comprehensive, transparent and accurate financial modeling and prediction approaches to capitalize on economic opportunities without incurring high levels of unexpected risk.

Many financial applications rely on the expertise of their staff to make a judgment call even when the factors in consideration are too broad and complex to be adequately assessed by the human brain. This

might result in the risks being assessed incompletely and inaccurately with a lack of decision consistency. For example, loan officers usually apply the rule of the five C principles (Capacity, Capital, Character, Collateral, and Conditions) to decide whether to grant a loan or not. In order to make accurate and consistent decision with this rule it would be necessary to have complete knowledge of the applicant and there is a need for consistency across the loan officers where each officer should make the same decision for the same applicant, which might not be the case.

In financial applications, as in many real-world problems, the data is highly imbalanced. For example, in a credit card application the number of good customers is much higher than that of bad customers and in fraud detection the majority of the data are normal transactions whereas a few fraudulent transactions are usually present. Most classifiers designed for minimizing the global error rate perform poorly on imbalanced datasets because they misclassify most of the data belonging to the class represented by few examples [1], [2]. To tackle this problem, pre-processing techniques like under-sampling or over-sampling are usually applied but both of them present problems. On the one hand, under-sampling techniques may increment the noise since they could eliminate some important patterns. On the other hand, over-sampling techniques, like Synthetic Minority Oversampling Technique (SMOTE) [3], may add noise for the original input data or violate the inherent geometrical structure of the minority and majority classes [4]. Hence, in financial applications it is not desirable to preprocess or sample the data as this could cause big problems.

Real-world financial problems have been tackled using several machine learning and artificial intelligence techniques. The majority of commercial financial systems rely on statistical regression techniques because they provide good results when facing prediction problems composed of two output categories. Other kinds of machine learning techniques applied in financial domains are support vector machines, which were applied to forecast financial time series [5] and to effectively manage governmental funds to small and medium enterprises [6]. Neural networks were applied in a big number of financial applications [7], [8], [9]. However, the drawback of such advanced machine learning techniques is that although they can give good prediction accuracies, they provide black box models which are very difficult to understand and analyze by a financial analyst. Therefore, they do not fulfil the current common requirements of having an explanation of the reasoning behind a given financial decision.

Trust is the main reason why it is important to provide the end user with easily interpretable models. Regardless of the degree of sophistication of our economies, all transactions still come down to trust. Therefore, transparency is required so that it is possible to know how the given financial models are operating. This need for transparency is reflected in legislations that force financial institutions to disclose the reasoning behind their financial decisions and models.

Decision trees and Fuzzy Rule-Based Systems (FRBSs) are examples of white box transparent models which have been applied for various financial applications **¡Error! No se encuentra el origen de la referencia.**. FRBSs have been successfully applied on credit approval, loan portfolio, bankruptcy prediction and security management **¡Error! No se encuentra el origen de la referencia.**. The main challenge faced in these works when working with real-world data is the high level of data imbalance. Most of the previous works use preprocessing techniques which might introduce noise and uncertainty.

Interval-Valued Fuzzy Rule-Based Classification Systems (IV-FRBCSs) [19], [20] are interpretable classifiers because they use linguistic terms, which are modeled with Interval-Valued Fuzzy Sets [21] (IVFSs), in the antecedents of the rules. IVTURS$_{FARC-HD}$ [22] (Interval-Valued fuzzy rule-based classification system with TUning and Rule Selection) is a novel IV-FRBCS that provides an accurate as well as a transparent model. The inference process of this system uses interval information in all the steps and it applies Interval-Valued Restricted Equivalence Functions (IV-REFs) [23], [24] to measure the equivalence between the interval membership degrees and the ideal interval membership degrees. Furthermore, IVTURS$_{FARC-HD}$ applies an evolutionary algorithm to modify the values used in the construction of the IV-REFs. This system is designed for standard classification problems, which means that it cannot easily deal with the imbalanced data available in financial applications without preprocessing techniques.

In this paper, we will present a compact evolutionary IV-FRBCS based on IVTURS$_{\text{FARC-HD}}$ for the modeling and prediction of financial applications with imbalanced data, with the aim of providing an accurate, comprehensible, transparent and interpretable model. In order to face the usual problems presented in financial domains with imbalanced data and trying to get a small set of short fuzzy rules (interpretable model), we introduce the following techniques:

- A rescaling method to balance the weights of the fuzzy rules associated with the different classes. In this manner, the imbalanced problem is faced internally by the proposed method, which means that sampling/preprocessing techniques are not needed.
- A technique to classify the incoming examples even if they do not match any fuzzy rule in the generated rule base. To do so, the similarity among the uncovered example and the rules is considered.
- An evolutionary process used to perform a rule selection process along with the tuning of both the shape and the lateral position of the IVFSs as well as the values used to construct the IV-REFs. This allows maximizing the classification performance and producing compact and interpretable models.

The quality of our new proposal, which is denoted IVTURS$_{\text{FARC-HD}}$ with Rescaling Rule Weight for Imbalanced classification (IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$), has been tested thorough various experiments using real-world data sets from eleven financial applications. The obtained results, which are statistically supported, show that IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$ outperforms the original C4.5 method [25] as well as type-1 and interval-valued fuzzy counterparts which used the SMOTE sampling technique to preprocess data. Furthermore, our proposal notably enhances the results achieved by the cost sensitive C4.5 [26] and it gives competitive results versus an approximative fuzzy classifier like FURIA [27] when it uses SMOTE. This fact strengthens the quality of our new method because it provides an accurate and interpretable model learned from the original data. Therefore, IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$ avoids using pre-processing techniques and give accurate results as well as producing a reduction in the number of generated rules, which implies providing more transparent and highly interpretable models.

This paper is arranged as follows: Section 2 provides some needed background material and the related work about FRBCSs. The proposed compact evolutionary IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$ is presented in Section 3. The experimental results are shown in Section 4 while the conclusions are drawn in Section 5.

2. BACKGROUND

In this section, we present the background needed to understand the remainder of the paper. We start by presenting some theoretical concepts about IVFSs. Then, we describe the problem of the imbalanced data-sets in classification and finally, we briefly introduce the fuzzy association rules for classification along with the two state-of-the-art fuzzy association rule-based classification models considered in this paper.

2.1 Interval-Valued Fuzzy Sets

Fuzzy Sets (FSs) [28] assign crisp values as membership degrees of the elements to the sets whereas IVFSs [21] assign intervals instead of numbers as membership degrees. IVFSs have been successfully applied in various applications including image processing [23], assessment of soil and water conservation [29] and classification [30] among others.

Let us denote by $L([0,1])$ the set of all closed subintervals in $[0,1]$, that is,

$$L([0,1]) = \left\{ \boldsymbol{x} = \left[\underline{x}, \overline{x}\right] \middle| 0 \leq \underline{x} \leq \overline{x} \leq 1 \right\}. \ (1)$$

We must remark that we will denote an interval in **bold-face** and a crisp value in normal-font, that is, $\boldsymbol{x}$ is an interval and x is a crisp value.

Definition 1 [21,31,32,33] An interval-valued fuzzy set A on the universe $U \neq \emptyset$ is a mapping $A_{IV} : U \rightarrow L([0,1])$, such that

$$A_{IV}(u_i) = \left[\underline{A}(u_i), \overline{A}(u_i)\right] \in L([0,1]), for \ all \ u_i \in U. \ (2)$$

Obviously, $\left[\underline{A}(u_i), \overline{A}(u_i)\right]$ is the interval membership degree of the element u_i to the IVFS A. Our interpretation of the interval membership degree is that the membership degree is a number within the interval but its real value it is not known. The length of the interval membership degree, $L\big(A_{IV}(u_i)\big) =$

$\overline{A}(u_i) - \underline{A}(u_i)$, can be seen as a representation of the ignorance related to the assignment of crisp values as the membership degrees of the elements to the set. Such ignorance degree can be quantified by means of weak ignorance functions [20].

In order to determine the largest interval membership degree, we need to use a total order relationship for intervals. A method to construct different linear orders between intervals can be found in [34], [35]. A particular case of these linear orders is the one defined by Xu and Yager in [36], which is based on the score and accuracy degrees as shown in Equation (3).

$$[\underline{x}, \overline{x}] \leq [\underline{y}, \overline{y}] \text{ if and only if } \underline{x} + \overline{x} < \underline{y} + \overline{y} \text{ or } \underline{x} + \overline{x} = \underline{y} + \overline{y} \text{ and } \overline{x} - \underline{x} \geq \overline{y} - \underline{y} \quad (3)$$

Using this total order relationship, it is easy to prove that $[0,0]$ and $[1,1]$ are the smallest and the largest element of $L([0,1])$, respectively. So, they are the smallest and the largest interval membership degrees.

In addition, we present the interval operations which will be used to make the computation with intervals both in the inference process and in the computation of the rule weight as an element of $L([0,1])$ instead of with numbers. Let $[\underline{x}, \overline{x}], [\underline{y}, \overline{y}]$ be two intervals, with $\underline{x}, \overline{x}, \underline{y}, \overline{y} \in \mathbb{R}^+$, so that $[\underline{x}, \overline{x}] \leq_L [\underline{y}, \overline{y}]$, which means that $\underline{x} \leq \underline{y}, \overline{x} \leq \overline{y}$ (this is not a total order relationship for intervals), and $\underline{y} > 0$ the rules of interval arithmetic are as follows [36]:

- Addition: $[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$ (4)

- Subtraction: $[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [|\underline{y} - \overline{x}|, |\overline{y} - \underline{x}|]$ (5)

- Product: $[\underline{x}, \overline{x}] * [\underline{y}, \overline{y}] = [\underline{x} * \underline{y}, \overline{x} * \overline{y}]$ (6)

- Division: $\frac{[\underline{x}, \overline{x}]}{[\underline{y}, \overline{y}]} = [\wedge\left(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right), \vee\left(\frac{\underline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right)]$ (7)

where $\wedge$ represents the t-norm (minimum) whereas $\vee$ represents the t-conorm (maximum).

Finally, we recall the definition of IV-REFs [23],[24], which are used to measure the similarity between two intervals. These functions are applied in the inference process of the method used as the base of our new proposal, i.e. IVTURS$_{\text{FARC-HD}}$. We also recall the construction method of these functions that is based on automorphisms, which are continuous and strictly increasing functions $\phi: [0,1] \to [0,1]$ so that $\phi(0) = 0$ and $\phi(1) = 1$. For example if we use $\phi(x) = x^a$, each value of $a \in (0, \infty)$ will generate an automorphism.

Definition 2 [23], [24] An interval-valued restricted equivalence function associated with an interval-valued negation $\boldsymbol{N}$ is a function

$$IV - REF: L([0,1])^2 \to L([0,1]) \quad (8)$$

So that:

> *IR1)* $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = IV - REF(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$;
> *IR2)* $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = [1,1]$ if and only if $\boldsymbol{x} = \boldsymbol{y}$;
> *IR3)* $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = [0,0]$ if and only if $\boldsymbol{x} = [1,1]$ and $\boldsymbol{y} = [0,0]$ or $\boldsymbol{x} = [0,0]$ and $\boldsymbol{y} = [1,1]$;
> *IR4)* $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = IV - REF(\boldsymbol{N}(\boldsymbol{x}), \boldsymbol{N}(\boldsymbol{y}))$ with $\boldsymbol{N}$ being an involutive interval-valued negation [31], [38];
> *IR5)* For all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in L([0,1])$, if $\boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$, then $IV - REF(\boldsymbol{x}, \boldsymbol{y}) \geq_L IV - REF(\boldsymbol{x}, \boldsymbol{z})$ and $IV - REF(\boldsymbol{y}, \boldsymbol{z}) \geq_L IV - REF(\boldsymbol{x}, \boldsymbol{z})$.

The construction method of IV-REFs used in this paper, which is based on the construction method of REFs introduced in [39], is the following one:

$$IV - REF\left([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]\right)$$
$$= \begin{bmatrix} T\left(\phi_1^{-1}\left(1 - |\phi_2(\underline{x}) - \phi_2(\underline{y})|\right), \phi_1^{-1}(1 - |\phi_2(\overline{x}) - \phi_2(\overline{y})|)\right), \\ S\left(\phi_1^{-1}\left(1 - |\phi_2(\underline{x}) - \phi_2(\underline{y})|\right), \phi_1^{-1}(1 - |\phi_2(\overline{x}) - \phi_2(\overline{y})|)\right) \end{bmatrix}$$
(9)

where $\phi_1(x) = x^a$ and $\phi_2(x) = x^b$ with $a, b \in [0.01, 100]$, T and S represent a t-norm and a t-conorm respectively. We must point out that this IV-REF is associated with the interval-valued negation $N([\underline{x}, \overline{x}]) = [\phi_2^{-1}(1 - \phi_2(\overline{x})), \phi_2^{-1}(1 - \phi_2(\underline{x}))]$.

Example 1 Let $a = 1$ and $b = 1$, the automorphisms used in Equation 9 are $\phi_1(x) = x$ and $\phi_2(x) = x$ respectively. Let T and S be the minimum and the maximum respectively, Equation 9 can be rewritten as

$$IV - REF\left([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]\right) = [\wedge\left(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|\right), \vee\left(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|\right)] \quad (10)$$

Using Equation (10) as the IV-REF associated

with the negation $N([\underline{x}, \overline{x}]) = [1 - \overline{x}, 1 - \underline{x}]$ and let $\boldsymbol{x} = [\underline{x}, \overline{x}] = [0, 0]$, $\boldsymbol{y} = [\underline{y}, \overline{y}] = [0.3, 0.6]$ and $\boldsymbol{z} = [\underline{z}, \overline{z}] = [1, 1]$ be three intervals, the conditions IR1- IR5 fulfill as shown below

IR1) $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = IV - REF(\boldsymbol{y}, \boldsymbol{x}) \Rightarrow [0.4, 0.7] = [0.4, 0.7]$

IR2) $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = [0.4, 0.7]$ whereas $IV - REF(\boldsymbol{y}, \boldsymbol{y}) = [1, 1]$

IR3) $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = [0.4, 0.7]$ whereas $IV - REF(\boldsymbol{x}, \boldsymbol{z}) = [0, 0]$

IR4) $IV - REF(\boldsymbol{x}, \boldsymbol{y}) = IV - REF(N(\boldsymbol{x}), N(\boldsymbol{y})) \Rightarrow IV - REF([0,0], [0.3, 0.6]) = IV - REF([1, 1], [0.4, 0.7]) \Rightarrow [0.4, 0.7] = [0.4, 0.7]$

IR5) $if \boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$, then $IV - REF(\boldsymbol{x}, \boldsymbol{y}) \geq_L IV - REF(\boldsymbol{x}, \boldsymbol{z})$ and $IV - REF(\boldsymbol{y}, \boldsymbol{z}) \geq_L IV - REF(\boldsymbol{x}, \boldsymbol{z}) \Rightarrow \boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$, therefore $[0.4, 0.7] \geq_L [0, 0]$ and $[0.3, 0.4] \geq_L [0, 0]$

2.2 Imbalanced Data-Sets in Classification

During the last years, as the popularity of data mining is growing, the machine learning techniques have been applied to several real-world problems, like financial problems. Real-world problems usually contain few examples of the concept to be described due to rarity or the cost to obtain it. The learning from these kinds of problems has been identified as one of the main challenges in data mining [40].

The imbalanced data-sets problem are very common in real world financial data-sets, where one or more classes are represented by a large number of examples (known as majority class) while the other classes are represented by only few examples (known as minority class) [41]. This problem causes the classifier to predict the samples of the majority class and completely ignore the minority ones.

An important aspect when dealing with imbalanced data-sets is the selection of an appropriate metric to measure the performance of the proposals. The most straightforward way to evaluate the performance of classifiers is the analysis based on the confusion matrix. Table 1 shows a confusion matrix for a two-class problem. From this table it is possible to extract a number of widely used metrics to measure the performance of learning systems, such as error rate defined in Equation (11) and accuracy defined in Equation (12) as follows:

$$Err = (FP + FN)/(TP + FP + TN + FN) \quad (11)$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} = 1 - Err \quad (12)$$

The accuracy is the most commonly used metric for empirical evaluations but for classification in this framework this metric might lead to erroneous conclusions since the minority class has little impact on accuracy compared to the majority class [42]. Therefore, in the framework of imbalanced problems there are more accurate metrics. For instance, from Table 1 four performance measures can be derived in order to take into account the classification rate of each class independently:

- True positive rate (TP_{rate}): is the percentage of correctly classified examples belonging to the minority class.
- True negative rate (TN_{rate}): is the percentage of correctly classified examples belonging to the majority class.
- False positive rate (FP_{rate}): is the percentage of misclassified examples belonging to the majority class.
- False negative rate (FN_{rate}): is the percentage of misclassified examples belonging to the minority class.

TABLE 1 CONFUSSION MATRIX FOR A TWO-CLASS PROBLEM.

	MINORITY PREDICTION	MAJORITY PREDICTION
MINORITY CLASS	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
MAJORITY CLASS	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

A well-known metric that attempts to maximize the accuracy of each class is the geometric mean, which is defined as follows [43]:

$$GM = \sqrt{TP_{rate} * TN_{rate}} \quad (13)$$

2.2.1 Fuzzy Association Rules for Classification

This section is aimed at providing a brief introduction of fuzzy association rule-based classifiers, since it is the methodology used by the state-of-the-art fuzzy classification techniques used in this paper, which are FARC-HD [44] and IVTURS [22] that are briefly in Sections 2.3.1 and 2.3.2, respectively.

Association discovery is widely used in data mining since it allows interesting knowledge to be discovered in large databases [45]. Association rules represent dependencies among items in a database using expression like $A \rightarrow B$, where A and B are sets of items and $A \cap B \neq \emptyset$ [46]. The use of fuzzy logic in association rules allows both dealing with uncertain and inaccurate data and introducing linguistic terms implying the generation of an interpretable model for the end users.

The task of classification [47], which aims at determining the class to which the patterns belong, can be tackled using fuzzy association rules. In this case, the antecedent part of the fuzzy rules is composed of fuzzy terms whereas the consequent part has the predicted class label and the rule weight which is written as follows:

R_j: If x_1 is A_{j1} and … and x_n is A_{jn} then Class = C_j
$$\text{with } RW_j, \quad (14)$$

where R_j is the label of the rule, $x = (x_1, \ldots, x_n)$ is a n-dimensional example vector, A_{ji} is an antecedent fuzzy set representing the variable i in rule j, C_j is a class label and $RW_j \in [0, 1]$ is the rule weight [48]. The fuzzy rule in Equation (14) can be represented as $A_j \rightarrow C_j$, where $A_j = (A_{j1}, \ldots, A_{jn})$.

Let $x_p = (x_{p1}, \ldots, x_{pn},), p = 1, 2, \ldots, N$ be a set of N labeled examples from M classes of an n-dimensional classification problem. The matching degree of each training example x_p with the antecedent A_q is defined as follows:

$$\mu_{A_j}(x_p) = T\left(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})\right), \quad (15)$$

where $\mu_{A_{ji}}(\cdot)$ is the membership function of the antecedent fuzzy set A_{ji} and T is a t-norm.

The support of the fuzzy rule $A_j \rightarrow C_j$, which can be viewed as the coverage of the training examples by the fuzzy rule, is written as follows

$$Support(A_j \rightarrow C_j) = \frac{\sum_{x_p \in Class\, C_j} \mu_{A_j}(x_p)}{|N|} \quad (16)$$

The confidence of the fuzzy rule $A_j \rightarrow C_j$, which can be viewed as the validity of the fuzzy rule, is written as follows

$$Confidence(A_j \rightarrow C_j) = \frac{\sum_{x_p \in Class\, C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{N} \mu_{A_j}(x_p)} \quad (17)$$

2.2.2 Fuzzy Association Rules for Classification for High Dimensional Problems

In [44], Alcalá-Fdez et al. defined the algorithm known as FARC-HD (Fuzzy Association Rule-based Classification model for High Dimensional problems). This method allowed outperforming the performance provided by ten well-known classifiers. Furthermore, it generates a compact set of rules with a small computational effort. This fuzzy classifier is composed of the following three stages:

1. *Fuzzy association rule extraction for classification*. In this step the rule base is generated. To do so, a search tree is generated for each class in which the frequent itemsets are computed by applying the support (see Equation (16)) and the confidence (see Equation (17)). In this method each item is represented by a fuzzy term and the depth of the tree is limited by a predefined parameter. Once the search tree is completed, a fuzzy association classification rule is generated for each frequent itemset. To do so, the path of the frequent itemset is assigned as the antecedent part, the class of the search tree is set as class label and the confidence is assigned as rule weight.

2. *Candidate rule prescreening*. In this step a pattern weighting scheme is carried out to select the most promising set of rules for each class, since in the first stage a huge number of fuzzy rules can be generated. To do so, each pattern has assigned a weight that is meant to be the strength in which it contributes in the computation of the quality of the rules. The improved weighted relative accuracy measure [44] is applied to compute the quality of the rules. An iterative process is carried out in which in each run the best rule is selected and the weights of the patterns are decreased based on the covering of such rule. This process is repeated until a stopping criterion is fulfilled.

3. *Rule selection and lateral tuning*. The final stage of the method consists of selecting and tuning a set of rules starting from the final rule base obtained in the second stage. To this aim, the tuning of the lateral position of the linguistic labels [49], which is based on the linguistic 2-tuples representation [50], combined with a genetic rule selection process is applied. For the lateral tuning the parameter α, which determines the position of the linguistic labels, is tuned for each linguistic label of the system. Figure 1 shows an example of the lateral displacement of the

linguistic label s_2 to the left since the value of the parameter α is negative (the value of α has to be positive for displacements to the right). Therefore, the number of parameters is the number of variables times the number of linguistic labels used to model each variable. The synergy between both tuning and rule selection enables to contextualize the membership functions to the problem that has been tackled and to obtain a compact rule set having a high degree of cooperation among its rules.
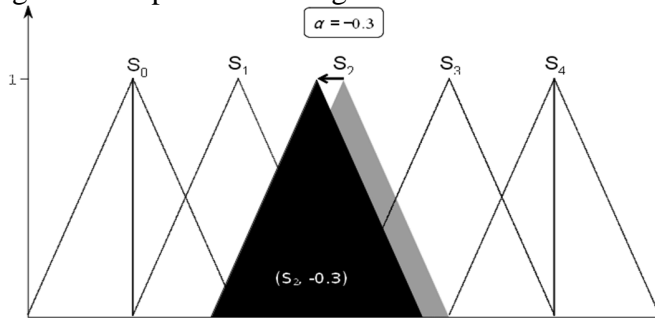

Figure 1 Lateral displacement of the membership function S_2. Grey and black triangles are its initial and final position respectively.

The Fuzzy Reasoning Method (FRM) [51] is the mechanism that uses the fuzzy rules to classify new examples. Specifically, in first place the total vote strength for each class is computed using Equation (18) and then, the example x_p is classified in the class having the maximum total strength of the vote.

$$V_{Class_k}(x_p) =$$
$$f_{R_j \in RB \text{ and } C_j = k}\left(\mu_{A_j}(x_p) * RW_j\right) \text{ with } j = 1, \dots, L \tag{18}$$

where L is the number of rules of the rule base (RB) and f is an aggregation function. The aggregation function can be the maximum or the sum leading to the FRMs of the wining rule or the additive combination, respectively.

2.3.2 Interval-Valued Fuzzy Rule-Based Classification System with Tuning and Rule Selection

IVTURS$_{FARC-HD}$ [22] is an extension of FARC-HD that models the linguistic terms with IVFSs instead of with FSs. IVTURS$_{FARC-HD}$ outperforms the performance of both the original FARC-HD and the Fuzzy Unordered Rule Induction Algorithm (FURIA) [27]. This method is composed of the following steps:

1. *Rule base generation*. This method learns an initial type-1 FRBCS so as to initialize the

parameters of the IV-FRBCS. Specifically, the initial rule base is the one obtained after the application of the two first stages of the FARC-HD algorithm (see stages 1 and 2 of Section 2.3.1).

2. *IVFSs construction*. The second step consists of modeling the linguistic labels which represent the antecedent part of the fuzzy rules with IVFSs. To do so, each IVFS is constructed as follows [19], [20]: 1) the lower bound is the fuzzy set used by the fuzzy learning algorithm to model the corresponding linguistic label and 2) the upper bound is centered around the same apex as the lower bound (being symmetrical in both sides) having a greater support, which is determined by the value of the parameter W. For the initial construction of each IVFS the value of W is set to 0.5 in order to make it 50% greater than that of the lower bound as depicted in Figure 2. In addition, the modeling of the linguistic labels by means of IVFSs implies that the rule weight has to be also an element of $L([0,1])$ instead of a number, which is computed applying Equation (17) using interval arithmetic.
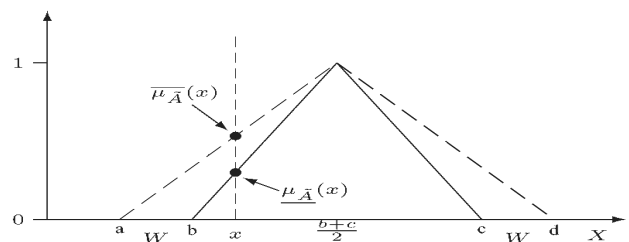

Figure 2 Initially constructed IVFS. The solid line is the initial fuzzy set and consequently it is the lower bound of the IVFS. The dashed line is the upper bound of the IVFS.

3. *Interval-Valued FRM (IV-FRM)*. The classic FRM described in Section 2.3.1 is extended in such a way that it can deal with the representation of the linguistic labels by means of IVFSs. All the steps composing the inference process make their computation using intervals. Specifically, t-representable interval-valued t-norms [33], [38], interval product (see Equation (6)), interval-valued aggregation functions [33], [34] and the total order relationship for intervals are applied to compute the matching degree, the product between the rule weight and the matching degree, the aggregation function f and the final prediction, respectively. Furthermore, the idea of the maximum similarity classifier is introduced in the computation of the
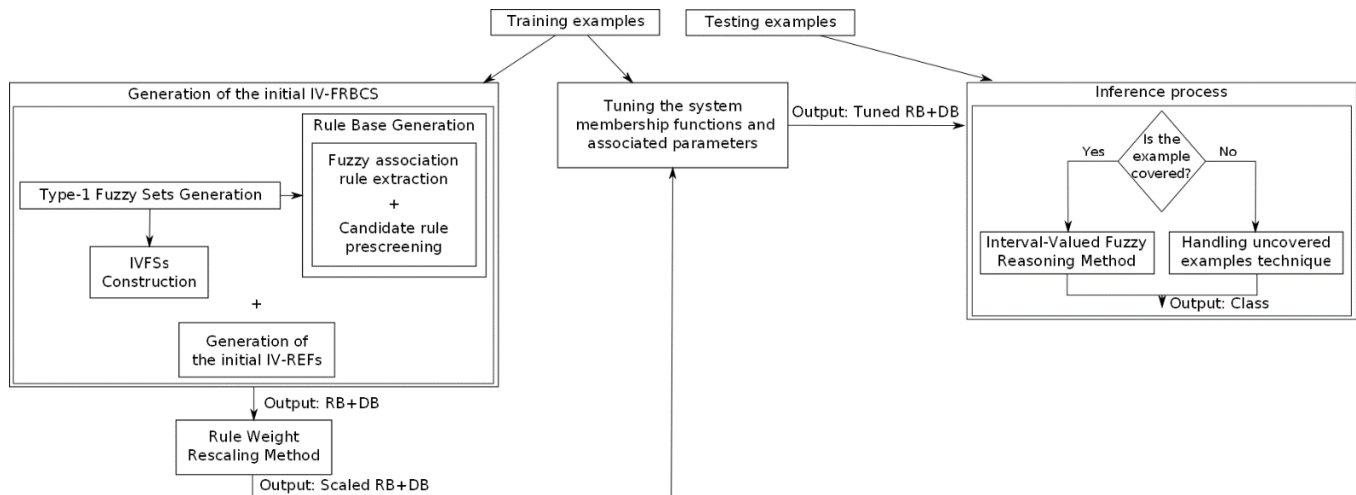
Figure 3 Overview of the proposed compact evolutionary IV-FRBCS for financial applications modeling and prediction.

matching degree. To do so, IV-REFs are applied to compute the equivalence between the interval membership degrees and the ideal membership degree, $[1,1]$, for each attribute of the problem. For the construction of each IV-REF it is necessary to set the value of the parameters a and b in Equation (9), which means that their results may vary depending on these values.

4. *Rule selection and genetic tuning*. The last step is an optimization stage, where a combination of a rule selection process and a tuning approach, as in the previously explained FARC-HD method, is carried out. However, this method does not tune the lateral position of the linguistic labels but the values of the parameters a and b used to construct the IV-REF associated with each variable of the problem. Therefore, the number of parameters to be tuned is two times the number of attributes of the problem [22].

3. THE PROPOSED COMPACT EVOLUTIONARY INTERVAL-VALUED FUZZY RULE-BASED CLASSIFICATION SYSTEM FOR FINANCIAL APPLICATIONS MODELING AND PREDICTION

In this section we present IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$, which is an IV-FRBCS built on the basis of IVTURS$_{\text{FARC-HD}}$ [22], for the modeling and prediction of financial applications which are characterized by highly imbalanced data. The proposed system will handle the imbalanced data (with no need for data pre-processing or sampling) and it is aimed to do the modeling and prediction based on a small set of short

rules which will help to increase the transparency and interpretability of the generated model to the user.

Figure 3 shows an overview of the proposed method, which shares two basic components of the IVTURS$_{\text{FARC-HD}}$ method, namely the generation of the initial IV-FRBCS (applying the two first steps introduced in Section 2.3.2) and the IV-FRM. Our new method generates an initial IV-FRBCS using the training examples and then, the created rule base is scaled using the process introduced in Section 3.1. After this steps, it is applied an evolutionary process to adapt the system's parameters to the problem (using the training examples again), which is described in Section 3.3. When new patterns arrive (testing examples) the method uses the tuned model to classify them: if the example is covered by any fuzzy rule the usual IV-RFM is applied; otherwise, the method defined in Section 3.2 to handle uncovered examples is used.

The novelty of our new method consists of a method to rescale the rule weights of the generated rule base in order to face the imbalance problem at algorithmic level. The rescaling is necessary because when dealing with imbalanced classification problems it is common that the confidence of the minority class rules is low. This fact implies that at classification time most of the examples are classified in the majority class leading to a lack of accuracy in the minority class, which might be the class of interest in financial applications. Additionally, we provide the IV-FRM with a technique that allows one to provide a classification for those examples not matching any rule in the generated rule base. Finally, we propose to use an

evolutionary algorithm to tune the values of the system' parameters in order to increase its performance as much as possible. Specifically, we tune the parameters defining the IVFSs, namely the lateral position using the parameter α as shown in Figure 1 and the amplitude of the support of the upper bound using the parameter W as shown in Figure 2, as well as the parameters a and b used to construct the IV-REFs (see Equation (9)) applied in the computation of the matching degree of the examples with the antecedent of the rules.

In the remainder of this section we describe in detail the three novel techniques, namely the rule weight rescaling method, the mechanism to handle uncovered examples and the tuning proposal.

3.1 Rule Weight Rescaling Method

This section is aimed at describing the method used to rescale the rule weight of the rules in order to handle the imbalanced data-sets faced in real world financial applications. The need for this procedure is easily shown in the following example in Table 2 which shows the rule base generated by the IVTURS$_{FARC-HD}$ method for the fourth partition of the Yeast-1-4-5-8_vs_7 dataset (obtained from the KEEL repository [52]). Five linguistic labels have been used for each variable: very low, low, medium, high and very high.

From Table 2 it is observed that there is only one rule belonging to the minority class and its rule weight is small whereas the rule weights of the rules belonging to the majority class are large. If the rule belonging to the minority class had a perfect matching degree ([1, 1] = [1, 1]* [1, 1]* [1, 1]), its association degree would be its rule weight ([0.06, 0.44] = [1, 1]*[0.06, 0.44]). In order to compute the classification soundness degree for each class we aggregate the association degrees (for instance with the maximum) of the rules having that class in their consequents. Following the previous example, the classification soundness degree for the minority class would be [0.06, 0.44]. Regarding the majority class, we have to apply the maximum of the association degrees of rules 2-5. If any of the association degrees of these four rules were greater than [0.06, 0.44] the example would be classified in the majority class, since the predicted class is the one having the greatest classification soundness degree. Therefore, the conditions necessary to classify an example in the minority class are difficult to be fulfilled, since it is

difficult to have a perfect matching degree and, even in this situation, it is easy that any of the rules of the majority class has a matching degree such that when multiplied by its rule weight the association degree is greater than [0.06, 0.44].

TABLE 2 RULES GENERATED FOR THE YEAST-1-4-5-8_VS_7 DATASET.

Rule Number	Rule	Rule Weight
1	IF Gvh IS High AND Pox IS Very Low AND Vac IS Very High Then Class IS Minority	[0.06, 0.44]
2	IF Nuc IS Medium Then Class IS Majority	[0.97, 0.99]
3	IF Pox IS Very High Then Class IS Majority	[1.0, 1.0]
4	IF Mcg IS Very High Then Class IS Majority	[0.95, 0.99]
5	IF Mcg IS Very Low Then Class IS Majority	[0.97, 0.98]

In order to deal with the previous problem, we propose a method to rescale the rule weights once the rule base has been generated. The procedure is composed of the following four steps:

- To compute the cumulative matching degrees, for each rule the matching degrees of examples belonging to the rule class are summed.

$$\left[CMD_j, \overline{CMD_j} \right] =$$
$$\left[\sum_{x_{p \in Class\ k}} \underline{\mu_{A_j}}(x_p), \sum_{x_{p \in Class\ k}} \overline{\mu_{A_j}}(x_p) \right], j = 1, \dots, L \tag{19}$$

- To compute the scaling factor for each class, the cumulative matching degrees of rules having the same class in the consequent are summed.

$$\left[\underline{SF_k}, \overline{SF_k} \right]$$
$$= \left[\sum_{j=1, Class(R_j)=k}^{L} \underline{CMD_j}, \sum_{j=1, Class(R_j)=k}^{L} \overline{CMD_j} \right],$$
$$k = 1, \dots, M \tag{20}$$

- To compute the scaled cumulative matching degree for each rule, the cumulative matching degree of each rule is divided (using the division of interval mathematics as explained in Equation (7)) by the scaling factor of the corresponding class as follows:

$$\underline{SCMD_j} = \wedge \left(\frac{CMD_j}{SF_{Class(R_j)}}, \frac{\overline{CMD_j}}{\overline{SF_{Class(R_j)}}} \right) j = 1, \dots, L$$

$$\overline{SCMD_j} = \vee\left(\frac{\underline{CMD_j}}{\underline{SF_{Class(R_j)}}}, \frac{\overline{CMD_j}}{\overline{SF_{Class(R_j)}}}\right) j = 1, \dots, L \tag{21}$$

$$\tag{22}$$

- To compute the scaled support and confidence of each rule, Equations (16) and (17) are applied using the results computed in the previous step (following the division of interval mathematics as explained in Equation (7)) to result in the scaled support and confidence values:

$$\left[\underline{Support_{Scaled}(A \rightarrow C_j)}, \overline{Support_{Scaled}(A \rightarrow C_j)}\right]$$
$$= \left[\frac{\underline{SCMD_j}}{N}, \frac{\overline{SCMD_j}}{N}\right] \tag{23}$$

$$\underline{Confidence_{Scaled}(A \rightarrow C_j)}$$
$$= \bigwedge\left(\frac{\underline{SCMD_j}}{\sum_{k=1}^{M}\underline{SCMD_k}}, \frac{\overline{SCMD_j}}{\sum_{k=1}^{M}\overline{SCMD_k}}\right) \tag{24}$$

$$\overline{Confidence_{Scaled}(A \rightarrow C_j)}$$
$$= \vee\left(\frac{\underline{SCMD_j}}{\sum_{k=1}^{M}\underline{SCMD_k}}, \frac{\overline{SCMD_j}}{\sum_{k=1}^{M}\overline{SCMD_k}}\right) \tag{25}$$

- To compute the rule weight, the scaled support and confidence of each rule are multiplied (using the multiplication of interval mathematics as explained in Equation (6)) and assigned as the rule weight.

$$\underline{RW_j} = \underline{Support_{Scaled}(A \rightarrow C_j)}$$
$$* \underline{Confidence_{Scaled}(A \rightarrow C_j)}, j$$
$$= 1, \dots, L \tag{26}$$
$$\overline{RW_j} = \overline{Support_{Scaled}(A \rightarrow C_j)}$$
$$* \overline{Confidence_{Scaled}(A \rightarrow C_j)}, j$$
$$= 1, \dots, L \tag{27}$$

3.2 Handling Inputs Not Matching Rules in the Rule Base

The rule learning method used by FARC-HD is able to create rules whose maximum number of antecedents can be programmatically limited using the *maximum tree depth*. Therefore, it creates a compact rule base composed of short rules, which helps increasing both the interpretability and the readability of the model and it also implies a reduction of the computational time needed to classify an example.

According to the fuzzy rule learning algorithm used by the FARC-HD method, if we set the *maximum tree depth* to the number of variables of the problem, a rule base composed of fuzzy rules whose antecedent length is equal to the number of variables could be generated. In this situation, the created fuzzy rules would cover very narrow areas of the solution space, which could provoke an increase on the system's accuracy at the expense of a reduction of the system's interpretability since both the rule length and the number of generated fuzzy rules would be greater. Figure 4a depicts a specific fuzzy rule (IF x_1 is Low and x_2 is Low) covering a narrow area whereas a generic fuzzy rule (IF x_1 is Low) is shown in Figure 4b. Although FARC-HD could generate both types of rules, it usually creates generic fuzzy rules like the later one.

A big problem encountered when producing specific rules is that some regions of the solution space could not be covered by any fuzzy rule. This situation happens in those cells shown in Figure 4a where there are no examples (like the cell {Low, Medium}), since specific rules (like Rule: {Low, Low}) are generated for cells having examples. This situation provokes the need of providing the system with a mechanism to classify examples that are not covered by any fuzzy rule in the rule base. There are two main approaches to handle this situation: 1) to reject the input without providing a prediction for the example and hence the example is not considered to compute the TP_{rate} and TN_{rate}; 2) to build a default rule that always classifies uncovered examples in the majority class.

The first approach is unacceptable solution for the financial domain where the prediction system should always be able to provide a prediction. The second approach avoids the problem of not providing a prediction but if the prediction capability over the uncovered examples is measured applying the geometric mean the achieved result will be always 0. This is due to the fact that the default rule correctly classifies all the examples of the majority class ($TN_{rate}=1$) whereas it misclassifies all the examples of the minority class ($TP_{rate} = 0$).
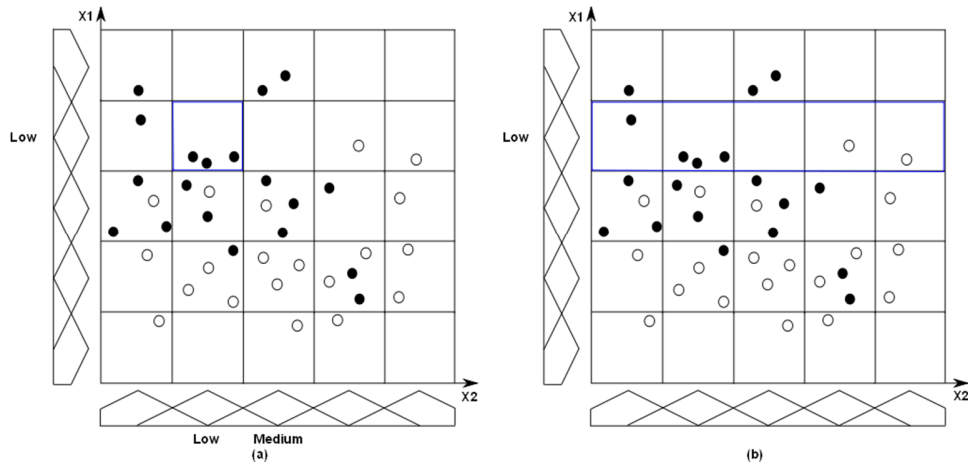
Figure 4 Solution space covered by fuzzy rules. a) A fuzzy rule using all the antecedents. b) A fuzzy rule not using all the antecedents.

In order to solve this problem we proposed a technique aimed at using a weighted combination of the most suitable rules in the rule base to classify the uncovered example x_p. This technique is composed of the following four steps:

- **Generation of a Set of Fuzzy Rules for the Uncovered example (SFRU)**. To do so, the membership functions providing a positive membership degree are found for each variable. Then, all the possible fuzzy rules are generated by performing all the combinations of the previously matched membership functions. These rules do not have in the consequent part either the class or the rule weight. For example, in a problem with two input variables $x1$ and $x2$ if the uncovered example x_p matches the fuzzy sets {Low, Medium} for the variable $x1$ and {Medium, High} for the variable $x2$, the generated fuzzy rules will be $SFRU(x_p) =$ {R1:{Low, Medium}, R2:{Low, High}, R3:{Medium, Medium}, R4:{Medium, High}}.

- **Obtaining the most similar fuzzy rules in the rule base.** In this step, for each fuzzy rule in $SFRU(x_p)$, the most similar fuzzy rule in the rule base is obtained. With this aim, in first place the fuzzy rules in the rule base and the ones in $SFRU(x_p)$ are decoded using an integer coding scheme. For example, the set of linguistic labels {low, medium, high} could be encoded as {1, 2,

3}. Then, Equation (28) is applied in order to measure the similarity between fuzzy rules.

$$Sim(R'_j, R_j) = \prod_{k=1}^{n} \left(1 - \frac{V(R'_{jk}) - V(R_{jk})}{NL_k} \right) \quad (28)$$

where R'_j is the j'-th rule in $SFRU(x_p)$ with $j' = \{1, ..., |SFRU(x_p)|\}$ and $|SFRU(x_p)|$ is the number of rules in $SFRU(x_p)$, R_j is the j-th rule in the rule base with $j = \{1, ..., L\}$, n is the number of variables of the problem, NL_k is the number of linguistic labels of the k-th variable and $V(\cdot)$ is the integer codification of the k-th linguistic label of the rule being analyzed. At this point each rule in $SFRU(x_p)$ has a similarity with every rule in the rule base. Finally, for each rule in $SFRU(x_p)$ the rule in the rule base having the greatest similarity is selected. Therefore, after this step, as many original rules in the rule base as rules in $SFRU(x_p)$ are taken, which are used (along with the rules in $SFRU(x_p)$) to make the final prediction as shown in the next step of the method.

- **Computing the vote strength for each class**. In this stage, both the previously selected original rules and the rules in $SFRU(x_p)$ are used to compute the vote strength, related to the lower and the upper bounds of the IVFSs, for each class using Equations (29) and (30) respectively. More specifically, we use the rule weights of the original rules and the antecedents of the rules in

J. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (+34 948 166 408; fax: +34 948 168 924; e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).

D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Wivenhoe Park, Colchester, UK.

F. Herrera is with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain and with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, 21589, Jeddah, Saudi Arabia (e-mail: Herrera@decsai.ugr.es).

$SFRU(x_p)$.

$$\underline{Vote}_h(x_p) = \frac{\sum_{j'=1,Class(R_{j'})=h}^{|SFRU(x_p)|} \underline{\mu_{A_{j'}}}(x_p)*\underline{RW}_{j'}}{\bigvee_{Class(j')=h}^{|SFRU(x_p)|} \underline{\mu_{A_{j'}}}(x_p)*\underline{RW}_{j'}} \quad (29)$$

$$\overline{Vote}_h(x_p) = \frac{\sum_{j'=1,Class(R_{j'})=h}^{|SFRU(x_p)|} \overline{\mu_{A_{j'}}}(x_p)*\overline{RW}_{j'}}{\bigvee_{Class(R_{j'})=h}^{|SFRU(x_p)|} \overline{\mu_{A_{j'}}}(x_p)*\overline{RW}_{j'}} \quad (30)$$

where $\underline{\mu_{A_{j'}}}(x_p)$ and $\overline{\mu_{j'}}(x_p)$ are the lower and upper matching degrees of the example x_p with the j'-th rule in $SFRU(x_p)$ whereas $\underline{RW}_{j'}$ and $\overline{RW}_{j'}$ are the lower and upper rule weights of the most similar rule in the rule base to the j'-th rule in $SFRU(x_p)$.

- **Final prediction of the class**. The uncovered example x_p will be classified in the class having largest vote strength according to Equation (31).

$$F(Y_1, ..., Y_M) = arg \max_{h=1,...,M} \left(\frac{\underline{Vote}_h(x_p) + \overline{Vote}_h(x_p)}{2} \right) \quad (31)$$

3.3 Tuning the System Membership Functions and the Associated Parameters

The last stage of the methodology consists of optimizing the values of the system's parameters. In this paper we propose to tune both the values determining both the shape and position of the IVFSs and the values of the parameters used to generate the IV-REF associated with each variable of the problem. In this manner, we combine the good features of two common tuning approaches [53], [54], namely the genetic tuning of the knowledge base parameters and the genetic adaptive inference system. Additionally, we perform simultaneously a rule selection process to decrease the system's complexity.

We use the CHC evolutionary model [55], which is short for Cross generational elitist selection, Heterogeneous recombination and Cataclysmic mutation, to carry out the optimization process, since it is the same method used in the two state-of-the-art fuzzy classifiers used in this paper and it is a good choice in problems having a complex search space. The specific components of our new proposal are:

- *Coding scheme*: The chromosomes use a double coding scheme. On the one hand, real codification is considered for the tuning

proposals and binary codification is used for the rule selection process. Equation (32) shows the structure of the whole chromosome.

$$C_{Total} = \{C_L, C_W, C_E, C_R\} \quad (32)$$

It can be observed that the chromosome is composed of four parts: the three first parts perform the tuning of the membership functions and the associated parameters and the last part carries out a rule selection process. These parts are described in detail below:

1. Tuning of the lateral position of the linguistic labels (depicted in Figure 2).

$$C_L = \{\alpha_{11}, ..., \alpha_{1l_1}, ..., \alpha_{n1}, ..., \alpha_{nl_n}\}, \quad (33)$$

where $\alpha_{is} \in [-0.5, 0.5]$, with $i = 1, ..., n$, $s = 1, ..., l_i$ and l_i represents the number of linguistic labels used in the i-th variable.

2. Amplitude of the support of the upper bound of the IVFSs.

$$C_W = \{W_{11}, ..., W_{1l_1}, ..., W_{n1}, ..., W_{nl_n}\}, \quad (34)$$

where $W_{is} \in [0, 1]$, with $i = 1, ..., n$, $s = 1, ..., l_i$ and l_i represents the number of linguistic labels used in the i-th variable.

3. Tuning of the equivalence: genes take values in the range [0.01, 1.99]

$$C_E = \{a_1, b_1, ..., a_n, b_n\}, \quad (35)$$

where $a_i, b_i \in [0.01, 1.99]$, with $i = 1, ..., n$.

4. Rule selection: this part is composed of as many genes as the number of rules in the rule base. A binary codification is used; therefore the possible values for the genes are 0 or 1, where the values 1 and 0 mean that the associated rule is used or not in the inference respectively.

$$C_R = \{R_1, ..., R_L\} \quad (36)$$

where $R_j \in \{0, 1\}$, with $j = 1, ..., L$.

- *Initialization of the population*: we initialize a chromosome representing the initial system, which is the chromosome that encodes the initial set-up of the IV-FRBCS. To do so, we set all the genes performing the lateral tuning to 0.0 (so that the membership functions do not have any lateral displacement), all the genes used to modify the amplitude of the support of the IVFSs are set to 0.5 (in order to make the amplitude of the upper bound 50% greater than that of the lower bound), the ones to carry out the tuning of the equivalence are set to 1.0 (with this setting the identity function is computed) and the genes used to make the rule selection process are set to 1 (to consider

all the rules in the rule base). The remainder individuals are randomly initialized within the ranges demanded in each part.

- *Chromosome evaluation*: we take the average mean between the accuracy achieved in both classes, which is the area under the ROC curve [56] considering a single point

$$Fitness = \frac{TP_{rate} + TN_{rate}}{2} \qquad (37)$$

We must point out that this fitness function provides a similar behavior to that of the geometric mean, since both of them take into account the accuracy obtained in each class of the problem. In this manner we approximate the results of the geometric mean using a less computational demanding metric.

- *Crossover operator*: for the part of the chromosome using a real codification we apply the Parent Centrix BLX operator [57] whereas the half uniform crossover [58] is used for the part using the binary codification.
- *Restarting approach*: the population is randomly initialized and the best individual found so far is included in the population as in the elitist scheme. In this manner, we get away from local optima.

4. EXPERIMENTS AND RESULTS

In this section, we present the experiments and results to validate our proposed system, which is denoted $IVTURS_{FARC-HD}^{RRW\_I}$, for financial applications. The experimental framework used to show the quality of the new method is described in Section 4.1. The study has a double aim, on the one hand, the suitability of the novelties introduced in $IVTURS_{FARC-HD}^{RRW\_I}$ needs to be justified (Section 4.2) and, on the other hand, the benefits of our new method against the use of SMOTE, which is a widely used pre-processing technique, and the cost sensitive C4.5 decision tree are analyzed (Section 4.3). In both scenarios, we present the evaluations carried over eleven different financial applications.

4.1 Experimental Framework

In this section we present the framework we have used to test the quality of our new approach. Specifically, we first describe the financial datasets used in the study. Next, we introduce the notation and configuration of the different classifiers and finally, the statistical tests used to validate our results are presented.

4.1.1 Financial datasets description

We have used eleven real-world datasets from various financial domains. The features of these datasets are summarized in Table 3, where we can see the names of the datasets, their total number of attributes (#Attr.), classes (#Cl.) and examples (data instances) (#Ex.). We also show the class distribution of the examples according to the classes (% Cl. 0 (% of Class 0) and % Cl. 1(% Class 1)) and the Imbalanced Ratio (IR) [59] of each dataset, which is defined as the ratio of the number of instances of the minority class and the majority class. These datasets and the tables, in which the obtained results are shown, are sorted incrementally according to the IR.

TABLE 3 FEATURES OF THE FINANCIAL DATA-SETS USED IN THE EXPERIMENTAL STUDY.

Dataset	#Attr.	#Cl.	Examples			IR
			#Ex.	% Cl. 0	% Cl. 1	
BI	12	2	1300	46.54	53.46	1.15
BC	52	2	1650	59.64	40.36	1.48
WSI	12	2	1298	39.29	60.71	1.55
FESI	12	2	1300	62.15	37.85	1.64
DT	12	2	1300	62.62	37.38	1.67
AL	12	2	1300	62.85	37.15	1.69
SL	21	2	1894	68.22	31.78	2.15
Arb	7	2	1641	75.56	24.44	3.09
FD	20	2	17795	78.94	21.06	3.75
Len	23	2	24476	81.61	18.39	4.44
LA	10	2	123115	1.46	98.54	67.66

The description of each financial dataset is as follows:

- **Bank Investment (BI) Data set**: is used to predict whether to invest or not by buying or not buying a bank share in stock market.
- **Bank Credit (BC) Data Set**: is a bank credit card approval application system which is in use by a real-world bank to identify good and bad customers where good customers are non-defaulting customers and bad customers are defaulting customers.
- **Western Stock Index (WSI) Data Set**: is used for predicting a major Western stock market composite index of whether the stock market index will increase/remain the same or decrease.

- **Far Eastern Stock Index (FESI) Data Set**: is used for predicting a far eastern stock market composite index of whether the stock market index will increase or remain the same or decrease.
- **Digital TV (DT) Data Set**: is used for predicting whether to invest or no by buying or not buying a digital TV network share in stock market.
- **Airline (AL) Data Set**: is applied for predicting whether to invest or no by buying or not buying an airline share in stock market.
- **Small Loans (SL) Data Set**: is used for the evaluation of customers (good or bad customer) for personal small loans applications where there is no knowledge of the customer full credit history.
- **Arbitrage (Arb) Data Set**: is used for spotting arbitrage opportunities in the London International Financial Futures Exchange (LIFFE) market [10-12]. The data reported in this paper was developed in [12] to identify arbitrage situations by analyzing option and futures prices in the LIFFE market.
- **Fraud Detection (FD) Data Set**: is used for fraud detection in personal loan application for small loan amounts.
- **Lending (Len) Data Set**: is used for evaluation of customers (good or bad customer) for personal small loans applications where there is knowledge of the customer full credit history.
- **Loan Authorisation (LA) Data Set**: is related to the prediction of good (profitable) or bad (non-profitable) customers for loan authorization.

In order to carry out the experimental study we have split the data using a random stratified scheme, where 70% of the examples are used to train the system and the remaining 30% of the examples are used to test the generated model.

4.1.2 The Used Configurations and Notations

In this paper, we compare our proposed compact evolutionary IV-FRBCS that we call $IVTURS_{FARC-HD}^{RRW\_I}$, which is highlighted in grey in Table 4, with the following algorithms:

- $IVTURS_{FARC-HD}$: it is a version of our new proposal that do not use the rule weight rescaling method. It will be used to determine the benefits of using the rule weight rescaling method.
- FARC-HD method [44] (FARCHD): it is a state-of-the-art type-1 fuzzy classifier. It will be used to show the suitability of the use of IVFSs.
- IVTURS_FS: it is the type-1 fuzzy counterpart of $IVTURS_{FARC-HD}$. It also will be considered to analyze the goodness of the use of IVFSs.
- C4.5 [25]: it is the classical C4.5 decision tree. It is included in the study since it is a widely used interpretable method when dealing with classification problems.
- C4.5_CS [26]: it is the C4.5 decision tree modified so that it uses a cost sensitive method to deal with the imbalanced problem at algorithmic level.
- FURIA [27]: it is a state-of-the art type-1 fuzzy approximative classifier.
- SMOTE [3]: it is one of the most used pre-processing techniques. It will be used to check whether the use of $IVTURS_{FARC-HD}^{RRW\_I}$ is competitive versus a state-of-the-art pre-processing technique.

We have to point out that all the classifiers used in the comparison provide an interpretable and transparent model in order to ease the understating of the system by financial analysts. The exception is FURIA, since its linguistic terms are not defined in the same way in the different rules. The notations and descriptions of the features of the classifiers used in the study are shown in Table 4, where the first column shows the name given for each approach, the second one describes the kind of set used to model the linguistic labels, the third column shows the method used to compute the rule weight and the last column specifies the parameters which are tuned in the optimization process.

TABLE 4 NOTATIONS AND DESCRIPTIONS OF THE FIVE APPROACHES.

Name	Linguistic Labels	Rule Weight	Tuned parameters
IVTURS$_{\text{FARC-HD}}$	IVFSs	Confidence	α, a, b, W
FARCHD	Type-1 FSs	Confidence	α
IVTURS_FS	Type-1 FSs	Confidence	α, a, b
C4.5	Crisp sets	Relative frequency	-
FURIA	Type-1 FSs	m-estimate ($m = 2$)	Left and right points of MFs
IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$	IVFSs	Rule Weight Rescaling Method	α, a, b, W

In order to conduct a fair comparison, we have considered the same configurations for the methods based on the FARC-HD algorithm (they can use either type-1 FSs or IVFSs): we have used five linguistic labels (using triangular shaped membership functions) per variable, the interval product (or product when using type-1 FSs) to model the conjunction operator (t-norm) and the FRM of the wining rule (the maximum as aggregation function). The thresholds used in the a-priori algorithm are introduced in Table 5. For the genetic tuning we have considered the following values for their parameters:

- Population size: 50 individuals.
- Number of evaluations: 20,000.
- Bits per gene for the Gray codification (for incest prevention): 30.

Regarding the C4.5 decision tree we have used 0.25 and 2 as confidence level and minimum number of examples per leaf, respectively. Finally, we have set the configuration of FURIA as recommended by the authors, that is, three folds and two optimizations.

TABLE 5 CONFIGURATION OF THE FARC-HD METHOD.

Parameter	Value
Minimum support	0.01
Minimum confidence	0.9
Maximum tree depth	3
k_t: number of covered times	2

4.1.3 Statistical tests

We will use hypothesis validation techniques in order to give statistical support to the analysis of the presented results [60], [61]. We use non-parametric tests because the initial conditions that guarantee the reliability of the parametric tests cannot be fulfilled, which implies that the statistical analysis loses credibility with these parametric tests [60].

Specifically, we use the Friedman aligned ranks test [62] to detect statistical differences among a group of results and the Holm post-hoc test [63] to find the algorithms that reject the equality hypothesis

with respect to a selected control method.

The post-hoc procedure allows us to know whether a hypothesis of comparison could be rejected at a specified level of significance α. Furthermore, we compute the Adjusted P-Value (APV) in order to take into account the fact that multiple tests are conducted. In this manner, we can directly compare the APV with respect to the level of significance α in order to be able to reject the null hypothesis.

Furthermore, we consider the method of aligned ranks of the algorithms in order to show graphically how good a method is with respect to the remainder ones. The first step to compute this ranking is to obtain the average performance of the algorithms in each data set. Next, we compute the subtractions between the accuracy of each algorithm minus the average value for each data-set. Then, we rank all these differences in descending order and finally, we then average the rankings obtained by each algorithm. In this manner, the algorithm which achieves the lowest average ranking is the best one.

4.2 Studying the effectiveness of the novelties introduced in IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$

In this part of the study, the analysis is conducted in order to justify empirically the novelties introduced in our new proposal. To do so, we first compare the behavior of IVTURS$_{\text{FARC-HD}}^{\text{RRW\_I}}$ versus IVTURS$_{\text{FARC-HD}}$ and the two classifiers using type-1 FSs (FARC-HD algorithm and IVTURS_FS). In this way, we show the necessity of applying two of the three novelties introduced in the new proposal, namely, the use of IVFSs and the rescaling of the rule weight method.

Table 6 contains the results obtained when applying these four classifiers both in training and in testing. These results are measured using the geometric mean (Eq. (13)) and they are grouped by two columns to show the results obtained in training

TABLE 6. RESULTS OBTAINED IN TRAINING (TR.)
AND TESTING (TST.) MEASURED USING THE GEOMETRIC MEAN.

Dataset	$IVTURS_{FARC-HD}^{RRW\_I}$		FARCHD		$IVTURS_{FARC-HD}$		IVTURS_FS		IR
	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	
BI	0.80	0.43	0.84	**0.51**	0.81	0.47	0.85	0.49	1.15
BC	0.70	**0.61**	0.83	0.54	0.77	0.59	0.80	0.56	1.48
WSI	0.80	**0.59**	0.84	0.55	0.83	0.58	0.83	0.54	1.55
FESI	0.79	**0.57**	0.81	0.45	0.76	0.50	0.78	0.45	1.64
DT	0.79	0.53	0.77	0.54	0.74	**0.59**	0.77	0.47	1.67
AL	0.79	**0.55**	0.75	0.53	0.73	0.49	0.75	0.47	1.69
SL	0.65	**0.59**	0.52	0.44	0.46	0.39	0.48	0.41	2.15
Arb	0.94	**0.94**	0.78	0.78	0.93	0.93	0.92	0.92	3.09
FD	0.63	**0.61**	0.24	0.19	0.19	0.15	0.21	0.16	3.75
Len	0.53	**0.50**	0.28	0.20	0.26	0.20	0.28	0.21	4.44
LA	0.73	**0.72**	0.29	0.31	0.73	0.72	0.31	0.30	67.66
Mean	0.74	**0.60**	0.63	0.46	0.66	0.51	0.63	0.45	

(Tr.), in testing (Tst.). The best (highest) testing result is highlighted in **bold-face**.

From the results shown in Table 6 it is observed that our proposed approach, highlighted in grey, provides clearly the best mean result (of geometric mean) in testing (beating the competing techniques by a big margin followed by $IVTURS_{FARC-HD}$) and it also reaches the best result in nine out of the eleven datasets. From these results we can stress two facts: on the one hand, the use of IVFSs ($IVTURS_{FARC-HD}^{RRW\_I}$ and $IVTURS_{FARC-HD}$) allows the results of the approaches using type-1 FSs (FARC-HD and IVTURS_FS) to be improved and, on the other hand, it is observed that the rule weight rescaling method has a beneficial effect, since the results of $IVTURS_{FARC-HD}^{RRW\_I}$ are better than those obtained when applying $IVTURS_{FARC-HD}$. The results of $IVTURS_{FARC-HD}^{RRW\_I}$ are especially better than the ones of the remainder approaches when the IR increases. Consequently, we can conclude that the new techniques introduced in our method are appropriate to deal with imbalanced data sets present in the vast majority of financial applications.

In order to support the superiority of $IVTURS_{FARC-HD}^{RRW\_I}$, we have applied the Friedman aligned ranks test. The obtained ranks are shown in the third column of Table 7 and the p-value is 0.036, which implies the existence of statistical differences among these four methods. This fact allows us to perform the Holm post-hoc test, whose obtained APVs are included in the last column of Table 7, to check whether $IVTURS_{FARC-HD}^{RRW\_I}$, which is used as control method because it is the best ranked one, statistically outperforms the remainder approaches. Looking at the statistical results shown in Table 7, we can conclude that our new approach is statistically better than FARC-HD, $IVTURS_{FARC-HD}$ and IVTURS_FS and consequently, the use of both IVFSs and the rescaling rule weight method allows handling the imbalanced financial data sets to give a superior performance.

TABLE 7. HOLM'S TEST TO COMPARE $IVTURS_{FARC-HD}^{RRW\_I}$ VERSUS FARC-HD, $IVTURS_{FARC-HD}$ AND IVTURS_FS. $IVTURS_{FARC-HD}^{RRW\_I}$ IS USED AS CONTROL METHOD.

No.	Algorithm	Ranking	APV
1	IVTURS_FS	29.82	0.0011
2	FARCHD	28.27	0.0020
3	$IVTURS_{FARC-HD}$	21.64	0.0380
4	$IVTURS_{FARC-HD}^{RRW\_I}$	10.27	-

J. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (+34 948 166 408; fax: +34 948 168 924; e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).

D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Wivenhoe Park, Colchester, UK.

F. Herrera is with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain and with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, 21589, Jeddah, Saudi Arabia (e-mail: Herrera@decsai.ugr.es).

Finally, we test the appropriateness of the proposed technique to deal with uncovered patterns. For the sake of generating specific fuzzy rules, we have run the fuzzy rule learning method using as *maximum tree depth* the number of attributes of the problem. For the FD dataset (we report only on this dataset as the similarity results are similar for the other datasets) we have selected the 10, 20, 30 and 40 fuzzy rules having a largest rule weight from each class, which implies obtaining rule bases composed of 20, 40, 60 and 80 rules respectively. The obtained results for the FD financial problem are presented in Table 8, where each row shows the number of rules in the rule base, the number of uncovered examples (using the rule base composed of as many rules as indicated in the first column) along with the accuracy achieved over the uncovered examples for each class (TP_{rate} and TN_{rate}) and the result of the geometric mean (GM). It can be observed that the use of the similarity technique is beneficial for the system since when using the default rule the result of the geometric mean is always 0. Furthermore, it is noticed that the larger the number of fuzzy rules in the rule base the better the result of this technique is. This is due to the fact that when using a larger number of fuzzy rules the solution space is better covered leading to obtaining more suitable similar rules. This shows the power of proposed technique where we produce predictions for uncovered patterns to give a reasonable geometric mean.

TABLE 8 RESULTS OBTAINED WITH THE SIMILARITY TECHNIQUE ON THE FD DATASET USING THE $IVTURS_{FARC-HD}^{RRW\_I}$ METHOD.

Number of rules	Number of uncovered examples	TP_{rate}	TN_{rate}	GM
20	1595	0.530	0.550	0.5399
40	933	0.570	0.590	0.5799
60	813	0.595	0.605	0.6000
80	315	0.602	0.638	0.6197

4.3 Analyzing the suitability of $IVTURS_{FARC-HD}^{RRW\_I}$ versus state-of-the-art techniques

The second part of the experimental study is aimed at analyzing the behavior of our new proposal when it is compared versus state-of-the-art techniques that deal with imbalanced data. More specifically, we study the behavior of our approach versus the cost sensitive C4.5 decision tree [26] and

several classifiers that receive pre-processed data by means of SMOTE, which is one of the most widely used pre-processing techniques. To do so, this study is divided in three parts because we consider three types of algorithms, which are based on fuzzy association rules, the C4.5 decision tree and the FURIA algorithm, respectively.

Table 9 shows the results obtained by our new method and the three classifiers based on the usage of fuzzy association rules using SMOTE both in training (Tr.) and in testing (Tst.), which are measured using the geometric mean. The best testing result for each dataset is highlighted in **bold-face**.

From results in Table 9, it is shown that the best mean testing result of geometric mean is provided by our new proposal. More specifically, we can observe that the results of the other method using IVFSs (IVTURS_{FARC-HD}+SMOTE) are improved by 1.4% and the improvement versus the two approaches using type-1 FSs (FARCHD+SMOTE and IVTURS_FS+SMOTE) is around 3%.

In order to find whether there are statistical differences among these methods we have applied the aligned Friedman ranks test. The provided p-value is 0.0291, which confirms the existence of statistical differences where $IVTURS_{FARC-HD}^{RRW\_I}$ was the best method since it is the best ranked one (see third column of Table 10). Next, we have carried out the Holm's post-hoc test to determine if $IVTURS_{FARC-HD}^{RRW\_I}$ is statistically better than the remainder methods. The results obtained in the statistical study are shown in Table 10, where it can be observed that our new proposal, which is used as control method because it obtains the best ranking, outperforms both the FARC-HD method and the fuzzy counterpart of the IVTURS_{FARC-HD} algorithms using SMOTE (FARCHD+SMOTE and IVTURS_FS+SMOTE). Finally, it is noteworthy that there are not statistical differences with respect to IVTURS_{FARC-HD}+SMOTE.

In the second part of the study we compare our proposed method against the approaches using decision trees, namely C4.5, C4.5 with SMOTE and the cost sensitive C4.5 (C45_CS), which widely used when dealing with imbalanced datasets. Table 11 shows the results provided by these four classifiers both in Training (Tr.) and testing (Tst.). It is noteworthy the notable improvement achieved by $IVTURS_{FARC-HD}^{RRW\_I}$ over the mean of the testing data,

TABLE 9 RESULTS OBTAINED BY THE APPROACHES BASED ON FUZZY ASSOCIATION RULES CLASSIFIERS IN TRAINING (TR.) AND TESTING (TST.).

SAMPLING	NO		SMOTE						IR
Dataset	IVTURS$_{FARC-HD}^{RRW\_I}$		FARCHD		IVTURS_FS		IVTURS$_{FARC-HD}$		
	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	
BI	0.80	0.43	0.83	0.51	0.84	**0.53**	0.81	0.47	1.15
BC	0.70	**0.61**	0.81	0.56	0.78	0.55	0.76	0.55	1.48
WSI	0.80	**0.59**	0.84	0.49	0.83	0.52	0.80	0.55	1.55
FESI	0.79	**0.57**	0.82	0.54	0.80	**0.57**	0.78	0.56	1.64
DT	0.79	0.53	0.83	0.52	0.80	0.44	0.78	**0.59**	1.67
AL	0.79	0.55	0.79	0.52	0.79	**0.56**	0.78	**0.56**	1.69
SL	0.65	**0.59**	0.67	0.53	0.66	0.56	0.64	0.55	2.15
Arb	0.94	0.94	0.90	0.90	0.87	0.88	0.91	**0.95**	3.09
FD	0.63	**0.61**	0.69	0.58	0.69	0.56	0.68	0.57	3.75
Len	0.53	**0.50**	0.78	0.44	0.79	0.38	0.80	0.41	4.44
LA	0.73	0.72	0.75	**0.73**	0.75	**0.73**	0.75	**0.73**	67.66
Mean	0.74	**0.60**	0.79	0.57	0.78	0.57	0.77	0.59	

TABLE 10 HOLM'S TEST TO COMPARE IVTURS$_{FARC-HD}^{RRW\_I}$ VERSUS THE APPROACHES BASED ON FUZZY ASSOCIATION RULES CLASSIFIERS. IVTURS$_{FARC-HD}^{RRW\_I}$ IS USED AS CONTROL METHOD.

No.	Algorithm	Ranking	APV
1	FARC+SMOTE	27.82	0.0366
2	IVTURS_FS+SMOTE	27.09	0.0366
3	IVTURS$_{FARC-HD}$+SMOTE	21	0.2071
4	IVTURS$_{FARC-HD}^{RRW\_I}$	14.09	-

since it improves in 4%, 7% and 12% the results obtained by C45_CS, C45+SMOTE and the original C4.5 decision tree, respectively.

In order to support the previous findings, we have followed the same statistical study carried out in the previous analysis. The obtained results are shown in Table 12, where it can be observed that IVTURS$_{FARC-HD}^{RRW\_I}$ is the best ranking method and it is statistically outperforming the original C4.5 decision tree with and without SMOTE. Regarding C45_CS, it can be noticed that there are not statistical differences. However, our proposal obtains a notable average improvement in performance based on the achievement of better results in seven out of the eleven financial problems, which manifest a superior behavior.

Finally, we have also tested our new method versus the state-of-the-art fuzzy classification known as FURIA with and without SMOTE. The performance of these methods is included in Table 13, where it can be observed that IVTURS$_{FARC-HD}^{RRW\_I}$ obtains the best mean result, which is based on the achievement of the best performance in half of the financial applications considered in this paper. The statistical results (see Table 14) confirm that our new approach outperforms the original FURIA but there are not statistical differences between IVTURS$_{FARC-HD}^{RRW\_I}$ and FURIA+SMOTE, although the average performance provided by our proposal is better than that of FURIA+SMOTE. Therefore, we can conclude that our new proposal provides competitive results versus an approximative fuzzy classifier that uses pre-processed data in the learning stage whereas our proposal uses the original data and consequently it keeps the original distribution of the classes.

J. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (+34 948 166 408; fax: +34 948 168 924; e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).

D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Wivenhoe Park, Colchester, UK.

F. Herrera is with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain and with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, 21589, Jeddah, Saudi Arabia (e-mail: Herrera@decsai.ugr.es).

TABLE 11 RESULTS OBTAINED BY $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$ AND THE APPROACHES BASED ON C4.5 IN TRAINING (TR.) AND TESTING (TST.).

Dataset	$\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$		C4.5		C4.5+SMOTE		C4.5_CS		IR
	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	
BI	0.80	0.43	0.96	**0.47**	0.97	0.41	0.93	0.45	1.15
BC	0.70	**0.61**	0.90	0.53	0.91	0.53	0.90	0.55	1.48
WSI	0.80	0.59	0.65	0.39	0.72	**0.62**	0.90	0.49	1.55
FESI	0.79	**0.57**	0.87	0.44	0.95	0.51	0.96	0.53	1.64
DT	0.79	0.53	0.96	**0.54**	0.93	0.48	0.95	0.52	1.67
AL	0.79	0.55	0.96	**0.57**	0.97	0.52	0.92	**0.57**	1.69
SL	0.65	**0.59**	0.59	0.40	0.78	0.52	0.73	0.49	2.15
Arb	0.94	0.94	0.98	0.96	0.98	**0.98**	0.99	0.97	3.09
FD	0.63	**0.61**	0.57	0.33	0.88	0.54	0.89	0.55	3.75
Len	0.53	0.50	0.68	0.41	0.91	0.40	0.96	**0.52**	4.44
LA	0.73	**0.72**	0.26	0.24	0.89	0.36	0.97	0.54	67.66
Mean	0.74	**0.60**	0.76	0.48	0.90	0.53	0.92	0.56	

TABLE 12 HOLM'S TEST TO COMPARE $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$ VERSUS THE APPROACHES BASED ON THE C4.5 DECISION TREE. $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$ IS USED AS CONTROL METHOD.

No.	Algorithm	Ranking	APV
1	C4.5	31.18	0.0030
2	C45+SMOTE	26.18	0.0352
3	C45_CS	19.45	0.2521
4	$\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$	13.18	-

Regarding the interpretability of the model, Tables 15 shows the number of rules along with their average number of antecedents (number in brackets) produced by the different techniques. This table is split into two groups, the proposals that do not use SMOTE (sampling: no) and the proposals that use it (sampling: SMOTE). For each problem, the less number of rules and its average number of antecedents are stressed in **bold-face**. It is remarkable to note the decrease of the number of rules produced by our new method when compared with the classifiers that receives pre-processed data by SMOTE. More specifically, our method produces an average of 71 rules whereas the approaches based on the usage of fuzzy association rules (FARCHD, IVTURS_FS and IVTURS_FARC-HD) using SMOTE generates an average number of rules ranged between 113 and 125, which means that at least 37% less number of rules than the competing methods are produced by $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$. Regarding the approaches based on the C4.5 decision tree, their average number of created rules is ranged between 160 and 735, that is, $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$ provides a more interpretable model. Finally, we have to point out that although FURIA produces a less number of rules, it is an approximative model, which means that the same linguistic label is defined in a different way in each rule. Due to this fact, the model provided by FURIA is far from being interpretable. In addition, our method produces shorter rule length as shown by the number in brackets next to the number of rules. Consequently, we can conclude that $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC}-\text{HD}}$ is producing a more interpretable model than the other competing methods that used SMOTE.

J. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (+34 948 166 408; fax: +34 948 168 924; e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).

D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Wivenhoe Park, Colchester, UK.

F. Herrera is with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain and with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, 21589, Jeddah, Saudi Arabia (e-mail: Herrera@decsai.ugr.es).

TABLE 13 RESULTS OBTAINED BY $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$ AND THE APPROACHES BASED ON FURIA IN TRAINING (TR.) AND TESTING (TST.).

Dataset	$\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$		FURIA		FURIA+SMOTE		IR
	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	
BI	0.80	0.43	0.95	0.47	0.95	**0.48**	1.15
BC	0.70	**0.61**	0.52	0.48	0.66	0.58	1.48
WSI	0.80	0.59	0.90	**0.60**	0.87	0.57	1.55
FESI	0.79	0.57	0.91	0.46	0.92	**0.60**	1.64
DT	0.79	0.53	0.91	**0.65**	0.90	0.55	1.67
AL	0.79	**0.55**	0.96	0.52	0.94	0.54	1.69
SL	0.65	**0.59**	0.43	0.50	0.56	0.56	2.15
Arb	0.94	0.94	0.98	**0.98**	0.98	0.97	3.09
FD	0.63	**0.61**	0.23	0.22	0.65	0.59	3.75
Len	0.53	**0.50**	0.30	0.28	0.85	0.24	4.44
LA	0.73	0.72	0.15	0.16	0.82	**0.73**	67.66
Mean	0.74	**0.60**	0.66	0.48	0.83	0.58	

TABLE 14 HOLM'S TEST TO COMPARE $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$ VERSUS THE APPROACHES BASED ON FURIA. $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$ IS USED AS CONTROL METHOD.

No.	Algorithm	Ranking	APV
1	FURIA	22.82	0.0345
2	FURIA+SMOTE	15.18	0.5967
3	$\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$	13	-

Table 16 shows the set of small number of rules and short rule lengths generated by our proposed method ($\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$) for the Arbitrage data set, where it can be seen that only 9 rules were generated and they have a short length which enable high degree of transparency and interpretability for the financial adviser while producing the best geometric mean when compared to the C4.5 decision tree, type-1 and IVFS based methods as shown in Table 9 with no need to use any preprocessing.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a compact evolutionary IV-FRBCS for the modeling and prediction of real-world financial problems, which is built on the basis of $\text{IVTURS}_{\text{FARC-HD}}$ and it is named $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$. The proposed system provides a small set of short linguistic fuzzy rules, which means obtaining a highly interpretable model in order to fulfill the current important requirements of transparency and interpretability needed by financial organizations. Furthermore, the proposed method deals directly with the imbalanced datasets problem (with no need for preprocessing or sampling), which is common in financial domains, by using a rule weight rescaling method. Finally, the system is enhanced with a mechanism to handle uncovered examples so it can always produce prediction with no need for default rules. These three properties make the new method highly suitable for real-world financial applications.

The quality of $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$ has been tested in eleven financial problems. From the obtained results we can stress the following lessons learned: 1) the use of both IVFSs and the rescaling rule weight method allows enhancing the results obtained with the counterparts of our proposal that do not apply them; 2) the similarity technique has shown to be an appropriate mechanism to handle examples which have not been covered by the fuzzy rules in the rule base; 3) $\text{IVTURS}_{\text{FARC}-\text{HD}}^{\text{RRW\_I}}$ achieves a better

J. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (+34 948 166 408; fax: +34 948 168 924; e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).
D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Wivenhoe Park, Colchester, UK.
F. Herrera is with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain and with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, 21589, Jeddah, Saudi Arabia (e-mail: Herrera@decsai.ugr.es).

TABLE 15 NUMBER OF RULES BESIDE THEIR AVERAGE NUMBER OF ANTECEDENTS PER RULE (IN BRACKETS).

SAMPLING	NO				SMOTE					IR
Dataset	$\text{IVTURS}_{FARC-HD}^{RRW\_I}$	C4.5	C45_CS	FURIA	FARCHD	IVTURS_FS	$\text{IVTURS}_{FARC-HD}$	C45	FURIA	
BI	**26 (2.57)**	63 (9.91)	57 (8.89)	36 (3.69)	29 (2.62)	29 (2.69)	30 (2.57)	77 (10.31)	36 (3.97)	1.15
BC	165 (2.97)	165 (11.99)	159 (12.31)	**7 (2)**	199 (2.96)	200 (2.96)	216 (2.96)	186 (13.52)	11 (2.36)	1.48
WSI	38 (2.58)	**27 (7)**	68 (10.23)	37 (3.76)	47 (2.66)	43 (2.6)	38 (2.63)	34 (8.08)	28 (3.43)	1.55
FESI	34 (2.53)	49 (8.24)	65 (9.37)	35 (3.4)	35 (2.54)	35 (2.46)	**24 (2.71)**	82 (9.69)	37 3.89)	1.64
DT	35 (2.54)	67 (10.22)	55 (9.64)	**30 (3.93)**	43 (2.77)	50 (2.66)	34 (2.59)	71 (10.29)	32 (4.06)	1.67
AL	**17 (2.59)**	60 (9.65)	48 (9.92)	40 (3.88)	21 (2.71)	19 (2.63)	20 (2.45)	78 (11.48)	44 (4.29)	1.69
SL	50 (2.95)	63 (7.52)	143 (11.27)	12 (4.67)	87 (2.98)	102 (2.97)	75 (2.92)	168 (10.31)	**3 (1.67)**	2.15
Arb	**9 (1.89)**	14 (5.14)	18 (6.44)	11 (3.1)	15 (2.6)	17 (2.59)	17 (2.53)	35 (7.94)	20 (3.1)	3.09
FD	102 (2.95)	495 (13.19)	1393 (15.32)	124 (6.68)	141 (2.88)	158 (2.87)	135 (2.85)	1172 (17.46)	**7 (2.43)**	3.75
Len	295 (2.99)	735 (13.97)	1763 (17.24)	**7 (3.14)**	608 (2.95)	674 (2.95)	613 (2.95)	844 (14.31)	28 (3.75)	4.44
LA	**10 (1.5)**	30 (6.1)	1735 (17.54)	220 (5.58)	47 (2.7)	42 (2.67)	43 (2.7)	5336 (20.79)	430 (6.33)	67.66
Mean	71 (2.55)	160.73 (9.36)	500.36 (11.65)	**50.82 (3.98)**	115.64 (2.76)	124.45 (2.73)	113.18 (2.71)	734.82 (12.2)	61.45 (3.57)	

TABLE 16 FUZZY RULE BASE GENERATED FOR THE ARBITRAGE DATA-SET.

R1: If ProftAfter Is VeryLow Then Class = 0 with RW = [1.5E-4, 2.0E-4]

R2: If InterRAte Is VeryLow Then Class = 0 with RW = [3.6E-5, 8.4E-5]

R3: If ProftAfter Is VeryHigh Then Class = 1 with RW = [5.5E-6, 1.0E-5]

R4: If ProftAfter Is Medium Then Class = 1 with RW = [1.3E-5, 3.5E-5]

R5: If Basis Is Medium And FutT-t Is Low And C-P IS Low Then Class = 1 with RW = [4.4E-5, 7.5E-5]

R6: If Und Is Medium And FutT-t Is Low And C-P IS Low Then Class = 1 with RW = [4.4E-5, 7.5E-5]

R7: If Und Is Medium And FutT-t Is Medium And ProftAfter Is Low Then Class = 1 with RW = [3.2E-5, 7.4E-5]

R8: If InterRAte Is Medium And ProftAfter Is Low Then Class =1 CF: [1.5E-4, 2.2E-4]

R9: If FutT-t Is High And C-P Is Medium And ProftAfter Is Low Then Class = 1 with RW = [2.0E-5, 4.4E-5]

performance, with respect to several state-of-the-art white box classifiers (C4.5 decision tree, type-1 and interval-valued fuzzy counterparts) using SMOTE as pre-processing technique; 4) Our new approach enhances the results provided by the cost sensitive C4.5 decision tree and 5) $\text{IVTURS}_{FARC-HD}^{RRW\_I}$ provides competitive results versus FURIA applied after applying SMOTE, while our system has the advantages of avoiding the use of pre-processing techniques and obtaining more interpretable and transparent models for the financial analysts.

For our future work, we will aim to investigate the use of type-2 fuzzy sets and the use of multi-objective multi-constraints evolutionary algorithms to optimize the type-2 fuzzy parameters and to generate a compact rule set satisfying various financial objectives. We will aim also to investigate methods allowing the fast update of the generated financial models to handle dynamic and fast changing financial markets. Furthermore, the proposed method could be extended to multi-class problems using decomposition techniques [64] and multiple classifier systems [65, 66].

REFERENCES

[1] N. Japkowicz and S. Stephen, "The class imbalance problem: a systematic study," Intelligent Data Analysis, vol. 6, no. 5, pp. 429--450, 2002

[2] C. Phua, D. Alahakoon and V. Lee, "Minority report in fraud detection: classification of skewed data," SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 50-59, 2004.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligent Research, vol. 16, pp. 321-357, 2002.

[4] J. Wang, M. Xu, H. Wang and J. Zhang, "Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding," In Proceedings of the 8th international Conference on Signal Processing, ICSP 3, pp. 1815-1818, 2006.

[5] K. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55 no. (1–2), pp. 307–319, 2003.

[6] H. S. Kim, S. Y. Sohn, "Support vector machines for default prediction of SMEs based on technology credit," European Journal of

Operational Research, vol. 201, no. 3, pp. 838–846, 2010.

[7] E. Giacomini, "Neural Networks in Quantitative Finance," Master Thesis, Humboldt-Universität zu Berlin, 2003.

[8] R. Lawrence, "Using Neural Networks to Forecast Stock Market Prices," University of Manitoba, 1997

[9] K. Kwong, "Financial Forecasting Using Neural Network or Machine Learning Techniques," University of Queensland, 2001.

[10] A. Garcia-Almanza and E. Tsang, "Forecasting stock prices using genetic programming and chance discovery", Proceedings of the 12th International Conference On Computing In Economics And Finance, 2006.

[11] A. Garcia-Almanza and E. Tsang, "Evolving decision rules to predict investment opportunities," The International Journal of Automation and Computers, vol. 5, no. 1, pp. 22-31, 2008.

[12] A. Garcia-Almanza, "New classification methods for gathering patterns in the context of genetic programming", PhD Thesis, Department of Computing and Electronic Systems, University of Essex, 2008.

[13] R. Nogueira, S. M. Vieira and J. M. C Sousa, "The prediction of bankruptcy using fuzzy classifiers," ICSC Congress on Computational Intelligence Methods and Applications, pp. 95-100, 2005.

[14] V. Snášel, P. Krömer, J. Platoš and A. Abraham, "The evolution of fuzzy classifier for data mining with applications," 8th international conference on Simulated evolution and learning, pp. 349-358, 2010.

[15] Z. Yang and G. Gan "Application of Fuzzy Classification in Bankruptcy Prediction," 4th international conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications - with Aspects of Theoretical and Methodological Issues, pp. 921-928, 2008.

[16] P. Bonissone, "Development and maintenance of fuzzy models in financial applications," in Soft Methodology and Random Information Systems, M. Lopez-Diaz, M. Gil, P. Grzegorzwski, O. Hyrniewicz, J. Lawry, Eds. Berlin, Germany: Springer Verlag, pp. 50-66, 2004.

[17] L. Angstenberger, "Dynamic Fuzzy Pattern Recognition with Applications to Finance and Engineering," Kluwer Academic Publishers, 2001.

[18] D. Bernardo, H. Hagras and E. Tsang "An Interval Type-2 Fuzzy Logic Based System for Model Generation and Summarization of Arbitrage Opportunities in Stock Markets" 12th UK Workshop on Computational Intelligence (UKCI), art. No. 6335765, 2012.

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

[20] J. Sanz, A. Fernández, H. Bustince, and F. Herrera, "A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position," International Journal of Approximate Reasoning, vol. 52, no. 6, pp. 751–766, 2011.

[21] R. Sambuc, "Function Φ-flous, application a l'aide au diagnostic en pathologie thyroidienne," Ph.D. dissertation, University of Marseille, 1975.

[22] J. Sanz, A. Fernández, H. Bustince and F. Herrera, "IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection," IEEE Transactions on Fuzzy Systems, vol. 21, no. 3, pp. 399–411, 2013.

[23] M. Galar, J. Fernandez, G. Beliakov, and H. Bustince, "Interval-valued fuzzy sets applied to stereo matching of color images," IEEE Transactions on Image Processing, vol. 20, no. 7, pp. 1949–1961, 2011.

[24] A. Jurio, M. Pagola, D. Paternain, C. Lopez-Molina, and P. Melo-Pinto, "Interval-valued restricted equivalence functions applied on clustering techniques," in Proceedings of the joint 2009 international fuzzy systems association world congress and 2009 european society of fuzzy logic and technology conference, J. Carvalho, D. Kaymak, and J. Sousa, Eds., 2009, pp. 831–836, Lisbon, Portugal, Jul. pp. 20-24, 2009.

[25] J. Quinlan, "C4.5: Programs for Machine Learning," San Mateo-California: Morgan Kaufmann Publishers, 1993.

[26] K. M. Ting. "An instance-weighting method to induce cost-sensitive trees," IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 3, pp. 659-665, 2002.

[27] J. Hühn and E. Hüllermeier, "FURIA: an algorithm for unordered fuzzy rule induction," Data Mining and Knowledge Discovery, vol. 19, no. 3, pp. 293–319, 2009.

[28] L. A. Zadeh, "Fuzzy sets," Information and control, vol. 8, no. 3, pp. 338-353, 1965.

[29] H. Wong, B.Q. Hu, "Application of interval clustering approach to water quality evaluation," Journal of Hydrology, vol. 491, no. 29, pp. 1–12, 2013.

[30] J. Sanz, A. Fernández, H. Bustince and F. Herrera, "IIVFDT: Ignorance functions based interval-valued fuzzy decision tree with genetic tuning," International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems, vol. 20, no. Suppl 2, pp. 1-30, 2012.

[31] M. B. Gorlzakczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets, " Fuzzy Sets and Systems, vol. 21, pp. 1–17, 1987.

[32] I. B. Turksen, "Interval valued fuzzy sets based on normal forms, " Fuzzy Sets and Systems, vol. 20, no. 2, pp. 191–210, 1986.

[33] H. Bustince, E. Barrenechea, and M. Pagola, "Generation of interval-valued fuzzy and atanassov's intuitionistic fuzzy connectives from fuzzy connectives and from $K\alpha$ operators: laws for conjunctions and disjunctions, amplitude," International Journal of Intelligent Systems, vol. 23, no. 6, pp. 680–714, 2008.

[34] H. Bustince, M. Galar, B. Bedregal, A. Kolesarova and R. Mesiar, "A new approach to interval-valued Choquet integrals and the problem of ordering in interval-valued fuzzy sets applications," IEEE Transactions on Fuzzy Systems, vol. 21, no. 6, pp. 1150–1162, 2013.

[35] H. Bustince, J. Fernandez, A. Kolesárová and R., Mesiar, "Generation of linear orders for intervals by means of aggregation functions," Fuzzy Sets and Systems, vol. 220, pp. 69–77, 2013.

[36] Z. S. Xu and R. R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," International Journal of General Systems, vol. 35, no. 4, pp. 417–433, 2006.

[37] L. Stefanini, "A generalization of Hukuhara difference and division for interval and fuzzy arithmetic," Fuzzy Sets and Systems, vol. 161, no. 11, pp. 1564–1584, 2010.

[38] G. Deschrijver, C. Cornelis, and E. Kerre, "On the representation of intuitionistic fuzzy t-norms and t-conorms," IEEE Transactions on Fuzzy Systems, vol. 12, no. 1, pp. 45–61, 2004.

[39] H. Bustince, E. Barrenechea, and M. Pagola, "Restricted equivalence functions," Fuzzy Sets and Systems, vol. 157, no. 17, pp. 2333–2346, 2006.

[40] Q. Yang and X. Wu, "10 challenging problems in data mining research," International Journal of Information Technology and Decision Making, vol. 5, no. 4, pp. 597–604, 2006.

[41] V. López, A. Fernández, S. García, V. Palade and F. Herrera, "An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics," Information Sciences, vol. 250, pp. 113–141, 2013.

[42] G. M. Weiss, "Mining with rarity: a unifying framework". SIGKDD Explorations, vol. 6, no. 1, pp. 7–19, 2004.

[43] M. Kubat, R. C. Holte and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," Machine Learning, vol. 30, no. 2–3, pp. 195–215, 1998.

[44] J. Alcalá-Fdez, R. Alcala, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," IEEE Transactions on Fuzzy Systems, vol. 19, no. 5, pp. 857–872, 2011.

[45] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Second Edition. San Fransisco: Morgan Kaufmann, 2006.

[46] C. Zhang and S. Zhang, "Association Rule Mining: Models and Algorithms Series," Lecture Notes in Computer Science, LNAI 2307, Berlin: Springer-Verlag, 2002.

[47] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification," Second Edition, John Wiley, 2001.

[48] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," IEEE Transactions on Fuzzy Systems, vol. 13, no. 4, pp. 428–435, 2005.

[49] R. Alcalá, J. Alcalá-Fdez, and F. Herrera, "A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection," IEEE Transactions on Fuzzy Systems, vol. 15, no. 4, pp. 616–635, 2007.

[50] F. Herrera and L. Martínez, "A 2-tuple fuzzy linguistic representation model for computing with words," IEEE Transactions on Fuzzy Systems, vol. 8, no. 6, pp. 746–752, 2000.

[51] O. Cordón, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," International Journal of Approximate Reasoning, vol. 20, no. 1, pp. 21–45, 1999.

[52] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. C. Fernández, and F. Herrera, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," Soft Computing, vol. 13, no. 3, pp. 307–318, 2009.

[53] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," Fuzzy Sets and Systems, vol. 141, no. 1, pp. 5–31, 2004.

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

[55] L. Eshelman, "Foundations of Genetic Algorithms," Morgan Kaufman, 1991, Ch. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, pp. 265–283.

[56] J. Huang, C.X. Ling, "Using AUC and accuracy in evaluating learning algorithms, " IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 3, pp. 299–310, 2005.

[57] F. Herrera, M. Lozano, and A. M. Sánchez, "A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study," International Journal of Intelligent Systems, vol. 18, no. 3, pp. 309–338, 2003.

[58] L. J. Eshelman and J. Schaffrer, "Foundations of Genetic Algorithms," Morgan Kaufman, 1993, Ch. Real coded genetic algorithms and interval schemata, pp. 187–202.

[59] A. Orriols-Puig and E. Bernadó-Mansilla, "Evolutionary rule-based systems for imbalanced datasets," Soft Computing, vol. 13, no. 3, pp. 213–225, 2009.

[60] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," Journal of Machine Learning Research, vol. 7, pp. 1–30, 2006.

[61] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," Information Sciences, vol. 180, no. 10, pp. 2044–2064, 2010.

[62] J. L. Hodges and E. L. Lehmann, "Ranks methods for combination of independent experiments in analysis of variance," Annals of Mathematical Statistics, vol. 33, pp. 482–497, 1962.

[63] S. Holm, "A simple sequentially rejective multiple test procedure," Scandinavian Journal of Statistics, vol. 6, pp. 65–70, 1979.

[64] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," Pattern Recognition, vol. 44, no. 8, pp. 1761-1776, 2011.

[65] A. Fernández, V. López, M. Galar, M. J. del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," Knowledge-Based Systems, vol. 42, pp. 97-110, 2013.

[66] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," Information Fusion, vol. 16, pp. 3-17, 2014.
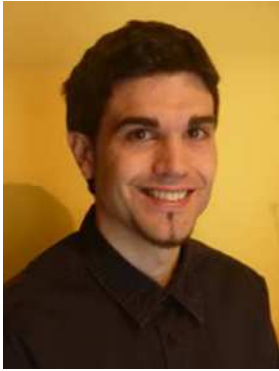
José Antonio Sanz received the M.Sc. degree in computer sciences and the Ph.D. degree, both form the Public University of Navarra in 2008 and 2011 respectively. He is currently an Associate Lecturer with the Department of Automatics and Computation, Public University of Navarra. He is the author of 14 published original articles and is involved in teaching artificial intelligence for students of computer sciences. He received the best paper award in the FLINS 2012 international conference and the Pepe Millá award in 2014. His research interests include fuzzy techniques for classification problems, interval-valued fuzzy sets, genetic fuzzy systems, data mining and medical applications of soft computing techniques. Dr. Sanz is member of the European Society for Fuzzy Logic and Technology.

Dario Bernardo received the B.Sc. and M.Sc. degrees in Computer Engineering for enterprise management from University of Pisa, Italy, and another M.Sc. degree in Computational and Software Techniques in Engineering from Cranfield University, UK. He is in his final year of his Ph.D in computer science at University of Essex, Colchester, UK, and has been a Knowledge Partnership Associate during the Ph.D duration. He is currently a data scientist working in London in the field of predictive analytics in finance. His major research interests are in computational intelligence, machine learning, type-2 fuzzy systems, fuzzy logic, genetic algorithms, and evolutionary computation. His research interests also include big data analytics, algorithm scalability and optimization.

Francisco Herrera received his M.Sc. in Mathematics in 1988 and Ph.D. in Mathematics in 1991, both from the University of Granada, Spain. He is currently a Professor in the Department of Computer Science and Artificial Intelligence at the University of Granada. He has been the supervisor of 34 Ph.D. students. He has published more than 270 papers in international journals. He currently acts as Editor in Chief of the international journals "Information Fusion" (Elsevier) and "Progress in Artificial Intelligence (Springer). He acts as editorial board member of a dozen of journals, among others: International Journal of Computational Intelligence Systems, IEEE Transactions on Fuzzy Systems, Information Sciences, Knowledge and Information Systems, Fuzzy Sets and Systems, Applied Intelligence, Knowledge-Based Systems, and Swarm and Evolutionary Computation. He received the following honors and awards: ECCAI Fellow 2009, IFSA Fellow 2013, 2010 Spanish National Award on Computer Science ARITMEL to the "Spanish Engineer on Computer Science", International Cajastur "Mamdani" Prize for Soft Computing (Fourth Edition, 2010), IEEE Transactions on Fuzzy System Outstanding 2008 Paper Award (bestowed in 2011), 2011 Lotfi A. Zadeh Prize Best paper Award of the International Fuzzy Systems Association, and 2013 AEPIA Award to a scientific career in Artificial Intelligence (September 2013). His current research interests include bibliometrics, computing with words in decision making, information fusion, evolutionary algorithms, evolutionary fuzzy systems, biometrics, data preprocessing, data mining, cloud computing and big data.

Humberto Bustince received his Bs. C. degree on Physics from the Salamanca University, Spain, in 1983 and his Ph.D. degree in Mathematics from the Public University of Navarra, Pamplona, Spain, in 1994. He has been a teacher at the Public University of Navarra since 1991, and he is currently a Full Professor with the Department of Automatics and Computation. He served as subdirector of the Technical School for Industrial Engineering and Telecommunications from 01/01/2003 to 30/10/2008 and he was involved in the implantation of Computer Science courses at the Public University of Navarra. He is currently involved in teaching artificial intelligence for students of computer sciences. Dr. Bustince has authored more than 100 journal papers (Web of Knowledge), and more than 90 contributions to international conferences. He has also been co-author of four books on fuzzy theory and extensions of fuzzy sets. He is fellow of the IEEE Computational Intelligence Systems society and Member of the board of the European Society for Fuzzy Logic and Applications (EUSFLAT). He currently acts as Editor in chief of the Mathware&Soft Computing Magazine and of Notes on Intuitionistic Fuzzy Sets. He is also guest editor of the Fuzzy Sets and Systems journal and member of the editorial board of the Journal of Intelligence&Fuzzy Systems, the International Journal of Computational Intelligence Systems and the Axioms Journal. His current research interests include interval-valued fuzzy sets, Atanassov's intuitionistic fuzzy sets, aggregation functions, implication operators, inclusion measures, image processing, decision making and approximate reasoning.

Hani Hagras (M'03–SM'05, F'13) received theB.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Alexandria, Egypt, and the Ph.D. degree in computer science from the University of Essex, Colchester, U.K. He is a Professor in the School of Computer Science and Electronic Engineering, Director of the Computational Intelligence Centre and the Head of the Fuzzy Systems Research Group in the University of Essex, UK. His major research interests are in computational intelligence, notably type-2 fuzzy systems, fuzzy logic, neural networks, genetic algorithms, and evolutionary computation. His research interests also include ambient intelligence, pervasive computing and intelligent buildings. He is also interested in embedded agents, robotics and intelligent control. He has authored more than 200 papers in international journals, conferences and books. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and he is also a Fellow of the Institution of Engineering and Technology (IET (IEE). He was the Chair of IEEE Computational Intelligence Society (CIS) Senior Members Sub-Committee. His research has won numerous prestigious international awards where most recently he was awarded by the IEEE Computational Intelligence Society (CIS), the 2013 Outstanding Paper Award in the IEEE Transactions on Fuzzy Systems and he was also awarded the 2006 Outstanding Paper Award in the IEEE Transactions on Fuzzy Systems. He is an Associate Editor of the IEEE Transactions on Fuzzy Systems. He is also an Associate Editor of the International Journal of Robotics and Automation, the Journal of Cognitive Computation and the Journal of Ambient Computing and Intelligence. He is a member of the IEEE Computational Intelligence Society (CIS) Fuzzy SystemsTechnical Committee. Prof. Hagras chaired several international conferences where he served as the General Co-Chair of the 2007 IEEE International Conference on Fuzzy systems London, July 2007 and he served as the Chair of 2009, 2011, 2013 and 2014 IEEE Symposium on Intelligent Agents, he has also served as the Chair of 2011 IEEE Symposium on Advanced to Type-2 FuzzyLogic Systems, Paris 2011.