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 to obtain more accurate prediction models in order to decrease 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 real-world imbalanced financial datasets without using sampling techniques that 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 interval-valued fuzzy rule-based classification system with tuning and rule selection (IVTURS$_{\mathrm{FARC\text{-}HD}}$) for the modeling and prediction of real-world financial applications. This proposed system allows obtaining good prediction 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 11 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 that use the synthetic minority oversampling technique (SMOTE) 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 preprocessing techniques, and it provides interpretable models that allow obtaining more accurate results.**

***Index Terms*—Evolutionary algorithms, financial applications, interval-valued fuzzy rule-based classification systems, interval-valued fuzzy sets (IVFSs).**

Manuscript received November 19, 2013; revised February 28, 2014; accepted June 7, 2014. Date of publication July 8, 2014; date of current version July 31, 2015. This work was supported in part by the Spanish Ministry of Science and Technology under Project TIN2011-28488 and Project TIN2013-40765-P.

J. A. Sanz and H. Bustince are with the Department of Automática y Computación, Universidad Publica de Navarra, 31006 Pamplona, Spain (e-mail: joseantonio.sanz@unavarra.es; bustince@unavarra.es).

D. Bernardo and H. Hagras are with the Computational Intelligence Centre, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: dbernaa@essex.ac.uk; hani@essex.ac.uk).

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2014.2336263

I. 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 are 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, preprocessing techniques like undersampling or oversampling are usually applied, but both of them present problems. On the one hand, undersampling techniques may increment the noise since they could eliminate some important patterns. On the other hand, oversampling techniques, such as 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]–[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 fulfill 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 that have been applied for various financial applications [13]. FRBSs have been successfully applied on credit approval, loan portfolio, bankruptcy prediction, and security management [13]. 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 that 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. $\text{IVTURS}_{\text{FARC-HD}}$ [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, $\text{IVTURS}_{\text{FARC-HD}}$ 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 $\text{IVTURS}_{\text{FARC-HD}}$ 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.

1) 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.
2) A technique to classify the incoming examples, even if they do not match any fuzzy rule in the generated rule base (RB). To do so, the similarity among the uncovered example and the rules is considered.
3) 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 $\text{IVTURS}_{\text{FARC-HD}}$ with rescaling rule weight for imbalanced classification ($\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC-HD}}$), has been tested through various experiments using real-world datasets from 11 financial applications. The obtained results, which are statistically supported, show that $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC-HD}}$ 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 such as Fuzzy Unordered Rule Induction Algorithm (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, $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC-HD}}$ avoids using preprocessing techniques and gives accurate results as well as produces a reduction in the number of generated rules, which implies that it provides more transparent and highly interpretable models.

This paper is organized as follows. Section II provides some needed background material and the related work about FRBCSs. The proposed compact evolutionary $\text{IVTURS}^{\text{RRW\_I}}_{\text{FARC-HD}}$ is presented in Section III. The experimental results are shown in Section IV, while the conclusions are drawn in Section V.

II. Background

In this section, we present the background needed to understand the remainder of this paper. We start by presenting some theoretical concepts about IVFSs. Then, we describe the problem of the imbalanced datasets 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 (FARC) considered in this paper.

A. *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,

$$L([0,1]) = \{\boldsymbol{x} = [\underline{x}, \bar{x}] | \, 0 \leq \underline{x} \leq \bar{x} \leq 1\}. \tag{1}$$

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]–[33]: An IVFS A on the universe $U \neq \emptyset$ is a mapping $A_{\mathrm{IV}}: U \to L([0,1])$, such that

$$A_{IV}(u_i) = [\underline{A}(u_i), \bar{A}(u_i)] \in L([0,1]) \quad \text{for all } u_i \in U. \tag{2}$$

Obviously, $[\underline{A}(u_i), \bar{A}(u_i)]$ 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 is not known. The length of the interval membership degree, i.e., $L(A_{IV}(u_i)) = \bar{A}(u_i) - \underline{A}(u_i)$, 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] and [35]. A particular case of these linear orders is the one defined by Xu and Yager [36], which is based on the score and accuracy degrees as in

$$\begin{aligned}&[\underline{x}, \bar{x}] \leq [\underline{y}, \bar{y}], \text{ if and only if } \underline{x} + \bar{x} < \underline{y} + \bar{y} \text{ or} \\ &\underline{x} + \bar{x} = \underline{y} + \bar{y} \text{ and } \bar{x} - \underline{x} \geq \bar{y} - \underline{y}.\end{aligned} \tag{3}$$

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. Therefore, 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 $[\underline{x}, \bar{x}]$ and $[\underline{y}, \bar{y}]$ be two intervals, with $\underline{x}, \bar{x}, \underline{y}, \bar{y} \in \mathbb{R}^+$ so that $[\underline{x}, \bar{x}] \leq_L [\underline{y}, \bar{y}]$, which means that $\underline{x} \leq \underline{y}$, $\bar{x} \leq \bar{y}$ (this is not a total order relationship for intervals), and $\underline{y} > 0$; the rules of interval arithmetic are as follows [36]:

1) Addition: $[\underline{x}, \bar{x}] + [\underline{y}, \bar{y}] = [\underline{x} + \underline{y}, \bar{x} + \bar{y}]$. (4)

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

3) Product: $[\underline{x}, \bar{x}] * [\underline{y}, \bar{y}] = [\underline{x} * \underline{y}, \bar{x} * \bar{y}]$. (6)

4) Division: $\dfrac{[\underline{x}, \bar{x}]}{[\underline{y}, \bar{y}]} = \left[\wedge\left(\dfrac{\underline{x}}{\underline{y}}, \dfrac{\bar{x}}{\bar{y}}\right), \vee\left(\dfrac{\underline{x}}{\underline{y}}, \dfrac{\bar{x}}{\bar{y}}\right)\right]$. (7)

where $\wedge$ represents the *t*-norm (minimum) and $\vee$ 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$_{\mathrm{FARC\text{-}HD}}$. 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 IV-REF associated with an interval-valued negation $\boldsymbol{N}$ is a function

$$\mathrm{IV-REF}: \; L([0,1])^2 \to L([0,1]) \tag{8}$$

So that

IR1) $\mathrm{IV-REF}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{IV-REF}(\boldsymbol{y}, \boldsymbol{x})$

for all $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$;

IR2) $\mathrm{IV-REF}(\boldsymbol{x}, \boldsymbol{y}) = [1,1]$, if and only if $\boldsymbol{x} = \boldsymbol{y}$;

IR3) $\mathrm{IV-REF}(\boldsymbol{x}, \boldsymbol{y}) = [0,0]$ if and only if $\boldsymbol{x} = [1,1]$

and $\boldsymbol{y} = [0,0]$ or $\boldsymbol{x} = [0,0]$ and $\boldsymbol{y} = [1,1]$

IR4) $\mathrm{IV-REF}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{IV-REF}(\boldsymbol{N}(\boldsymbol{x}), \boldsymbol{N}(\boldsymbol{y}))$ 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 $\boldsymbol{I}V -$ 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 as follows:

$$\begin{aligned}&\mathrm{IV-REF}([\underline{x}, \bar{x}], [\underline{y}, \bar{y}]) \\ &= \left[T\left(\phi_1^{-1}\left(1 - \left|\phi_2(\underline{x}) - \phi_2(\underline{y})\right|\right), \phi_1^{-1}\left(1 - \left|\phi_2(\bar{x}) - \phi_2(\bar{y})\right|\right)\right)\right. \\ &\quad \left. S\left(\phi_1^{-1}\left(1 - \left|\phi_2(\underline{x}) - \phi_2(\underline{y})\right|\right), \phi_1^{-1}\left(1 - \left|\phi_2(\bar{x}) - \phi_2(\bar{y})\right|\right)\right)\right]\end{aligned} \tag{9}$$

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 $\boldsymbol{N}([\underline{x}, \bar{x}]) = [\phi_2^{-1}(1 - \phi_2(\bar{x}))], \phi_2^{-1}(1 - \phi_2(\underline{x}))]$.

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

$$\begin{aligned}\mathrm{IV-REF}([\underline{x}, \bar{x}], [\underline{y}, \bar{y}]) = \; &[\wedge(1 - |\underline{x} - \underline{y}|, 1 - |\bar{x} - \bar{y}|), \\ &\vee(1 - |\underline{x} - \underline{y}|, 1 - |\bar{x} - \bar{y}|)].\end{aligned} \tag{10}$$

Using (10) as the IV-REF associated with the negation $N([\underline{x}, \bar{x}]) = [1 - \bar{x}, 1 - \underline{x}]$ and letting $\boldsymbol{x} = [\underline{x}, \bar{x}] = [0,0]$, $\boldsymbol{y} = [\underline{y}, \bar{y}] = [0.3, 0.6]$ and $\boldsymbol{z} = [\underline{z}, \bar{z}] = [1,1]$ be three intervals, the following conditions IR1–IR5 fulfill.

IR1) $\mathrm{IV-REF}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{IV-REF}(\boldsymbol{y}, \boldsymbol{x})$

$\Rightarrow [0.4, 0.7] = [0.4, 0.7]$

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

$\mathrm{IV-\ REF}(\boldsymbol{y}, \boldsymbol{y}) = [1,1]$

TABLE I
CONFUSION 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)

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

$\mathrm{IV}-\ \mathrm{REF}(\boldsymbol{x},\boldsymbol{z}) = [0,0]$

IR4) $\mathrm{IV}-\mathrm{REF}(\boldsymbol{x},\boldsymbol{y}) = \mathrm{IV}-\mathrm{REF}(\boldsymbol{N}(\boldsymbol{x}),\boldsymbol{N}(\boldsymbol{y}))$

$\Rightarrow \boldsymbol{IV}-\mathrm{REF}([0,0],[0.3,0.6]) =$

$\mathrm{IV}-\ \mathrm{REF}([1,1],[0.4,0.7]) \Rightarrow [0.4,0.7] = [0.4,0.7]$

IR5) if $\boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$, then $\mathrm{IV}-\mathrm{REF}(\boldsymbol{x},\boldsymbol{y})$

$\geq_L\ \mathrm{IV}-\mathrm{REF}(\boldsymbol{x},\boldsymbol{z})$ and $\mathrm{IV}-\mathrm{REF}(\boldsymbol{y},\boldsymbol{z})$

$\geq_L\ \mathrm{IV}-\mathrm{REF}(\boldsymbol{x},\boldsymbol{z})$

$\Rightarrow \boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$; therefore, $[0.4, 0.7]$

$\geq_L\ [0,0]$ and $[0.3,0.4] \geq_L\ [0,0]$.

B. Imbalanced Datasets 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, such as 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 dataset problems are very common in real-world financial datasets, 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 datasets 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 I 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 (11) and accuracy defined in (12) as follows:

$$\mathrm{Err} = \frac{\mathrm{FP}+\mathrm{FN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{11}$$

$$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} = 1-\mathrm{Err}. \tag{12}$$

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 with the majority class [42]. Therefore, in the framework of imbalanced problems, there are more accurate metrics. For instance, from Table I, four performance measures can be derived in order to take into account the classification rate of each class independently.

1) True positive rate ($\mathrm{TP}_{\mathrm{rate}}$) is the percentage of correctly classified examples belonging to the minority class.
2) True negative rate ($\mathrm{TN}_{\mathrm{rate}}$) is the percentage of correctly classified examples belonging to the majority class.
3) False positive rate ($\mathrm{FP}_{\mathrm{rate}}$) is the percentage of misclassified examples belonging to the majority class.
4) False negative rate ($\mathrm{FN}_{\mathrm{rate}}$) is the percentage of misclassified examples belonging to the minority class.

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

$$\mathrm{GM} = \sqrt{\mathrm{TP}_{\mathrm{rate}} * \mathrm{TN}_{\mathrm{rate}}}. \tag{13}$$

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 Section II-C1 and 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 dependences among items in a database using expression like $A \to B$, 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:

$$\begin{aligned} R_j: &\ \text{If } x_1 \text{ is } A_{j1} \text{ and} \ldots \text{and } x_n \text{ is } A_{jn}, \text{ then} \\ &\ \mathrm{Class} = C_j \text{ with } \mathrm{RW}_j \end{aligned} \tag{14}$$

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

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

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

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

$$\text{Support}(A_j \to C_j) = \frac{\sum_{x_p \in \text{Class } C_j} \mu_{A_j}(x_p)}{|N|}. \tag{16}$$

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

$$\text{Confidence}(A_j \to C_j) = \frac{\sum_{x_p \in \text{Class } C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{N} \mu_{A_j}(x_p)}. \tag{17}$$

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 RB is generated. To do so, a search tree is generated for each class in which the frequent item sets are computed by applying the support [see (16)] and the confidence [see (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 item set. 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 RB 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 two-tuple 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. Fig. 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.

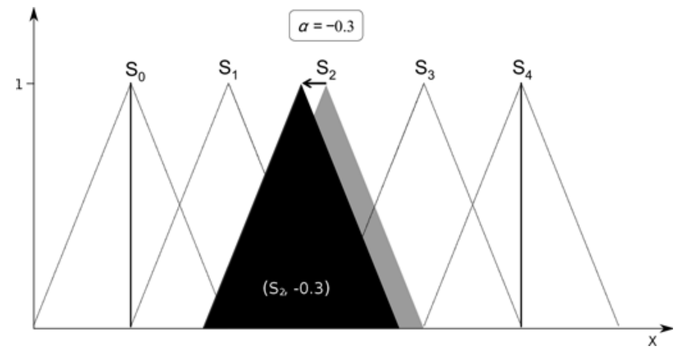


Fig. 1. Lateral displacement of the membership function s_2. Gray and black triangles are its initial and final positions, respectively.

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 (18), and then, the example x_p is classified in the class having the maximum total strength of the vote

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

where L is the number of rules of the 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.

3) Interval-Valued Fuzzy Rule-Based Classification System With Tuning and Rule Selection: IVTURS$_{\text{FARC-HD}}$ [22] is an extension of FARC-HD that models the linguistic terms with IVFSs instead of with FSs. IVTURS$_{\text{FARC-HD}}$ outperforms the performance of both the original FARC-HD and the FURIA [27]. This method is composed of the following steps.

1) *RB generation:* This method learns an initial type-1 FRBCS so as to initialize the parameters of the IV-FRBCS. Specifically, the initial RB is the one obtained after the application of the two first stages of the FARC-HD algorithm (see stages 1 and 2 of Section II-C1).
2) *IVFS 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 FS 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

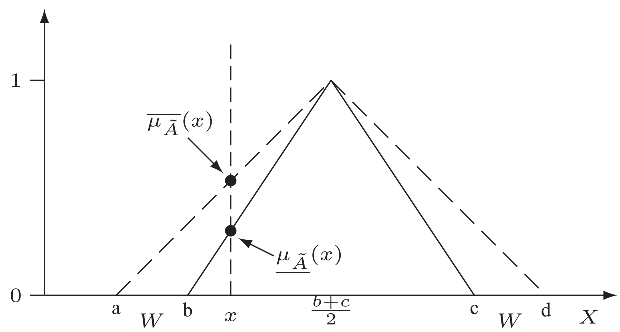


Fig. 2. Initially constructed IVFS. The solid line is the initial fuzzy set, and consequently, it is the lower bound of the IVFS. The dashed line is the upper bound of the IVFS.

bound as depicted in Fig. 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 (17) using interval arithmetic.

3) *Interval-valued FRM (IV-FRM):* The classic FRM described in Section II-C1 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 (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 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 (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].

III. Proposed Compact Evolutionary Interval-Valued Fuzzy Rule-Based Classification System for Financial Applications Modeling and Prediction

In this section, we present $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}^{\mathrm{RRW\_I}}$, which is an IV-FRBCS built on the basis of $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$ [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 preprocessing 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.

Fig. 3 shows an overview of the proposed method, which shares two basic components of the $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$ method, namely the generation of the initial IV-FRBCS (applying the two first steps introduced in Section II-C2) and the IV-FRM. Our new method generates an initial IV-FRBCS using the training examples, and then, the created RB is scaled using the process introduced in Section III-A. After these steps, an evolutionary process is applied to adapt the system's parameters to the problem (using the training examples again), which is described in Section III-C. 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 III-B to handle uncovered examples is used.

The novelty of our new method consists of a method to rescale the rule weights of the generated RB in order to face the imbalance problem at the 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 RB. 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 Fig. 1, and the amplitude of the support of the upper bound using the parameter W, as shown in Fig. 2, as well as the parameters a and b used to construct the IV-REFs [see (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.

A. 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 datasets faced in real-world financial applications. The need for this procedure is easily shown in the following example in Table II, which shows the RB generated by the $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$ method for the fourth partition of the Yeast-1-4-5-8_versus_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.

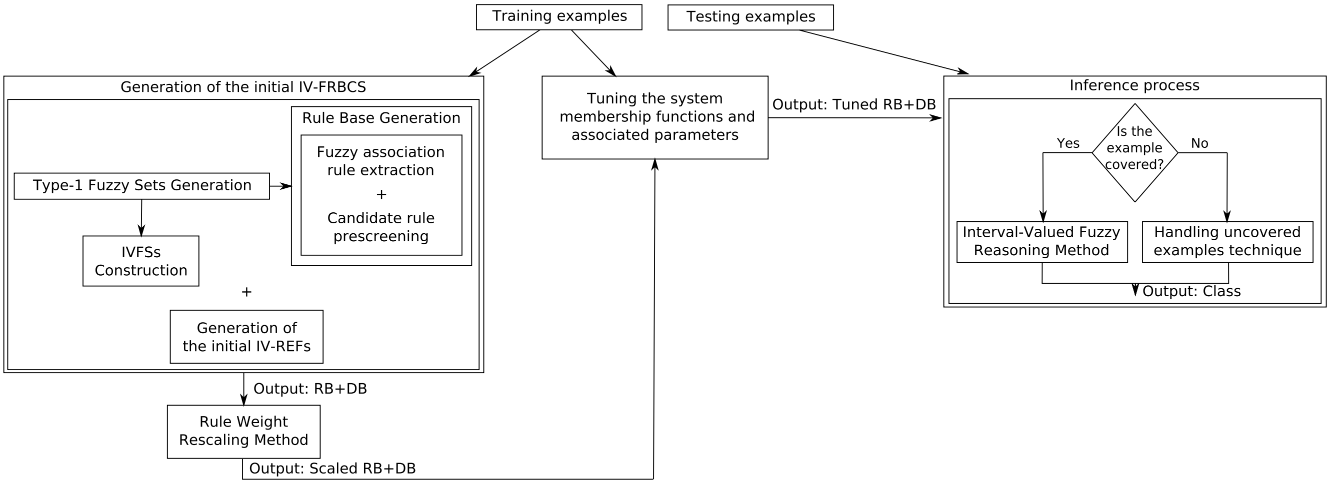


Fig. 3. Overview of the proposed compact evolutionary IV-FRBCS for financial applications modeling and prediction.

TABLE II
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]

From Table II, 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].

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

1) To compute the cumulative matching degrees, for each rule, the matching degrees of examples belonging to the rule class are summed:

$$[\underline{\mathrm{CMD}_j}, \overline{\mathrm{CMD}_j}] = \left[\sum_{x_p \in \mathrm{Class}\, k} \underline{\mu_{Aj}}(x_p), \sum_{x_p \in \mathrm{Class}\, k} \overline{\mu_{A_j}}(x_p)\right] \quad j = 1, \ldots, L. \tag{19}$$

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

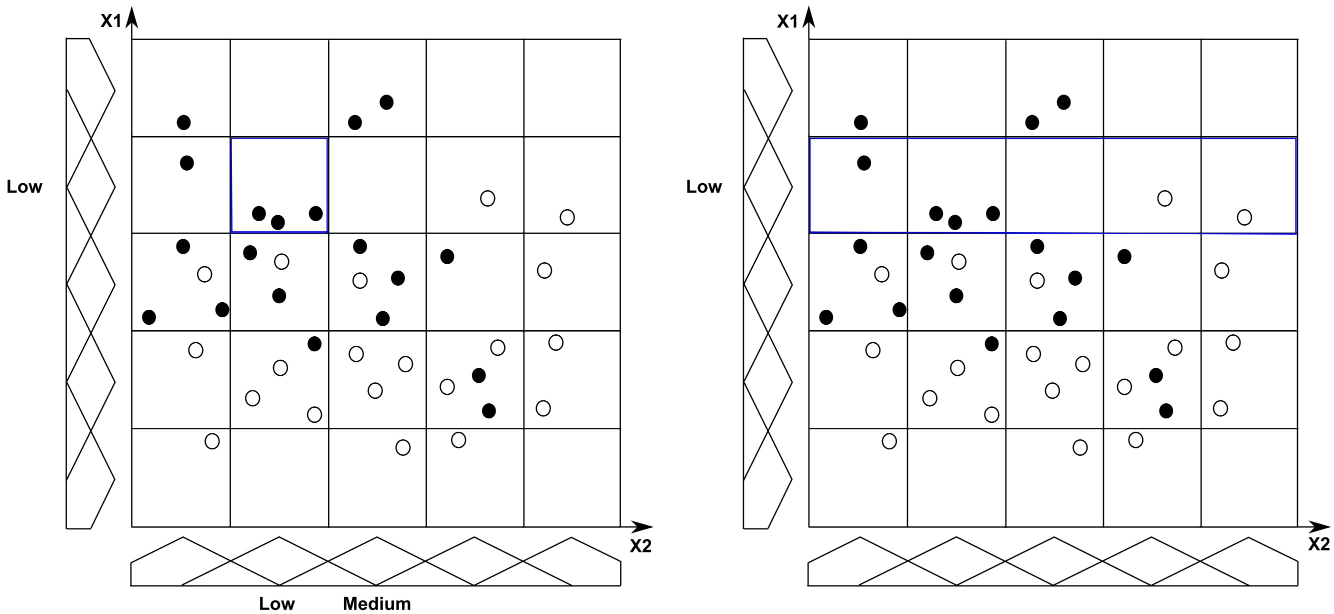
$$[\underline{\mathrm{SF}_k}, \overline{\mathrm{SF}_k}] = \left[\sum_{j=1, \mathrm{Class}(R_j)=k}^{L} \underline{\mathrm{CMD}_j}, \sum_{j=1, \mathrm{Class}(R_j)=k}^{L} \overline{\mathrm{CMD}_j}\right] \quad k = 1, \ldots, M. \tag{20}$$

3) 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 (7)] by the scaling factor of the corresponding class as follows:

$$\underline{\mathrm{SCMD}_j} = \wedge \left(\frac{\underline{\mathrm{CMD}_j}}{\underline{\mathrm{SF}_{\mathrm{Class}(R_j)}}}, \frac{\overline{\mathrm{CMD}_j}}{\overline{\mathrm{SF}_{\mathrm{Class}(R_j)}}}\right) \quad j = 1, \ldots, L \tag{21}$$

$$\overline{\mathrm{SCMD}_j} = \vee \left(\frac{\underline{\mathrm{CMD}_j}}{\underline{\mathrm{SF}_{\mathrm{Class}(R_j)}}}, \frac{\overline{\mathrm{CMD}_j}}{\overline{\mathrm{SF}_{\mathrm{Class}(R_j)}}}\right) \quad j = 1, \ldots, L. \tag{22}$$

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



(a) (b)

Fig. 4. Solution space covered by fuzzy rules. (a) Fuzzy rule using all the antecedents. (b) Fuzzy rule not using all the antecedents.

support and confidence values:

$$\left[\underline{\text{Support}_{\text{Scaled}}(A \rightarrow C_j)}, \overline{\text{Support}_{\text{Scaled}}(A \rightarrow C_j)}\right] = \left[\frac{\underline{\text{SCMD}_j}}{N}, \frac{\overline{\text{SCMD}_j}}{N}\right] \tag{23}$$

$$\underline{\text{Confidence}_{\text{Scaled}}(A \rightarrow C_j)} = \bigwedge\left(\frac{\underline{\text{SCMD}_j}}{\sum_{k=1}^{M} \underline{\text{SCMD}_k}}, \frac{\overline{\text{SCMD}_j}}{\sum_{k=1}^{M} \overline{\text{SCMD}_k}}\right) \tag{24}$$

$$\overline{\text{Confidence}_{\text{Scaled}}(A \rightarrow C_j)} = \bigvee\left(\frac{\underline{\text{SCMD}_j}}{\sum_{k=1}^{M} \underline{\text{SCMD}_k}}, \frac{\overline{\text{SCMD}_j}}{\sum_{k=1}^{M} \overline{\text{SCMD}_k}}\right). \tag{25}$$

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

$$\underline{\text{RW}_j} = \underline{\text{Support}_{\text{Scaled}}(A \rightarrow C_j)} \times \underline{\text{Confidence}_{\text{Scaled}}(A \rightarrow C_j)}, \quad j = 1, \ldots, L \tag{26}$$

$$\overline{\text{RW}_j} = \overline{\text{Support}_{\text{Scaled}}(A \rightarrow C_j)} \times \overline{\text{Confidence}_{\text{Scaled}}(A \rightarrow C_j)}, \quad j = 1, \ldots, L. \tag{27}$$

B. 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 RB 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, an RB 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. Fig. 4(a) 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 Fig. 4(b). 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 Fig. 4(a), 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 RB. 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} and 2) to build a default rule that always classifies uncovered examples in the majority class.

The first approach is an 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 GM, the achieved result will always be 0. This is due to the fact that the default rule correctly classifies all the examples of the majority class ($\text{TN}_{\text{rate}} = 1$), whereas it misclassifies all the examples of the minority class ($\text{TP}_{\text{rate}} = 0$).

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

1) *Generation of a set of fuzzy rules for the uncovered example:* 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 FSs {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}}.
2) *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 RB is obtained. With this aim, in first place, the fuzzy rules in the RB 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, (28) is applied in order to measure the similarity between fuzzy rules

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

where R'_j is the j'th rule in SFRU (x_p) with $j' = \{1, \ldots, |\text{SFRU}(x_p)|\}$ and $|\text{SFRU}(x_p)|$ being the number of rules in SFRU(x_p), R_j is the jth rule in the RB with $j = \{1, \ldots, L\}$, n is the number of variables of the problem, NL_k is the number of linguistic labels of the kth variable, and $V(\cdot)$ is the integer codification of the kth linguistic label of the rule being analyzed.

   At this point, each rule in SFRU(x_p) has a similarity with every rule in the RB. Finally, for each rule in SFRU(x_p), the rule in the RB having the greatest similarity is selected. Therefore, after this step, as many original rules in the RB 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.
3) *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 (29) and (30), respectively. More specifically, we use the rule weights of the original rules and the antecedents of the rules in SFRU(x_p).

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

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

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 $\underline{\text{RW}}_{j'}$ and $\overline{\text{RW}}_{j'}$ are the lower and upper rule weights of the most similar rule in the RB to the j'th rule in SFRU(x_p).
4) *Final prediction of the class:* The uncovered example x_p will be classified in the class having largest vote strength according to

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

C. 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 as follows.

1) *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. The following equation shows the structure of the whole chromosome:

$$C_{\text{Total}} = \{C_L, C_W, C_E, C_R\}. \tag{32}$$

   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:

a) Tuning of the lateral position of the linguistic labels (see Fig. 2)

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

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

b) Amplitude of the support of the upper bound of the IVFSs

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

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

c) Tuning of the equivalence: Genes take values in the range [0.01, 1.99]

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

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

d) Rule selection: This part is composed of as many genes as the number of rules in the RB. 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:

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

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

2) *Initialization of the population*: We initialize a chromosome representing the initial system, which is the chromosome that encodes the initial setup 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 RB). The remainder individuals are randomly initialized within the ranges demanded in each part.

3) *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

$$\text{Fitness} = \frac{\text{TP}_{\text{rate}} + \text{TN}_{\text{rate}}}{2}. \quad (37)$$

We must point out that this fitness function provides a similar behavior to that of the GM, 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 GM using a less computational demanding metric.

4) *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.

5) *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.

TABLE III
FEATURES OF THE FINANCIAL DATASETS USED IN THE EXPERIMENTAL STUDY

			Examples			
Dataset	#Attr.	#Cl.	#Ex.	% Cl. 0	% Cl. 1	IR
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

IV. EXPERIMENTS AND RESULTS

In this section, we present the experiments and results to validate our proposed system, which is denoted IVTURS$^{\text{RRW_I}}_{\text{FARC-HD}}$ for financial applications. The experimental framework used to show the quality of the new method is described in Section IV-A. The study has a double aim, on the one hand, the suitability of the novelties introduced in IVTURS$^{\text{RRW_I}}_{\text{FARC-HD}}$ needs to be justified (see Section IV-B), and on the other hand, the benefits of our new method against the use of SMOTE, which is a widely used preprocessing technique, and the cost-sensitive C4.5 decision tree are analyzed (see Section IV-C). In both scenarios, we present the evaluations carried over 11 different financial applications.

A. 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.

1) Financial Datasets Description: We have used 11 real-world datasets from various financial domains. The features of these datasets are summarized in Table III, 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 IV
NOTATIONS AND DESCRIPTIONS OF THE FIVE APPROACHES

Name	Linguistic Labels	Rule Weight	Tuned parameters
$\text{IVTURS}_{\text{FARC-HD}}$	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
$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$	IVFSs	Rule Weight Rescaling Method	α, a, b, W

The description of each financial dataset is as follows.

1) *Bank Investment (BI) Dataset* is used to predict whether to invest or not by buying or not buying a bank share in stock market.
2) *Bank Credit (BC) Dataset* 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 nondefaulting customers and bad customers are defaulting customers.
3) *Western Stock Index (WSI) Dataset* is used for predicting a major Western stock market composite index of whether the stock market index will increase/remain the same or decrease.
4) *Far Eastern Stock Index (FESI) Dataset* 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.
5) *Digital TV (DT) Dataset* is used for predicting whether to invest or no by buying or not buying a digital TV network share in stock market.
6) *Airline (AL) Dataset* is applied for predicting whether to invest or no by buying or not buying an airline share in stock market.
7) *Small Loan (SL) Dataset* 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.
8) *Arbitrage (Arb) Dataset* is used for spotting arbitrage opportunities in the London International Financial Futures Exchange (LIFFE) market [10]–[12]. The data reported in this paper were developed in [12] to identify arbitrage situations by analyzing option and futures prices in the LIFFE market.
9) *Fraud Detection (FD) Dataset* is used for fraud detection in personal loan application for small loan amounts.
10) *Lending (Len) Dataset* 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.
11) *Loan Authorization (LA) Dataset* is related to the prediction of good (profitable) or bad (nonprofitable) 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.

2) Used Configurations and Notations: In this paper, we compare our proposed compact evolutionary IV-FRBCS that we call $\mathbf{IVTURS}_{\mathbf{FARC\text{-}HD}}^{\mathbf{RRW\_I}}$, which is highlighted in gray in Table IV, with the following algorithms.

1) *$\text{IVTURS}_{\text{FARC-HD}}$:* 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.
2) *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.
3) *IVTURS_FS:* It is the type-1 fuzzy counterpart of $\text{IVTURS}_{\text{FARC-HD}}$. It also will be considered to analyze the goodness of the use of IVFSs.
4) *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.
5) *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.
6) *FURIA [27]:* It is a state-of-the art type-1 fuzzy approximative classifier.
7) *SMOTE [3]:* It is one of the most used preprocessing techniques. It will be used to check whether the use of $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ is competitive versus a state-of-the-art preprocessing 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 IV, 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 that are tuned in the optimization process.

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

TABLE V
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

introduced in Table V. For the genetic tuning, we have considered the following values for their parameters:

1) population size: 50 individuals;
2) number of evaluations: 20 000;
3) 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.

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 nonparametric 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 posthoc test [63] to find the algorithms that reject the equality hypothesis with respect to a selected control method.

The posthoc 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 dataset. Next, we compute the subtractions between the accuracy of each algorithm minus the average value for each dataset. Then, we rank all these differences in descending order, and finally, we average the rankings obtained by each algorithm. In this manner, the algorithm that achieves the lowest average ranking is the best one.

B. Studying the Effectiveness of the Novelties Introduced in $IVTURS^{RRW\_I}_{FARC\text{-}HD}$

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 $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ versus $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$ and the two classifiers using type-1 FSs (FARC-HD algorithm and IVTURS_FS). 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 VI contains the results obtained when applying these four classifiers both in training and in testing. These results are measured using the GM (see (13)), and they are grouped by two columns to show the results obtained in training (Tr.), in testing (Tst.). The best (highest) testing result is highlighted in **bold face**.

From the results shown in Table VI, it is observed that our proposed approach, highlighted in gray, provides clearly the best mean result (of GM) in testing (beating the competing techniques by a big margin followed by $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$), and it also reaches the best result in nine out of the 11 datasets. From these results, we can stress two facts: on the one hand, the use of IVFSs ($\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ and $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$) 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 $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ are better than those obtained when applying $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$. The results of $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ 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 datasets present in the vast majority of financial applications.

In order to support the superiority of $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$, we have applied the Friedman aligned ranks test. The obtained ranks are shown in the third column of Table VII, 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 posthoc test, whose obtained APVs are included in the last column of Table VII, to check whether $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$, 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 VII, we can conclude that our new approach is statistically better than FARC-HD, $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$, and IVTURS_FS, and consequently, the use of both IVFSs and the rescaling rule weight method allows handling the imbalanced financial datasets to give a superior performance.

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 10, 20, 30, and 40 fuzzy rules having a largest rule weight from each class, which implies obtaining RBs composed of 20, 40, 60, and 80 rules, respectively. The obtained results for the FD financial problem are presented in Table VIII, where each row shows the number of rules in the RB, the number of uncovered examples (using the RB composed of as many rules as indicated in the first column) along with the accuracy achieved over the uncovered examples for each class ($\mathrm{TP}_{\mathrm{rate}}$ and $\mathrm{TN}_{\mathrm{rate}}$), and the result of the GM. It can be observed that the use of the similarity technique is beneficial for the system

TABLE VI
RESULTS OBTAINED IN TRAINING (TR.) AND TESTING (TST.) MEASURED USING THE GM

	$\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$		FARCHD		$\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$		IVTURS_FS		
Dataset	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.	IR
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	

TABLE VII
HOLM'S TEST TO COMPARE $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ VERSUS FARC-HD, $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$, AND IVTURS_FS

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

$\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ is used as a control method.

TABLE VIII
RESULTS OBTAINED WITH THE SIMILARITY TECHNIQUE ON THE FD DATASET USING THE $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ METHOD

Number of rules	Number of uncovered examples	$\mathrm{TP}_{\mathrm{rate}}$	$\mathrm{TN}_{\mathrm{rate}}$	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

since, when using the default rule, the result of the GM is always 0. Furthermore, it is noticed that the larger the number of fuzzy rules in the RB, 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 GM.

C. *Analyzing the Suitability of $IVTURS^{RRW\_I}_{FARC\text{-}HD}$ 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 preprocessed data by means of SMOTE, which is one of the most widely used preprocessing techniques. To do so, this study is divided into 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 IX 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 GM. The best testing result for each dataset is highlighted in **bold face**.

From results in Table IX, it is shown that the best mean testing result of GM is provided by our new proposal. More specifically, we can observe that the results of the other method using IVFSs ($\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$+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 $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ was the best method since it is the best ranked one (see the third column of Table X). Next, we have carried out the Holm's posthoc test to determine if $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ is statistically better than the remainder methods. The results obtained in the statistical study are shown in Table X, 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 $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$ algorithms using SMOTE (FARCHD+SMOTE and IVTURS_FS+SMOTE). Finally, it is noteworthy that there are not statistical differences with respect to $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$+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 XI shows the results provided by these four classifiers both in Training (Tr.) and testing (Tst.). It is noteworthy the

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

SAMPLING	NO		SMOTE						IR
Dataset	**$\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$**		FARCHD		IVTURS_FS		$\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$		
	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 X
HOLM'S TEST TO COMPARE $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ VERSUS THE APPROACHES BASED ON FUZZY ASSOCIATION RULES CLASSIFIERS

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

$\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ is used as a control method.

notable improvement achieved by $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ over the mean of the testing data, since it improves in 4%, 7%, and 12% the results obtained by C45_CS, C45+SMOTE, and the original C4.5 decision tree, respectively.

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 XIII, where it can be observed that $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ 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 XIV) confirm that our new approach outperforms the original FURIA, but there are not statistical differences between $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ 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 preprocessed data in the learning stage, whereas our proposal uses the original data, and consequently, it keeps the original distribution of the classes.

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 XII, where it can be observed that $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ 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 the performance based on the achievement of better results in seven out of the 11 financial problems, which manifest a superior behavior.

Regarding the interpretability of the model, Table XV 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 preprocessed 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 $\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$) using SMOTE generate 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 $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$. Regarding the approaches based on the C4.5 decision tree, their average number of created rules is ranged between 160 and 735, that is, $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ 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 $\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$ is producing a more interpretable model than the other competing methods that used SMOTE.

Table XVI shows the set of small number of rules and short rule lengths generated by our proposed method ($\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$) for the Arbitrage dataset, where it can be seen that only nine rules were generated, and they have a short

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

Dataset	$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$		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 XII
HOLM'S TEST TO COMPARE $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ VERSUS THE APPROACHES BASED ON THE C4.5 DECISION TREE

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	$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$	13.18	-

$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ is used as a control method.

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

Dataset	$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$		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	

length, which enable high degree of transparency and interpretability for the financial adviser while producing the best GM when compared with the C4.5 decision tree, type-1, and IVFS-based methods as shown in Table IX with no need to use any preprocessing.

V. CONCLUSION 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 $\text{IVTURS}_{FARC\text{-}HD}$, and

TABLE XIV
HOLM'S TEST TO COMPARE $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ VERSUS THE APPROACHES BASED ON FURIA

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

$\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ is used as a control method.

it is named $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$. 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; therefore, 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 $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ has been tested in 11 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 RB; 3) $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ achieves a better 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 preprocessing technique; 4) our new approach enhances the results provided by the cost-sensitive C4.5 decision tree; and 5) $\text{IVTURS}_{\text{FARC-HD}}^{\text{RRW\_I}}$ provides competitive results versus FURIA applied after applying SMOTE, while our system has the advantages of avoiding the use of preprocessing techniques and obtaining more interpretable and transparent models for the financial analysts.

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

SAMPLING	NO				SMOTE					IR
Dataset	$\mathrm{IVTURS}^{\mathrm{RRW\_I}}_{\mathrm{FARC\text{-}HD}}$	C4.5	C45_CS	FURIA	FARCHD	IVTURS_FS	$\mathrm{IVTURS}_{\mathrm{FARC\text{-}HD}}$	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 XVI
FUZZY RB GENERATED FOR THE ARBITRAGE DATASET

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]

For our future work, we will aim to investigate the use of type-2 FSs and the use of multiobjective multiconstraint 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 multiclass 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," *Intell. Data Anal.*, 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 Explor. Newslett.*, 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," *J. Artif. Intell. Res.*, 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 *Proc. 8th Int. Conf. Signal Process.*, vol. 3, 2006, pp. 1815–1818.

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

[6] H. S. Kim and S. Y. Sohn, "Support vector machines for default prediction of SMEs based on technology credit," *Eur. J. Oper. Res.*, vol. 201, no. 3, pp. 838–846, 2010.

[7] E. Giacomini, "Neural networks in quantitative finance," Master thesis, Humboldt-Universität zu Berlin, Berlin, Germany, Inst. Statist. Econometr., 2003.

[8] R. Lawrence, "Using neural networks to forecast stock market prices," Dept. Comput. Sci., Univ. Manitoba, Winnipeg, MB. Canada, Dec. 12, 1997.

[9] K. Kwong, "Financial forecasting using neural network or machine learning techniques," Elect. Eng. Thesis Project, Univ. Queensland, Sch. Inform. Technol. Elect. Eng., Oct. 2001.

[10] A. Garcia-Almanza and E. Tsang, "Forecasting stock prices using genetic programming and chance discovery," in *Proc. 12th Int. Conf. Comput. Economics Finance*, 2006.

[11] A. Garcia-Almanza and E. Tsang, "Evolving decision rules to predict investment opportunities," *Int. J. Autom. Comput.*, vol. 5, no. 1, pp. 22–31, 2008.

[12] A. Garcia-Almanza, "New classification methods for gathering patterns in the context of genetic programming," Ph.D. dissertation, Dept. Comput. Electron. Syst., Univ. Essex, Colchester, U.K., 2008.

[13] R. Nogueira, S. M. Vieira, and J. M. C. Sousa, "The prediction of bankruptcy using fuzzy classifiers," in *Proc. ICSC Congr. Comput. Intell. Methods Appl.*, 2005, pp. 95–100.

[14] V. Snášel, P. Krömer, J. Platoš, and A. Abraham, "The evolution of fuzzy classifier for data mining with applications," in *Proc. 8th Int. Conf. Simulated Evol. Learn.*, 2010, pp. 349–358.

[15] Z. Yang and G. Gan, "Application of fuzzy classification in bankruptcy prediction," in *Proc. 4th Int. Conf. Intell. Comput.: Adv. Intell. Comput. Theories Appl.—Aspects Theor. Methodol. Issues*, 2008, pp. 921–928.

[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, and J. Lawry, Eds. Berlin, Germany: Springer-Verlag, 2004, pp. 50–66.

[17] L. Angstenberger, *Dynamic Fuzzy Pattern Recognition With Applications to Finance and Engineering*. Norwell, MA, USA: Kluwer, 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" in *Proc. 12th U.K. Workshop Comput. Intell.*, 2012, pp. 1–7.

[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," *Inf. Sci.*, 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," *Int. J. Approx. Reason.*, vol. 52, no. 6, pp. 751–766, 2011.

[21] R. Sambuc, "Function Φ-flous, application a l'aide au diagnostic en pathologie thyroidienne," These de Doctorat en Medicine, Univ. Marseille, Marseille, France, 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 Trans. Fuzzy Syst.*, vol. 21, no. 3, pp. 399–411, Jun. 2013.

[23] M. Galar, J. Fernandez, G. Beliakov, and H. Bustince, "Interval-valued fuzzy sets applied to stereo matching of color images," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1949–1961, Jul. 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 *Proc. Joint Int. Fuzzy Syst. Assoc. World Congr. Eur. Soc. Fuzzy Logic Technol. Conf.*, 2009, pp. 831–836.

[25] J. Quinlan, "C4.5: Programs for machine learning," San Mateo, CA, USA: Morgan Kaufmann, 1993.

[26] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 3, pp. 659–665, May/Jun. 2002.

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

[28] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[29] H. Wong and B. Q. Hu, "Application of interval clustering approach to water quality evaluation," *J. Hydrol.*, 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," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, 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 Syst.*, vol. 21, pp. 1–17, 1987.

[32] I. B. Turksen, "Interval valued fuzzy sets based on normal forms, " *Fuzzy Sets Syst.*, 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 Kα operators: Laws for conjunctions and disjunctions, amplitude," *Int. J. Intell. Syst.*, 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 Trans. Fuzzy Syst.*, vol. 21, no. 6, pp. 1150–1162, Dec. 2013.

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

[36] Z. S. Xu and R. R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," *Int. J. General Syst.*, 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 Syst.*, 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 Trans. Fuzzy Syst.*, vol. 12, no. 1, pp. 45–61, Feb. 2004.

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

[40] Q. Yang and X. Wu, "10 challenging problems in data mining research," *Int. J. Inf. Technol. 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," *Inf. Sci.*, 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," *Mach. Learn.*, 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 Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, Oct. 2011.

[45] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd Ed. San Francisco, CA, USA: Morgan Kaufmann, 2006.

[46] C. Zhang and S. Zhang, "Association rule mining: Models and algorithms series," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002.

[47] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.

[48] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 428–435, Aug. 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 Trans. Fuzzy Syst.*, vol. 15, no. 4, pp. 616–635, Aug. 2007.

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

[51] O. Cordón, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *Int. J. Approx. Reason.*, 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 Comput.*, 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 Syst.*, vol. 141, no. 1, pp. 5–31, 2004.

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

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

[56] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms, " *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.

[57] F. Herrera, M. Lozano, and A. M. Sánchez, "A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study," *Int. J. Intell. Syst.*, vol. 18, no. 3, pp. 309–338, 2003.

[58] L. J. Eshelman and J. Schafffer, "Real coded genetic algorithms and interval schemata," in *Foundations of Genetic Algorithms*. San Mateo, CA, USA: Morgan Kaufman, 1993, pp. 187–202.

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

[60] J. Demar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, 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," *Inf. Sci.*, 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," *Ann. Math. Statist.*, vol. 33, pp. 482–497, 1962.

[63] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian J. Statist.*, 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 Recog.*, 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," *Knowl.-Based Syst.*, vol. 42, pp. 97–110, 2013.

[66] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, 2014.

José Antonio Sanz received the M.Sc. degree in computer sciences in 2008 and the Ph.D. degree in 2011, both from the Public University of Navarra, Pamplona, Spain.

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 in international journals and is involved in teaching artificial intelligence for students of computer sciences. 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 a Member of the European Society for Fuzzy Logic and Technology. He received the Best Paper Award at the FLINS 2012 International Conference and the Pepe Millá Award in 2014.

Dario Bernardo received the B.Sc. and M.Sc. degrees in computer engineering for enterprise management from the University of Pisa, Pisa, Italy, and another M.Sc. degree in computational and software techniques in engineering from Cranfield University, Cranfield, U.K. He is currently pursuing the Ph.D. degree in computer science with the University of Essex, Colchester, U.K., during which he has been a Knowledge Partnership Associate.

He is currently a Data Scientist working in London, U.K., in the field of predictive analytics in finance. His major research interests include 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 (M'10) received the M.Sc. and Ph.D. degrees in mathematics from the University of Granada, Granada, Spain, in 1988 and 1991, respectively.

He is currently a Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. He has been the Supervisor of 34 Ph.D. students. He has published more than 270 papers in international journals. 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.

Dr. Herrera is the Editor-in-Chief of *Information Fusion* (Elsevier) and *Progress in Artificial Intelligence* (Springer). He acts as an Editorial Board Member of a dozen journals, including the *International Journal of Computational Intelligence Systems*, the 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 has received the following honors and awards: ECCAI Fellow in 2009, IFSA Fellow in 2013, the 2010 Spanish National Award on Computer Science ARITMEL to the "Spanish Engineer on Computer Science," the International Cajastur "Mamdani" Prize for Soft Computing (Fourth Edition, 2010), the IEEE TRANSACTIONS ON FUZZY SYSTEM Outstanding 2008 Paper Award (bestowed in 2011), the 2011 Lotfi A. Zadeh Prize Best Paper Award from the International Fuzzy Systems Association, and the 2013 AEPIA Award for his scientific career in Artificial Intelligence.

Humberto Bustince (M'06) received the B.Sc. degree in physics from Salamanca University, Salamanca, Spain, in 1983 and the Ph.D. degree in mathematics from the Public University of Navarra, Pamplona, Spain, in 1994.

He has been a Teacher with the Public University of Navarra since 1991, where he is currently a Full Professor with the Department of Automatics and Computation. He served as the Subdirector of the Technical School for Industrial Engineering and Telecommunications from 2003 to 2008, and he was involved in the implantation of Computer Science courses with the Public University of Navarra. He is currently involved in teaching artificial intelligence for students of computer sciences. He 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.

Dr. Bustince is a 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 the Editor-in-Chief of the *Mathware & Soft Computing Magazine* and *Notes on Intuitionistic Fuzzy Sets*. He is also the Guest Editor of the *Fuzzy Sets and Systems* and a 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 the B.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 currently a Professor with the School of Computer Science and Electronic Engineering, Director of the Computational Intelligence Centre, and the Head of the Fuzzy Systems Research Group with the University of Essex. His major research interests include 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.

Dr. Hagras is a Fellow of the Institution of Engineering and Technology. He was the Chair of IEEE Computational Intelligence Society (CIS) Senior Members' Subcommittee. He received numerous prestigious international awards for his research, where most recently, he received, from the IEEE CIS, the 2013 Outstanding Paper Award and the 2006 Outstanding Paper Award, both for 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 CIS Fuzzy Systems Technical Committee. He has chaired several international conferences. He served as the General Co-Chair of the 2007 IEEE International Conference on Fuzzy Systems, London, U.K., July 2007; the Chair of the 2009, 2011, 2013, and 2014 IEEE Symposium on Intelligent Agents; and the Chair of 2011 IEEE Symposium on Advanced to Type-2 Fuzzy Logic Systems, Paris, France, in 2011.