Information Sciences 577 (2021) 265–279

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A fuzzy association rule-based classifier for imbalanced classification problems

J. Sanz *, M. Sesma-Sara, H. Bustince

Department of Statistics, Computer Science and Mathematics, Universidad Publica de Navarra, Campus Arrosadia s/n, Pamplona, Navarra 31006, Spain
Institute of Smart Cities, Universidad Publica de Navarra, Campus Arrosadia s/n, Pamplona, Navarra 31006, Spain

ARTICLE INFO

Article history:
Received 27 November 2020
Received in revised form 30 June 2021
Accepted 3 July 2021
Available online 6 July 2021

Keywords:
Fuzzy association rule-based classifier
Imbalanced classification problems
Lift
Averaging aggregation functions

ABSTRACT

Imbalanced classification problems are attracting the attention of the research community because they are prevalent in real-world problems and they impose extra difficulties for learning methods. Fuzzy rule-based classification systems have been applied to cope with these problems, mostly together with sampling techniques. In this paper, we define a new fuzzy association rule-based classifier, named FARCI, to tackle directly imbalanced classification problems. Our new proposal belongs to the algorithm modification category, since it is constructed on the basis of the state-of-the-art fuzzy classifier FARC–HD. Specifically, we modify its three learning stages, aiming at boosting the number of fuzzy rules of the minority class as well as simplifying them and, for the sake of handling unequal fuzzy rule lengths, we also change the matching degree computation, which is a key step of the inference process and it is also involved in the learning process. In the experimental study, we analyze the effectiveness of each one of the new components in terms of performance, $F - score$, and rule base size. Moreover, we also show the superiority of the new method when compared versus FARC–HD alongside sampling techniques, another algorithm modification approach, two cost-sensitive methods and an ensemble.

1. Introduction

Classification problems have been widely tackled using fuzzy techniques [1,2]. In the last years, imbalance classification problems [3,4] have gained attention from the research community, since they are common in real-world problems [5,6]. There exist multi-class imbalanced classification problems [7] and binary ones, where the examples of one class (known as majority or negative class) outnumber those of the other one (known as minority or positive class), which entails difficulties for the learning of the classifiers [8]. In this paper, we focus on binary imbalanced classification problems as they are frequent in real-world problems.

Techniques usually applied to tackle this problem can be grouped in four categories: 1) data preprocessing methods [9,10], which are independent of the classifier because they are focused on sampling the dataset, either to balance it or easing the subsequent learning of the classifier; 2) algorithm modification methods [11–13], which consist in modifying existing algorithms or creating new ones to deal directly with imbalance problems; 3) cost-sensitive methods [14–16], where a larger cost is assigned during the learning process to the misclassified examples of the minority class to enhance the classifier

* Corresponding author at: Department of Statistics, Computer Science and Mathematics, Universidad Publica de Navarra, Campus Arrosadia s/n, Pamplona, Navarra 31006, Spain.

E-mail addresses: joseantonio.sanz@unavarra.es (J. Sanz), mikel.sesma@unavarra.es (M. Sesma-Sara), bustince@unavarra.es (H. Bustince).

https://doi.org/10.1016/j.ins.2021.07.019
0020-0255/© 2021 The Authors. Published by Elsevier Inc.

performance and 4) ensemble approaches [17,18], which are composed of several classifiers usually combined with sampling techniques to improve the performance of individual classifiers.

Fuzzy Rule-Based Classification Systems (FRBCSs) [19] are a type of fuzzy classifiers that have been widely used to cope with classification problems [20]. FRBCSs usually obtain accurate results and interpretable models as a result of using linguistic terms in the antecedent of IF-THEN rules. A state-of-the-art FRBCS is the Fuzzy Association Rule-based Classification model for High-Dimensional problems (FARC–HD) [21], which has been used as the basis of many proposals to cope with classification problems [22,23]. FARC–HD, like all the FRBCSs, can deal with uncertainty, ambiguity or vagueness in a very effective way. These properties are interesting when tackling imbalanced problems, as uncertainty is inherent to them and, consequently, FRBCSs have shown to perform well when they are combined with preprocessing methods [24] or without them [25].

The aim of this paper is to design a new FRBCS that is able to deal directly with imbalanced classification problems. We develop modifications in all the stages of FARC–HD and, therefore, our new method is embraced in the algorithm modification category. We name our proposal FARCI as it is a Fuzzy Association Rule-based Classifier for Imbalanced classification problems. Specifically, the novelties of FARCI are:

- The fuzzy association rule learning is no longer driven by the confidence [26] but the lift [27] to generate only fuzzy rules with a positive relationship between the antecedent and the consequent.
- The philosophy of the usage of the pattern weighting scheme applied to select the most interesting fuzzy rules [28] as well as the initialization of the hyper-parameters involved in this process are changed. As a consequence, the number of the generated fuzzy rules for the minority class is boosted.
- The fitness function used in the evolutionary process is modified so that it uses a proper metric to measure the performance in imbalanced scenarios.
- The computation of the matching degree is carried out by appropriate operations so that they are not affected by the unequal number of antecedents on the fuzzy rules learned from different classes.

All these new components are aimed at making FARCI obtain accurate results without using preprocessing techniques, which could lead to under or over fitting problems [9]. This enhancement is possible because FARCI is designed so that the fuzzy rule base is better suited for these problems. Specifically, the number of fuzzy rules of the minority class is enlarged and, additionally, the number of antecedents that compose the rules is reduced, making the fuzzy rules simpler.

To support the quality of our new method, we conduct an experimental study where: 1) we analyze the influence of each one of the new components of FARCI in terms of performance and rule base size; 2) we study whether our new method, when applying all the components, is able to improve the results of FARC–HD when it is used alongside preprocessing approaches and 3) we compare the results of FARCI versus those obtained by GP-COACH-H [12] as a representative of algorithm modification methods, C45CS [15] and C-SVMCS [14] as representatives of cost-sensitive methods and EUSBoost [18] as a representative of ensemble approaches. To do so, we consider 66 imbalanced datasets selected from the KEEL dataset repository [29], we measure the performance of the methods by means of the $F-score$ and we conduct an appropriate statistical study as suggested in the specialized literature [30].

The remainder of the paper is organized as follows: in Section 2 we recall some preliminary concepts related to both imbalanced classification problems and FRBCSs and, then, we describe in detail our new method in Section 3. The experimental framework and the obtained results with an analysis are presented in Sections 4 and 5, respectively. The main conclusions are drawn in Section 6.

2. Preliminaries

In this section, we first recall some concepts about imbalanced classification problems (Section 2.1) and then we describe FRBCSs focusing on the FARC–HD algorithm [21], since it is the basis of our new method (Section 2.2).

2.1. Imbalanced classification problems

A classification problem consists in learning a mapping function named *classifier* from a *training set*, which is a set of P training examples, (x_p, y_p), where $p \in \{1, \ldots, P\}$. Each example x_p is composed of n attributes, $(x_{p1}, \ldots, x_{pn})$ being x_{pi} the value of the i-th attribute ($i \in \{1, 2, \ldots, n\}$), and it belongs to a class $y_p \in \mathbb{C} = \{1, 2, \ldots, m\}$, where m coincides with the number of classes of the problem. The learned classifier allows to classify previously unknown examples.

When the number of examples belonging to each of the classes is different, the classification problem suffers from the problem of data imbalance [3,4]. Learning from this kind of problem has been identified as one of the main challenges in data mining [8]. Binary imbalanced classification problems are very common in real-world applications [6], where one of the classes is represented by a large number of examples (known as majority or negative class) and the other one is represented by only a few ones (known as minority or positive class). The Imbalance Ratio (IR) is equal to the number of examples of the majority class divided by the number of examples of the minority class and it characterizes the imbalance degree of the problem. In this scenario, classifiers tend to predict the examples as majority class, ignoring the minority class. However, the IR is not the only difficulty that should be taken into account during learning. These problems are also characterized by small disjuncts [31], overlapping among classes and other problems that may provoke a degradation of the system's performance [8]. To cope with them, four different strategies are usually applied: 1) to preprocess the data using sampling methods [9], which have been successfully applied to boost the performance of FRBCSs [24]; 2) to modify existing classifiers so that they can internally take into account this problem [11,32], which has also applied to rule-based classifiers [13,15]; 3) to consider cost-sensitive solutions [33] which were used to improve FRBCSs in imbalanced big data problems [16] and 4) to use ensembles of classifiers designed for imbalanced scenarios [17], where the base classifier is usually a rule based one like a decision tree.

When dealing with imbalanced datasets, a key point is the choice of an appropriate metric to measure the performance of the classifiers, since the accuracy rate may lead to obtaining erroneous conclusions [34]. Consequently, different metrics, derived from a confusion matrix (see Table 1), are usually applied in this framework, such as the geometric mean (Eq. (1)), the balanced accuracy (Eq. (2)) or the $F-score$ (Eq. (3)).

$$GM = \sqrt{TP_{rate} \cdot TN_{rate}} \tag{1}$$

$$Bal_{acc} = \frac{TP_{rate} + TN_{rate}}{2} \tag{2}$$

$$F-score = 2 \cdot \frac{Precision \cdot TP_{rate}}{Precision + TP_{rate}}, \tag{3}$$

where TP_{rate} and TN_{rate} are the percentages of correctly classified examples belonging to the minority and majority classes, respectively, and *Precision* is the percentage of correctly classified examples of the minority class from those predicted as the minority class.

Both *AUC* and *GM* are common criteria to measure the performance of a classifier in a data imbalance framework. However, in this paper we consider the usage of the $F-score$ as the performance metric as it does not consider the TN_{rate} and, consequently, it is less sensitive to the IR [35].

2.2. Fuzzy rule-based classification systems and FARC–HD

Among the many existing methods to tackle classification problems, FRBCSs are widely used because they provide accurate results as well as an interpretable model [19]. Specifically, this interpretability is obtained from the usage of linguistic terms in the antecedents of their rules as shown in Eq. (4).

$$\text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j \tag{4}$$

where R_j is the label of the j-th rule, $x = (x_1, \dots, x_n)$ is an n-dimensional pattern vector representing the example, A_{ji} is a fuzzy set, $C_j \in \mathbb{C}$ is the class label and RW_j is the rule weight [26].

FARC–HD [21] is currently one of the most accurate and interpretable FRBCSs in the literature. Our new method is constructed on its basis and, consequently, we remind both its three learning stages and its inference process.

The first stage of the learning consists in applying the Apriori algorithm [36] to learn the initial fuzzy rules. In this stage, triangular shaped membership functions are used to model the linguistic labels, which are obtained performing an homogeneous partition of the input space of each input attribute. In this case, each item is a linguistic label and the support and confidence[1] are computed using Eqs. (5) and (6), respectively:

$$Supp(A_j) = \frac{1}{P}\sum_{p=1}^{P}\mu_{A_j}(x_p), \tag{5}$$

$$Conf(A_j \rightarrow C_j) = \frac{\sum_{x_p \in ClassC_j}\mu_{A_j}(x_p)}{\sum_{p=1}^{P}\mu_{A_j}(x_p)}, \tag{6}$$

where $\mu_{A_j}(x_p)$ is the matching degree of the example x_p to the antecedent part of the fuzzy rule R_j, which is computed using Eq. (7):

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \dots, \mu_{A_{jn_j}}(x_{pn_j})), \tag{7}$$

where A_j is the antecedent of the fuzzy rule R_j, $\mu_{A_{ji}}(x_{pi})$ is the membership degree of the example with the i-th antecedent of the rule R_j, T is a T-norm and n_j is the number of antecedents of the rule.

In this stage, it is necessary to specify the minimum support, *MinSupp*, and confidence, *MinConf*. Specifically, the minimum support is weighted by the distribution of the classes of the problem. In this manner, the minimum support value

[1] We must point out that the support and the confidence are computed for the items included in A_i and the fuzzy rule $A_j \rightarrow C_j$, respectively.

Table 1
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)

can be different for each class: $MinSupp_{C_j} = MinSupp \cdot f_{C_j}$, where f_{C_j} is the proportion of examples belonging to class C_j in the dataset. Consequently, it is suitable to deal with imbalanced classification problems.

The learning process of the original Apriori algorithm is adapted to deal with classification problems. Specifically, the Apriori algorithm is applied as many times as the number of classes. For a specific class, the Apriori algorithm obtains a set of frequent itemsets and we generate as many fuzzy rules as frequent itemsets by setting: 1) the antecedent to the linguistic labels composing the itemset; 2) the consequent to the class under study and 3) the rule weight to the confidence of the generated fuzzy rule. The process to obtain the itemsets for each class starts by initializing a search tree with an empty node. Then, it creates the first level of the tree by listing all the possible itemsets composed of 1 item (nodes of the tree). For each itemset the support of the itemset and the confidence of the fuzzy rule that would be generated are computed. An itemset is called frequent itemset when its support is larger than *MinSupp*. Otherwise, the itemset does not need to be further extended (its node) due to the anti-monotone property of the support. When the confidence is larger than *MinConf*, it does not need to be extended as it has reached a satisfactory level of quality. In order to create the second (and subsequent) level(s) of the tree, this method combines the itemset in a node with those of nodes that can be further extended as long as they can be combined.[2] Note that the depth of the tree is limited (parameter *MaxDepth*) to obtain short rules that are easy to understand for the users.

After applying the Apriori algorithm, a lot of fuzzy rules are created. Consequently, the second stage of FARC–HD consists in selecting the most interesting rules by means of the application of a pattern weighting scheme [28]. Specifically, the quality of each fuzzy rule, $R_j : A_j \rightarrow C_j$, is measured using (8), which considers both the weights of the examples, w_p, and the matching degrees of the examples with the fuzzy rule. The larger the value of this equation the better the quality of the fuzzy rule. We have to point out that this scheme of weighted rules with moving example cover has also been used in other works like [37].

$$wWRAcc''(R_j) = \frac{n''(A_j \cdot C_j)}{n'(C_j)} \cdot \left(\frac{n''(A_j \cdot C_j)}{n''(A_j)} - \frac{n(C_j)}{P} \right), \tag{8}$$

where $n(C_j)$ is the number of examples of class C_j and

$$\begin{aligned} n''(A_j) &= \sum_{p=1}^{P} w_p \cdot \mu_{A_j}(x_p), \\ n''(A_j \cdot C_j) &= \sum_{x_p \in ClassC_j} w_p \cdot \mu_{A_j}(x_p), \\ n'(C_j) &= \sum_{x_p \in ClassC_j} w_p. \end{aligned}$$

The key point of this method is the assignment of the weights. Initially, the weight of each example is set to 1. Then, the best rule according to (8) is selected and the weights of the examples covered by that rule are decreased according to the formula $w_p = \frac{1}{c_p}$, where $p \in \{1, \ldots, P\}$ and c_p represents the number of times the p-th example has been covered by any of the already selected rules. When an example has been covered k_t times its weight is set to 0 (it is no longer considered). Obviously, the selected fuzzy rule cannot be selected anymore. This process is repeated until one of the stopping criteria is fulfilled: 1) all the examples have been covered more than k_t times or 2) all the fuzzy rules have been selected. This weighting scheme implies that examples with larger weights have a greater chance of being covered by the next fuzzy rules and, consequently, difficult examples are also covered as the iterative process runs.

This second stage, like the Apriori algorithm, is repeated for each class. That is, first the Apriori algorithm learns fuzzy rules for a specific class and then the most interesting ones are selected according to this procedure. In this manner, the second stage does not take into account fuzzy rules from the remaining classes.

In the third stage, once fuzzy rules from all the classes are learned and selected, an evolutionary process is carried out in order to optimize the lateral position of the membership functions [38] and to perform a rule selection process. The synergy between tuning and rule selection enables to contextualize the membership functions to the problem and to obtain a compact fuzzy rule set with a high cooperation degree among them.[3]. Eq. (9) is used as the fitness function to measure the quality of each chromosome (solution), C.

[2] Two itemsets cannot be combined when they share items (linguistic labels) belonging to the same attribute. Consequently, this process avoids the generation of fuzzy rules where an attribute is used more than once.

[3] Fuzzy rules that cooperate properly among them when used (combined) in the inference process are able to obtain a good classification performance.

$$Fitness(C) = \frac{\#Hits}{P} - \delta \cdot \frac{NR_{initial}}{NR_{initial} - NR + 1}, \tag{9}$$

where $\frac{\#Hits}{P}$ is the accuracy rate, $NR_{initial}$ is the number of fuzzy rules before applying the evolutionary process and NR is the number of rules selected by solution C.

Finally, to classify a new example x_p, FARC–HD applies the Fuzzy Reasoning Method (FRM) [39] known as additive combination, which uses the information of all the fired fuzzy rules to predict the class of the example: First, the total vote strength for each class is computed (Eq. (10)) and, then, the example x_p is classified in the class yielding the largest value.

$$V_{Class_k}(x_p) = \underset{R_j \in RB \text{ and } C_j = k}{f} \mu_{A_j}(x_p) \cdot RW_j \tag{10}$$

where $j \in \{1, \ldots, L\}$, L is the number of fuzzy rules in the fuzzy rule base and f is an aggregation function like the maximum or the normalized sum, which lead to the winning rule or the additive combination FRMs [39], respectively. The Choquet integral and its generalizations [22] can be also used as the aggregation function f.

3. A fuzzy association rule-based classifier for imbalanced classification problems

In this section we describe in detail our new method, FARCI, which consists in a modification of all the stages of FARC–HD [21] to tackle imbalanced classification problems. First, we present the change in the matching degree computation (Section 3.1), then we explain the modifications in the fuzzy rule generation process (Section 3.2) and in the selection of the generated fuzzy rules (Section 3.3). Finally, we describe the change in the fitness function for the evolutionary process (Section 3.4).

3.1. Matching degree computation

The matching degree between an example and the antecedent of a fuzzy rule (Eq. (7)) is used in several components of the method, namely, in the computation of the support and confidence in the fuzzy rule learning process, in the equation used to measure the quality of fuzzy rules (Eq. (8)) as well as in the FRM of the method. Consequently, it plays a key role for the success of this method.

However, the product is considered as the T-norm in FARC–HD, which may not be the best choice for imbalanced problems as it penalizes fuzzy rules having a larger number of antecedents. This is a key point because a specific feature of imbalanced classification problems is that examples of the minority class are usually gathered in small groups (known as small disjuncts), which usually implies that the length of fuzzy rules of the minority class is larger than that of those of the majority class. For example, we can think of two fuzzy rules composed by one (fuzzy rule A) and three (fuzzy rule B) linguistic terms in their antecedents, respectively. If the membership degree of an example to the antecedent of fuzzy rule A is 0.5 and the membership degrees of that example to the three antecedents of fuzzy rule B are all 0.75, the resulting matching degrees are 0.5 and 0.42 for fuzzy rules A and B, respectively. Consequently, one may be mislead to think that fuzzy rule A is more suitable, when B is a better option.

To deal with this problem, in the last years, there have been contributions where the conjunction among fuzzy sets in FRBCSs is modeled by means of different operators like overlap functions [40] and their usage even improves performance of the system when they have an averaging behavior [23,41]. For this reason, we propose the usage of averaging aggregation functions instead of T-norms to compute the matching degree. Specifically, we consider the usage of: 1) the geometric mean, as a representative of n-dimensional overlap functions and 2) the arithmetic mean, since it is the classical mean operator, with a modification: if any of the input membership degrees is 0, we set the result of the matching degree to 0 as well, since it means that this example does not match all the antecedents of the fuzzy rule.

3.2. Generation of fuzzy association rules

The criteria for evaluating the quality of the fuzzy association rules by the Apriori algorithm, the support and confidence (Eqs. (5)), have some well-known problems [27]. On the one hand, the support may discard useful fuzzy rules derived from low support items. On the other hand, the confidence may generate spurious fuzzy rules and reject interesting ones as a result of not taking into account the class distribution of examples. It is obvious that low support items would correspond to those from minority class examples and we hypothesise that interesting fuzzy association rules from the minority class may be rejected (or becoming complicated by adding more antecedents) because of the confidence.

The original FARC–HD method already takes into account the problem derived from the support by using a different *MinSupp* threshold for each class, which depends on the class distribution. However, it does not address problems derived from the confidence and, to cope with them, we propose the usage of the lift [27] (Eq. (11)), which is also known as interest, adapted to work in a fuzzy setting.

$$Lift(A_j \to C_j) = \frac{\sum_{x_p \in ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{P} \mu_{A_j}(x_p) \cdot f_{C_j}} \tag{11}$$

From an statistical point of view, the lift measures the deviance of the support of the fuzzy rule against its support under the assumption of statistical independence. Consequently, we can interpret the result of Eq. (11) as follows:

$$Lift(A_j \rightarrow C_j)\begin{cases} < 1, & \text{if } A_j \text{ and } C_j \text{ are negatively related;} \\ = 1, & \text{if } A_j \text{ and } C_j \text{ are independent;} \\ > 1, & \text{if } A_j \text{ and } C_j \text{ are positively related.} \end{cases}$$

Thus, in the fuzzy rule generation process, instead of checking whether the confidence is greater than *MinConf* to decide that an itemset does not need to be extended, we check whether the lift is greater than 1 to obtain exclusively the rules with a positive relationship between the antecedent and the consequent. In this case, we generate the corresponding fuzzy rule and we assign the confidence of the itemset as the rule weight. Note that any of the confirmation measures studied in [42] could be used for the rule generation process, by setting 0 instead of 1 as the value of the threshold, to obtain the same set of rules.

We must point out that when computing the lift we use the matching degree and, consequently, the modification proposed in Section 3.1 is also involved in the fuzzy association rule generation process. In [43], Burda points out that not every T-norm is suitable for the fuzzy lift computation. However, in our case, as the itemset in the consequent of fuzzy rules is a class, we have no restrictions.

3.3. Selection of fuzzy association rules

In the rule selection process, FARC–HD quantifies the interest of the rules of a specific class (locally to that class) by applying Eq. (8) and consequently, it does not take into account fuzzy rules form other classes. In our new proposal, we use Eq. (8) as well but we do the selection globally, i.e., when the fuzzy rules of all classes have been learned by the Apriori algorithm. This may seem like a small change, but it boosts and prioritizes the selection of fuzzy rules of the minority class because:

- The first term of the product is a sort of support of the fuzzy rule but focused on a specific class. Therefore, both classes are under the same conditions as the class distribution does not affect the result.
- The penalizing factor of the second term of the product implies a larger penalization for the fuzzy rules of the majority class and, consequently, fuzzy rules from the minority class have more chances of being selected (and earlier) than those of the majority class.

Moreover, for the sake of boosting the selection of fuzzy rules of the minority class, we also propose the following modifications on the hyper-parameters of this stage:

- To use a different value for K_t for each class, which determines the number of times an example needs to be covered before setting its weight to 0. Specifically, we set it to $K_{t,C_j} = \lfloor \frac{K_t}{f_{C_j}} \rfloor$, which implies a larger value for examples of the minority class. In this manner, examples of the minority class need to be covered more times than those of the majority class, giving more chances to fuzzy rules of the minority class to be selected.
- To modify the updating of the counter that quantifies the number of times an example is covered (c_p). We propose to increase it only when the class of the example coincides with that of the fuzzy rule. Consequently, we assure that examples are covered by K_{t,C_j} fuzzy rules of its class before setting their weights to 0. Consequently, the weights of the examples of the minority class will not be set to 0 because they are only covered by fuzzy rules of the majority class.
- To set the initial values of the weights of the examples depending on their class. We propose the initial weights to range between the value of an example that has been covered by one fuzzy rule and the initial weight value in the original setting, i.e., to be within $[0.5, 1.0]$. Specifically, we set the weights of examples of the minority class to 1.0 and that of those of the majority class to the result of $1.0/IR$ translated to $[0.5, 1.0] : \frac{1}{2 \cdot IR} + 0.5$. As a result, the larger the *IR* the closer the weight of examples of the majority class will be to 0.5, which implies that fuzzy rules of the minority class would have more chances of being selected.

3.4. Fitness function of the evolutionary process

The quality of the chromosomes of the evolutionary process in the last stage of FACR-HD is measured by Eq. (9), which is based on the accuracy rate ($\#Hits$) and the ratio of selected fuzzy rules.

It is well-known that the accuracy rate is not a suitable metric for imbalanced problems and, therefore, we propose to use the $F - score$ (Eq. (3)))) to measure the performance of each solution. Furthermore, as the number of generated rules in this setting is usually not as large as in standard classification problems, we propose to omit the ratio of selected fuzzy rules. Consequently, the fitness function used in FARCI is

$$Fitness(C) = F - score$$

Different metrics for imbalanced problems [44] could be used to guide to optimization process by changing the $F - score$ by the desired metric in the fitness function.

4. Experimental framework

In this section, we present the set-up of the experimental framework for the experiments conducted in this paper. First, we introduce the selected datasets for the experimental study and the considered statistical tests to support the quality of our proposal (Section 4.1). Finally, we show the methods used in the comparative study and their set-up (Section 4.2).

4.1. Datasets and statistical tests

We have selected 66 datasets from the KEEL dataset repository [29]. Specifically, 22 of the datasets have an IR less than 9.0 and the remaining 44 datasets have an IR larger than 9.0, belonging to Part I and II of this repository. Features of these datasets as well as the detailed results obtained from the approaches considered in this paper can be shown in https://github.com/JoseanSanz/FARCI.

To carry out the different experiments we consider a *stratified 5-fold cross-validation model*, i.e., we take 5 random partitions of the data (each with 20%) and, for each, consider the combination of 4 of them (80%) as training and the remaining one as test. For each dataset we show the average results of the five partitions.

To give statistical support to the analysis of the results, we conduct a series of non-parametric tests as suggested in the specialized literature [30]. Specifically, for pairwise and group comparisons we consider the Wilcoxon signed-ranks test and the Aligned Friedman test besides the Holm *post hoc* test, respectively. Holm's test is applied to detect the algorithms rejecting the null hypothesis of equivalence (we report the Adjusted P-Value, APV) against the control method determined by the Aligned Friedman test.

4.2. Methods and set-up for comparison

The base set-up of our new method is shown in Table 2.[4] Linguistic labels are modeled using uniformly distributed triangular membership functions, which form a strong partition. In our experiments, we study the influence of the operation used to compute the matching degree (product, arithmetic mean and geometric mean) as well as the criterion used in the fuzzy association rule generation process (confidence and lift).

We confront our proposal versus the original FARC–HD fuzzy classifier [21] after processing the data by sampling methods. The set-up of FARC–HD is the one shown in Table 2 in order to make a fair comparison. Specifically, we used:

- Under-sampling methods: Random Under-Sampling (RUS) [9], Tomek's Links (TL) [45], Condensed Nearest Neighbor (CNN) [46], One Sided Selection (OSS) [47] and Neighborhood Cleaning Rule (NCL) [48].
- Over-sampling methods: Random over-sampling (ROS) [9] and Synthetic Minority Over-sampling TEchnique (SMOTE) [49].
- Hybrid methods: SMOTE besides TL (SMOTE-TL) [9] and SMOTE besides Edited Nearest Neighbor (SMOTE-ENN) [9].

We will only show the results of those sampling methods that help FARC–HD enhance its results with respect to those obtained without applying any sampling method.

Furthermore, we compare our method to all internal approaches and ensembles provided in the KEEL software tool [50]:

- Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems Hierarchical [12], *GP-COACH-H*: it is a genetic programming approach that creates a hierarchical knowledge base, which is subsequently optimized by an evolutionary process that performs both a fuzzy rule selection process and a tuning of the lateral position of the membership functions. This process is carried out after preprocessing the data by means of SMOTE.
- C4.5 Cost-Sensitive [15], *C45CS*: it is a generalization of the standard C4.5 decision tree induction algorithm where an instance-weighting method is used to change the class distribution, which implies to learn trees in favour to the class associated with larger costs (weights).
- C-SVM Cost-Sensitive [14], *C-SVMCS*: it is a modification of support vector machines, where a guided repetitive under-sampling strategy is implemented to "rebalance" the dataset at hand. It allows both to extract informative examples and to remove redundant and noisy ones.

[4] We have to point out that the minimum confidence hyper-parameter is not used in the versions of FARCI using the lift in the fuzzy association rules learning stage.

Table 2
Setup of FARCI.

Parameters
Num. of linguistic labels per variable: 5
Minimum Support: 0.05
Minimum Confidence: 0.8
Maximum depth: 3
Parameter k: 2
Evaluations: 20000
Number of individuals: 50
α parameter: 0.02
Bits per gen: 30
Rule weight: confidence
Inference: additive combination

- Ensemble learning classifier with under-sampling preprocessing with C4.5 Decision Tree as Base Classifier [18], *EUSBoost*: this approach combines the Boosting algorithm with the evolutionary under-sampling technique to construct an ensemble of C4.5 decision trees. Furthermore, it promotes diversity by favoring the usage of different subsets of majority class instances to train each base classifier, which makes it outperform state-of-the-art ensemble methods in imbalanced domains.

The configuration of the previous methods is the default one provided in the KEEL software tool according to the suggestion of the authors of the different techniques [50].

5. Experimental results and analysis

This section is aimed at showing the results of our new approach with a triple objective:

- To analyze the influence of the different components and the hyper-parameters of the fuzzy association rule learning stage of FARCI in terms of performance as well as rule base size (Section 5.1).
- To study whether our new method (when applying all the components) is able to enhance the results of the FARC–HD fuzzy classifier, since it is the basis of FARCI, alongside sampling methods (Section 5.2).
- To check whether it provides competitive results versus well-known methods developed to avoid the usage of sampling techniques (Section 5.2).

5.1. Analyzing the influence of the components of FARCI

Though our final proposal makes usage of all the modifications described in Section 3, in this section we study the suitability of each one by showing the results of FARCI in an incremental way. That is, we start showing the results of the original FARC–HD, which is the basis of our new method, and then we include each component one by one. First, we only include the modification of the Fitness function ($FARCI_F$), then, we include the modification of the fuzzy association rule learning stage by using the Lift ($FARCI_{FL}$) and, finally, we add the changes made in the fuzzy rules Selection stage ($FARCI_{FLS}$). For each component, we also show the influence of the matching degree computation, namely, when using the product, the arithmetic mean and the geometric mean (they are identified using P, AM and GM as superscripts, respectively).

In Table 3 we show the results obtained in testing by the different versions of *FARCI*, which are measured in terms of the $F-score$. We only show the mean result of each version (and the standard deviation $\pm$), that is, for each dataset we compute the mean of the 5 folds and we average the result of the 66 datasets.[5] This table is horizontally split in groups according to the components of the new approach. Specifically, we show the global mean (in the 66 datasets), $Mean_G$, besides the averaged results in datasets whose IR is less, $Mean_{IR<9}$, and larger or equal, $Mean_{IR\geqslant 9}$, than 9. In each scenario, we highlight in **bold-face** the best result.

To study how the different components affect the rule base size, we consider the average number of: 1) fuzzy rules for each class (*nR*+ and *nR*- for the positive and negative classes, respectively) and 2) antecedents by fuzzy rule for each class (*nA*+ and *nA*-) for each component. We show these results in Table 4.

According to the results shown in Table 3 we can observe a leap in the performance of the system each time a new component is introduced. This improvement is clearly shown both in the global behaviour of the method as well as for datasets whose IR is larger or equal than 9. For low imbalanced datasets ($IR < 9$), the main contribution is provided by the inclusion of the fitness function, another leap can be observed when introducing the modification in the fuzzy rules selection ($FARCI_{FLS}$)

[5] The results obtained for each dataset and fold by each version can be found in https://github.com/JoseanSanz/FARCI

Table 3
Averaged testing results ($F - score$) obtained by the different versions of FARCI (and the standard deviation $\pm$).

Method	$Mean_G$	$Mean_{IR<9}$	$Mean_{IR\geqslant 9}$
FARC–HD	0.6073 $\pm$ 0.1157	0.7404 $\pm$ 0.0562	0.5407 $\pm$ 0.1448
$FARCI_F^P$	0.6377 $\pm$ 0.1207	0.7760 $\pm$ 0.0524	0.5685 $\pm$ 0.1542
$FARCI_F^{AM}$	0.6392 $\pm$ 0.1160	0.7836 $\pm$ 0.0430	0.5670 $\pm$ 0.1517
$FARCI_F^{GM}$	0.6425 $\pm$ 0.1181	0.7832 $\pm$ 0.0502	0.5722 $\pm$ 0.1513
$FARCI_{FL}^P$	0.6600 $\pm$ 0.1271	0.7826 $\pm$ 0.0486	0.5986 $\pm$ 0.1655
$FARCI_{FL}^{AM}$	0.6608 $\pm$ 0.1299	0.7828 $\pm$ 0.0522	0.5997 $\pm$ 0.1679
$FARCI_{FL}^{GM}$	0.6576 $\pm$ 0.1313	0.7775 $\pm$ 0.0513	0.5977 $\pm$ 0.1704
$FARCI_{FLS}^P$	0.6666 $\pm$ 0.1210	0.7876 $\pm$ 0.0479	0.6062 $\pm$ 0.1569
$FARCI_{FLS}^{AM}$	0.6708 $\pm$ 0.1150	0.7886 $\pm$ 0.0475	0.6119 $\pm$ 0.1480
$FARCI_{FLS}^{GM}$	**0.6768** $\pm$ **0.1205**	**0.7917** $\pm$ **0.0452**	**0.6194** $\pm$ **0.1573**

Table 4
Average rule base sizes of the different versions of FARCI. *nR+* and *nR-* are the number of rules of the minority and majority classes whereas *nA+* and *nA-* are their average number of antecedents.

Method	*nR+*	*nR-*	*nA+*	*nA-*
FARC–HD	4.04	4.22	2.14	1.16
$FARCI_F^P$	4.68	5.09	2.14	1.18
$FARCI_F^{AM}$	4.64	5.43	2.15	1.14
$FARCI_F^{GM}$	4.73	5.40	2.14	1.14
$FARCI_{FL}^P$	4.47	5.14	1.04	1.07
$FARCI_{FL}^{AM}$	4.47	5.02	1.04	1.06
$FARCI_{FL}^{GM}$	4.44	5.01	1.04	1.06
$FARCI_{FLS}^P$	8.17	4.66	1.06	1.05
$FARCI_{FLS}^{AM}$	8.70	4.80	1.11	1.07
$FARCI_{FLS}^{GM}$	8.59	4.74	1.10	1.06

and the remainder components do not provide a significant improvement. Therefore, the analysis made in this section is focused in the behavior of FARCI at global level, that is, we will use only the results shown in the second column of Table 3 ($Mean_G$).

The first component we analyze is the modification of the fitness function $FARCI_F$. We can observe in Table 3 that, when using the product ($FARCI_F^P$) as it is the operation used in FARC–HD to compute the matching degree, it produces and important leap in the performance. We show, in the second row of Table 5, the results of Wilcoxon's test to compare both approaches, where the superiority of using the modified fitness function is statistically confirmed.

Looking at the results shown in Table 4, we can observe that this component increases the number of fuzzy rules of both classes, which seem to allow FARCI to improve the results of FARC–HD. On the other hand, this component does not have an impact the number of antecedents. However, we can see that the average number of antecedents on the fuzzy rules from both classes is different, which is the main motivation of the change developed in the matching degree computation. Consequently, if we analyze the behavior in this component of the three operations ($FARCI_F^P$, $FARCI_F^{AM}$ and $FARCI_F^{GM}$) we can see that the two averaging aggregation functions allow to improve the results of the product because they do not penalize fuzzy rules having a larger number of antecedents, which is the case of those of the minority class. To study whether there are statistical differences among these three approaches, in the second column of Table 6 we show the results of the Aligned Friedman test, where we report both the obtained rank (the less the better) obtained and the APV provided by the Holm's post hoc test (value in parenthesis) when comparing the control method (the one associated with the less rank) versus the approach in the row. From these results we can see that the best option is to use the geometric mean but there are no statistical differences among them.

Then, we analyze the results of the usage of the lift in the fuzzy rule learning stage, $FARCI_{FL}$. According to the results in Table 3 (third group of results) we can observe that this component produces a rise of between 1.5% and 2.0% depending on the operation used for the matching degree computation regarding the results of the first modification ($FARCI_F$). This increase is even more accentuated in the highly imbalanced datasets. In order to support the quality of this component, we have made three Wilcoxon's tests (one for each matching degree option) to compare it versus the previous modification. These results can be seen in second group of results (third, fourth and fifth rows) of Table 5, where we can find statistical differences when using the product and the arithmetic mean and a relatively low p-value when using the geometric mean. Therefore, the suitability of this component is proved. In the third column of Table 6 we can find the statistical comparison of the three options

Table 5
Wilcoxon's tests to compare the different FARCI versions. *R+* and *R-* represent the ranks for the first and the second method of the comparison, respectively.

Comparison	*R+*	*R-*	p-value
FARC–HD vs. $FARCI_F^P$	721.5	1489.5	0.02
$FARCI_F^P$ vs. $FARCI_{FL}^P$	783	1428	0.04
$FARCI_F^{AM}$ vs. $FARCI_{FL}^{AM}$	826.5	1384.5	0.07
$FARCI_F^{GM}$ vs. $FARCI_{FL}^{GM}$	889	132	0.16
$FARCI_{FL}^P$ vs. $FARCI_{FLS}^P$	935	1276	0.27
$FARCI_{FL}^{AM}$ vs. $FARCI_{FLS}^{AM}$	946	1265	0.30
$FARCI_{FL}^{GM}$ vs. $FARCI_{FLS}^{GM}$	677.5	1533.5	< 0.01

Table 6
Ranks obtained by the Aligned Friedman ranks test and APVs computed with the Holm's post hoc test (in parenthesis). The different matching degrees operations are compared for each component of FARCI.

Operation/ Version	$FARCI_F$	$FARCI_{FL}$	$FARCI_{FLS}$
P	107.53 (0.33)	95.11	104.80 (0.09)
AM	97.33 (0.71)	99.61 (0.65)	106.92 (0.09)
GM	93.64	103.79 (0.65)	86.78

for the matching degree computation in this component, $FARCI_{FL}$, where we observe that they do not have statistical differences among themselves. An important fact implied by the modification of the fuzzy rule learning stage is that the generated fuzzy rules are simpler as can be observed in Table 4. The number of fuzzy rules is not boosted because of the usage of the lift but they have been simplified as the average number of antecedents is nearly one. Consequently, our hypothesis to develop this change is corroborated by the experimental results.

Next, we study the influence of the fuzzy rule selection stage, $FARCI_{FLS}$. This component produces an increase of up to 2% as can be observed in the last group of results of Table 3 and this superiority is confirmed by the Wilcoxon's test as observed in the three last rows of Table 5. The reason behind the notable enhancement of this component can be explained by the final number of fuzzy rules of the minority class, which is almost twice as that of the remainder approaches (see the last three rows of Table 4 versus the remainder ones). From these results, we highlight the behavior of the geometric mean to compute the matching degree as its results excel those of the remainder approaches, which is supported in the group statistical test shown in the last column of Table 6.

All in all, we have proved the appropriateness of each component, where the best configuration of FARCI is achieved using all the new components and the geometric mean to perform the matching degree computation: $FARCI_{FLS}^{GM}$.

We must point out that in https://github.com/JoseanSanz/FARCI we also show the same statistical study carried out in the two remainder scenarios ($IR < 9$ and $IR \geqslant 9$), where similar results are found for the latter scenario whereas for low imbalanced datasets there are not statistical differences among the different components, except when including the modification of the fitness function and the usage of the whole system applying the geometric mean. This is the expected behaviour as the new method is developed to face imbalanced problems and consequently, the larger the *IR* the better the expected enhancement of the results is.

Finally, once we have determined the best configuration of FARCI, we want to study the influence of the hyper-parameters of the fuzzy rule association learning process in the results of the new approach. On the one hand our new method does not use the confidence, which implies that we do not have to set the value of the minimum confidence threshold, *MinConf*. On the other hand, the hyper-parameters that could have a potential influence on the generated fuzzy rules are the minimum support, *MinSupp*, and the maximum depth of the search tree, *MaxDepth*. For this reason, we are going to study their impact on the performance results and the rule base size using the best configuration of FARCI, $FARCI_{FLS}^{GM}$.

We start by studying the effect of the *MinSupp* hyper-parameter by changing the default value (0.05) to 0.03 and 0.07. The testing results and the rule base sizes of these three approaches are introduced in Tables 7 and 8, respectively. The effect of this hyper-parameter in the rule base size is the expected one because the larger it gets the less the number of generated fuzzy rules becomes. Looking at the testing results, we can observe a notable impact because setting 0.05 as the value is the best option in the three scenarios. This fact is confirmed in the statistical study carried out by means of the aligned Friedman test comparing these three approaches using all the datasets, whose results are shown in Table 9. Therefore, we can conclude that the default value was appropriate.

In order to study the effect of the maximum depth of the trees, *MaxDepth*, we use the best configuration found so far, that is, $FARCI_{FLS}^{GM}$ using 0.05 as *MinSupp*. In this part of the study, we have considered the values 2, 3, 4 and 5 as value of the hyper-parameter under study. The testing performance and the size of the rule bases are reported in Tables 10 and 11, respectively.

Table 7
Averaged testing results ($F - score$), and the standard deviation $\pm$, obtained by the different values of *MinSupp* in $FARCI_{FLS}^{GM}$.

MinSupp	$Mean_G$	$Mean_{IR<9}$	$Mean_{IR\geqslant 9}$
0.03	0.6697 $\pm$ 0.1197	0.7868 $\pm$ 0.0470	0.6112 $\pm$ 0.1552
0.05	**0.6768** $\pm$ **0.1205**	**0.7917** $\pm$ **0.0452**	**0.6194** $\pm$ **0.1573**
0.07	0.6628 $\pm$ 0.1215	0.7860 $\pm$ 0.0470	0.6011 $\pm$ 0.1579

Table 8
Average rule base sizes using the different different values of *MinSupp* in $FARCI_{FLS}^{GM}$.

MinSupp	*nR+*	*nR-*	*nA+*	*nA-*
0.03	9.58	4.96	1.16	1.08
0.05	8.59	4.74	1.10	1.06
0.07	8.24	4.52	1.09	1.05

Table 9
Ranks obtained by the Aligned Friedman ranks test and APVs computed with the Holm's post hoc test (in parenthesis). The different values of *MinSupp* in $FARCI_{FLS}^{GM}$.

MinSupp	Rank (APV)
0.05	83.39
0.03	101.98 (0.06)
0.07	113.12 (<0.01)

Table 10
Averaged testing results ($F - score$), and the standard deviation $\pm$, obtained by the different values of *MaxDepth* in $FARCI_{FLS}^{GM}$.

MaxDepth	$Mean_G$	$Mean_{IR<9}$	$Mean_{IR\geqslant 9}$
2	**0.6773** $\pm$ **0.1157**	0.7893 $\pm$ 0.0458	**0.6212** $\pm$ **0.1499**
3	0.6768 $\pm$ 0.1205	0.7917 $\pm$ 0.0452	0.6194 $\pm$ 0.1573
4	0.6767 $\pm$ 0.1185	0.7932 $\pm$ 0.0453	0.6184 $\pm$ 0.1543
5	0.6769 $\pm$ 0.1185	**0.7935** $\pm$ **0.0442**	0.6186 $\pm$ 0.1548

Table 11
Average rule base sizes using the different different values of *MaxDepth* in $FARCI_{FLS}^{GM}$.

MaxDepth	*nR+*	*nR-*	*nA+*	*nA-*
2	8.44	4.74	1.06	1.04
3	8.59	4.74	1.10	1.06
4	8.74	4.74	1.12	1.07
5	8.77	4.75	1.14	1.07

Looking at these results, we can observe that the performance of the system is not influenced to a large degree by this hyper-parameter. Regarding the impact on the rule base size, it is again the expected one because we can observe a slight increase on the number of antecedents when the maximum depth of the trees is also increased. The results of the statistical study reported in Table 12 confirm that all these results are equivalent.

5.2. Comparative study versus sampling techniques and related methods

In this section, we conduct an experimental study to determine the quality of our new method using the combination of all the proposals and the default values for *MinSupp* and *MaxDepth*, $FARCI_{FLS}^{GM}$, which is shown to be the best option in the previous section. Specifically, first, we compare the performance of *FARCI* versus that obtained by the original FARC–HD [21] applied using the data preprocessed by well-known sampling algorithms such as TL, NCL, ROS, SMOTE, SMOTE-TL

Table 12
Ranks obtained by the Aligned Friedman ranks test and APVs computed with the Holm's post hoc test (in parenthesis). The different values of *MaxDepth* in $FARCI_{FLS}^{GM}$.

MaxDepth	Rank (APV)
5	130.55
4	130.96 (1.00)
2	131.26 (1.00)
3	137.23 (1.00)

and SMOTE-ENN. The results of FARC–HD alongside these six sampling methods are reported in Table 13, which has the same structure of Table 3. From these results, we can observe that our new method achieves a notable enhancement versus the results provided by all sampling methods, being ROS the one allowing FARC–HD to achieve the best results. We have conducted an statistical test among these seven methods, whose results are shown in Table 14. This table is sorted according to the ranks, obtained in the Aligned Friedman ranks test, from the best to the worst one. From these results, we can see statistical differences in favour to $FARCI_{FLS}^{GM}$ with respect to all approaches except ROS. In this case, we have carried out a Wilcoxon test to compare both methods (Table 15), where we observe statistical differences in favour to our new method.

We have also obtained the rule base sizes (Table 16) obtained by FARC–HD applied besides these sampling techniques. From these results we can observe the following facts:

- Under-sampling methods (TL and NCL) make FARC–HD obtain a low number of fuzzy rules (and balanced among the two classes). However, the average number of antecedents of fuzzy rules from each class is not equal, which can imply that fuzzy rules of the minority class are in an inferiority position with respect to those of the majority class, since FARC–HD uses the product to compute the matching degree.
- Over-sampling (ROS and SMOTE) and hybrid (SMOTE-TL and SMOTE-ENN) methods allow FARC–HD to obtain fuzzy rules whose average number of antecedents depending on the class is similar. However, in this case, the number of fuzzy rules belonging to the majority class is larger than that of the minority class, which can make the prediction of the minority class less probable.

Finally, we have also compared the performance of FARCI versus some related methods. Specifically, we consider another algorithm modification method like GP-COACH-H, two cost-sensitive approaches as C45CS and C-SVMCS and an ensemble

Table 13
Averaged testing results ($F - score$) obtained by FARC–HD alongside sampling techniques (and the standard deviation $\pm$).

Method	$Mean_G$	$Mean_{IR<9}$	$Mean_{IR\geqslant 9}$
TL	0.6355 ± 0.1187	0.7783 ± 0.0521	0.5641 ± 0.1513
NCL	0.6442 ± 0.1164	0.7730 ± 0.0503	0.5798 ± 0.1486
ROS	0.6603 ± 0.1083	0.7749 ± 0.0455	0.6029 ± 0.1390
SMOTE	0.6304 ± 0.0780	0.7690 ± 0.0395	0.5611 ± 0.1000
SMOTE-TL	0.6138 ± 0.0805	0.7570 ± 0.0446	0.5422 ± 0.0981
SMOTE-ENN	0.6126 ± 0.0914	0.7606 ± 0.0486	0.5386 ± 0.1123
$FARCI_{FLS}^{GM}$	**0.6768 ± 0.1205**	**0.7917 ± 0.0452**	**0.6194 ± 0.1573**

Table 14
Aligned Friedman ranks test and Holm's post hoc test to compare FARCI versus FARC–HD alongside sampling methods.

Method	Rank (APV)
$FARCI_{FLS}^{GM}$	151.86
ROS	183.28 (0.18)
TL	213.25 (0.02)
NCL	217.76 (0.01)
SMOTE	267.14 (< 0.01)
SMOTE-ENN	292.40 (< 0.01)
SMOTE-TL	294.83 (< 0.01)

Table 15
Wilcoxon's test to compare FARCI versus FARC–HD alongside ROS.

Comparison	*R+*	*R-*	p-value
$FARCI^{GM}_{FLS}$ vs. FARC–HD + ROS	1424.5	786.5	0.04

Table 16
Rule base size of FARC–HD when applied alongside sampling techniques.

Method	*nR+*	*nR-*	*nA+*	*nA-*
TL	4.11	3.85	2.17	1.15
NCL	4.00	3.95	2.15	1.17
ROS	4.34	7.87	1.90	1.84
SMOTE	5.66	8.52	2.07	1.87
SMOTE-TL	5.48	8.00	1.96	1.86
SMOTE-ENN	5.57	7.62	1.98	1.82

Table 17
Averaged testing results ($F - score$) obtained by the related methods (and the standard deviation $\pm$).

Method	$Mean_G$	$Mean_{IR<9}$	$Mean_{IR\geqslant 9}$
GP-COACH-H	0.6338 $\pm$ 0.0876	0.7710 $\pm$ 0.0402	0.5652 $\pm$ 0.1107
C45CS	0.6417 $\pm$ 0.1042	0.7729 $\pm$ 0.0450	0.5762 $\pm$ 0.1331
C-SVMCS	0.6165 $\pm$ 0.0759	0.7740 $\pm$ 0.0379	0.5377 $\pm$ 0.0944
EUSBoost	0.6062 $\pm$ 0.0886	0.7719 $\pm$ 0.0537	0.5233 $\pm$ 0.1057
$FARCI^{GM}_{FLS}$	**0.6768** $\pm$ **0.1205**	**0.7917** $\pm$ **0.0452**	**0.6194** $\pm$ **0.1573**

Table 18
Aligned Friedman ranks test and Holm's post hoc test to compare FARCI versus the related methods.

Method	Rank (APV)
GP-COACH-H	164.11 ($<$ 0.01)
C45CS	157.38 ($<$ 0.01)
C-SVMCS	193.76 ($<$ 0.01)
EUSBoost	204.11 ($<$ 0.01)
$FARCI^{GM}_{FLS}$	108.14

such as EUSBoost. The performance of these four methods and that of $FARCI^{GM}_{FLS}$ is reported in Table 17 and the statistical test to compare these five approaches can be seen in Table 18. From these results, the quality of our new method is clearly supported.

6. Conclusions

In this paper we have proposed a Fuzzy Association Rule-based Classifier for Imbalanced classification problems, FARCI. The new method belongs to the algorithm modification methods category, since it is constructed on the basis of the FARC–HD classifier. Specifically, we have modified all the stages of this classifier so that it can cope directly with imbalance classification problems. In the fuzzy rule generation step, we have proposed to use the lift as the criterion to evaluate the quality of fuzzy rules. We have also changed the philosophy and the initialization of the hyper-parameters of the fuzzy rule selection component as well as the fitness function of the evolutionary process. Furthermore, we have proposed the usage of averaging aggregation functions to compute the matching degree, which is an operation used in all the components of the classifier.

The suitability of each modification has been empirically proved in the experimental study. We can stress the following facts:

1. The usage of a proper metric to measure the quality of the solutions in the evolutionary process leads to a notable enhancement of the obtained results.
2. Averaging aggregation functions allow fuzzy rules with variable antecedent lengths to be equally taken into account, which positively affects the performance of the system.
3. The lift makes possible to not only improve the system's performance but to simplify the fuzzy rules.

4. The new philosophy behind the fuzzy rule selection stage allows FARCI to boost the number of fuzzy rules of the minority class, which implies a better discrimination capability of the classifier.
5. When combining all the modifications we obtain the final classifier, $FARCI_{FLS}^{GM}$, which provides better results than those of the original FARC–HD classifier combined with sampling techniques because both the rule structure as well as the number of rules of the majority class are more suitable for this type of problems.
6. $FARCI_{FLS}^{GM}$ also achieves a more accurate performance than that of four related methods that do not make usage of sampling techniques to tackle imbalanced classification problems.

All in all, our new method provides a competitive performance as well as a simple and interpretable rule base due to the new design of the components of the classifier.

As future research lines, we intend to study the behavior of different generalizations of the Choquet integral [22] in order to fuse the local information given by the fired fuzzy rules, since they have been shown to be competitive in standard classification problems. We are also interested in studying the usage of interval-valued fuzzy sets in this framework as they have been successfully applied in classification problems [32]. Moreover, it would be also interesting to study other pattern weighting schemes to select the most interesting fuzzy rules as well as analyzing different data properties in order to try to improve the performance of the system under specific circumstances.

CRediT authorship contribution statement

J. Sanz: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **M. Sesma-Sara:** Methodology, Investigation, Writing - original draft, Writing - review & editing, Visualization. **H. Bustince:** Investigation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the Spanish Ministry of Economy and Competitiveness through the Spanish National Research (project PID2019-108392GB-I00/ AEI/ 10.13039/501100011033) and by the Public University of Navarre under the project PJUPNA1926.

References

[1] I. Škrjanc, J.A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, F. Gomide, Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey, Information Sciences 490 (2019) 344–368.
[2] Q. Huang, B. Hu, F. Zhang, Evolutionary optimized fuzzy reasoning with mined diagnostic patterns for classification of breast tumors in ultrasound, Information Sciences 502 (2019) 525–536.
[3] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review, International Journal of Pattern Recognition and Artificial Intelligence 23 (4) (2009) 687–719.
[4] H. He, Y. Ma, Imbalanced Learning: Foundations, Algorithms, and Applications, 1st Ed., Wiley-IEEE Press, 2013.
[5] A. de Haro-García, G. Cerruela-García, N. García-Pedrajas, Ensembles of feature selectors for dealing with class-imbalanced datasets: A proposal and comparative study, Information Sciences 540 (2020) 89–116.
[6] J.A. Sanz, M. Galar, H. Bustince, T. Belzunegui, An evolutionary underbagging approach to tackle the survival prediction of trauma patients: a case study at the hospital of navarre, IEEE Access 7 (2019) 76009–76021.
[7] J.S.M. Lango, Multi-class and feature selection extensions of roughly balanced bagging for imbalanced data, Journal of Intelligent Information Systems 50 (2018) 97–127.
[8] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Transactions on Knowledge and Data Engineering 21 (9) (2009) 1263–1284.
[9] G. Batista, R. Prati, M. Monard, A study of the behavior of several methods for balancing machine learning training data, SIGKDD Explorations 6 (1) (2004) 20–29.
[10] A. Fernandez, S. Garcia, F. Herrera, N.V. Chawla, SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary, Journal of Artificial Intelligence Research 61 (2018) 863–905.
[11] R. Barandela, J. Sanchez, V. Garcia, E. Rangel, Strategies for learning in class imbalance problems, Pattern Recognition 36 (3) (2003) 849–851.
[12] V. Lopez, A. Fernandez, M. Jose del Jesus, F. Herrera, A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline data-sets, Knowledge-Based Systems 38 (SI) (2013) 85–104.
[13] K. Napierala, J. Stefanowski, BRACID: A comprehensive approach to learning rules from imbalanced data, Journal of Intelligent Information Systems 39 (2) (2012) 335–373.
[14] Y. Tang, Y.-Q. Zhang, N.V. Chawla, S. Krasser, SVMs Modeling for Highly Imbalanced Classification, IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics 39 (1) (2009) 281–288.
[15] K. Ting, An instance-weighting method to induce cost-sensitive trees, IEEE Transactions on Knowledge and Data Engineering 14 (3) (2002) 659–665.
[16] V. López, S. del Río, J.M. Benítez, F. Herrera, Cost-sensitive linguistic fuzzy rule based classification systems under the mapreduce framework for imbalanced big data, Fuzzy Sets and Systems 258 (2015) 5–38.
[17] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 42 (4) (2012) 463–484.

[18] M. Galar, A. Fernandez, E. Barrenechea, F. Herrera, EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling, Pattern Recognition 46 (12) (2013) 3460–3471.
[19] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer-Verlag, Berlin, 2004.
[20] A. Fernandez, M. Elkano, M. Galar, J. Sanz, S. Alshomrani, H. Bustince, F. Herrera, Enhancing evolutionary fuzzy systems for multi-class problems: Distance-based relative competence weighting with truncated confidences (DRCW-TC), International Journal of Approximate Reasoning 73 (2016) 108–122.
[21] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Transactions on Fuzzy Systems 19 (5) (2011) 857–872.
[22] G. Lucca, G.P. Dimuro, J. Fernandez, H. Bustince, B. Bedregal, J.A. Sanz, Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-integrals named C-F1(F2)-integrals, IEEE Transactions on Fuzzy Systems 27 (1) (2019) 124–134.
[23] M. Elkano, M. Galar, J. Sanz, A. Fernandez, E. Barrenechea, F. Herrera, H. Bustince Sola, Enhancing multi-class classification in FARC-HD fuzzy classifier: On the synergy between n-dimensional overlap functions and decomposition strategies, IEEE Transactions on Fuzzy Systems 23 (5) (2014) 1562–1580.
[24] M. Antonelli, P. Ducange, F. Marcelloni, An experimental study on evolutionary fuzzy classifiers designed for managing imbalanced datasets, Neurocomputing 146 (SI) (2014) 125–136..
[25] J. Leski, R. Czabanski, M. Jezewski, J. Jezewski, Fuzzy ordered c-means clustering and least angle regression for fuzzy rule-based classifier: Study for imbalanced data, IEEE Transactions on Fuzzy Systems 28 (11) (2020) 2799–2813.
[26] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, IEEE Transactions on Fuzzy Systems 9 (4) (2001) 506–515.
[27] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison Wesley, 2005.
[28] B. Kavšek, N. Lavrač, Apriori-sd: Adapting association rule learning to subgroup discovery, Applied Artificial Intelligence 20 (7) (2006) 543–583.
[29] 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, F. Herrera, KEEL: a software tool to assess evolutionary algorithms for data mining problems, Soft Computing 13 (3) (2009) 307–318.
[30] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability, Soft Computing 13 (10) (2009) 959–977.
[31] T. Jo, N. Japkowicz, Class imbalances versus small disjuncts, SIGKDD Explorations Newsletter 6 (1) (2004) 40–49.
[32] J. Sanz, D. Bernardo, F. Herrera, H. Bustince, H. Hagras, A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data, IEEE Transactions on Fuzzy Systems 23 (4) (2015) 973–990.
[33] V. López, A. Fernández, J. Moreno-Torres, F. Herrera, Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics, Expert Systems with Applications 39 (7) (2012) 6585–6608.
[34] M. Kubat, R. Holte, S. Matwin, Machine learning for the detection of oil spills in satellite radar images, Machine Learning 30 (2–3) (1998) 195–215.
[35] G.Y. Wong, F.H.F. Leung, S.-H. Ling, A hybrid evolutionary preprocessing method for imbalanced datasets, Information Sciences 454 (2018) 161–177.
[36] R.S.R. Agrawal, Fast algorithms for mining association rules, in: International Conference on Very Large Data Bases, Santiago de Chile, Chile, 1994, p. 487–499..
[37] K. Napierala, J. Stefanowski, I. Szczech, Increasing the interpretability of rules induced from imbalanced data by using bayesian confirmation measures, in: A. Appice, M. Ceci, C. Loglisci, E. Masciari, Z.W. Ras (Eds.), New Frontiers in Mining Complex Patterns, Springer International Publishing, Cham, 2017, pp. 84–98.
[38] R. Alcalá, J. Alcalá-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Transactions on Fuzzy Systems 15 (4) (2007) 616–635.
[39] O. Cordon, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, International Journal of Approximate Reasoning 20 (1) (1999) 21–45.
[40] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, Nonlinear Analysis: Theory, Methods & Applications 72 (3) (2010) 1488–1499.
[41] T. da Cruz Asmus, G.P. Dimuro, B. Bedregal, J.A. Sanz, S. Pereira, H. Bustince, General interval-valued overlap functions and interval-valued overlap indices, Information Sciences 527 (2020) 27–50.
[42] D.H. Glass, Confirmation measures of association rule interestingness, Knowledge-Based Systems 44 (2013) 65–77.
[43] M. Burda, Lift Measure for Fuzzy Association Rules, Springer International Publishing, Cham, 2015, pp. 249–260.
[44] A. Luque, A. Carrasco, A. Martín, A. de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognition 91 (2019) 216–231.
[45] I. Tomek, 2 modifications of CNN, IEEE Transactions on Systems Man and Cybernetics 6 (11) (1976) 769–772.
[46] P. Hart, The condensed nearest neighbour rule, IEEE Transactions on Information Theory 14 (5) (1968) 515–516.
[47] M. Kubat, S. Matwin, Addressing the curse of imbalanced training sets: one-sided selection, in: 14th International Conference on Machine Learning (ICML97), 1997, pp. 179–186.
[48] J. Laurikkala, Improving identification of difficult small classes by balancing class distribution, in: 8th Conference on AI in Medicine in Europe (AIME01), Vol. 2001 of Lecture Notes on Computer Science, Springer, Berlin/ Heidelberg, 2001, pp. 63–66.
[49] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, Journal of Artificial Intelligence Research 16 (2002) 321–357.
[50] I. Triguero, S. Gonzalez, J. Moyano, S. Garcia, J. Alcala-Fdez, J. Luengo, A. Fernandez, M. del Jesus, L. Sanchez, F. Herrera, KEEL 3.0: An open source software for multi-stage analysis in data mining, International Journal of Computational Intelligence Systems 10 (2017) 1238–1249.