Knowledge-Based Systems 73 (2015) 1–17

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A proposal for evolutionary fuzzy systems using feature weighting: Dealing with overlapping in imbalanced datasets

Saleh Alshomrani [a], Abdullah Bawakid [a], Seong-O Shim [a], Alberto Fernández [b,*], Francisco Herrera [a,c]

[a] Faculty of Computing and Information Technology – North Jeddah, King Abdulaziz University (KAU), Jeddah, Saudi Arabia
[b] Department of Computer Science, University of Jaén, Jaén, Spain
[c] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, Granada, Spain

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form 1 August 2014
Accepted 7 September 2014
Available online 16 September 2014

Keywords:
Imbalanced datasets
Overlapping
Feature weighting
Evolutionary fuzzy systems
Fuzzy rule based classification systems

ABSTRACT

In a general scenario of classification, one of the main drawbacks for the achievement of accurate models is the presence of high overlapping among the concepts to be learnt. This drawback becomes more severe when we are addressing problems with an imbalanced class distribution. In such cases, the minority class usually represents the most important target of the classification. The failure to correctly identify the minority class instances is often related to those boundary areas in which they are outnumbered by the majority class examples.

Throughout the learning stage of the most common rule learning methodologies, the process is often biased to obtain rules that cover the largest areas of the problem. The reason for this behavior is that these types of algorithms aim to maximize the confidence, measured as a ratio of positive and negative covered examples. Rules that identify small areas, in which minority class examples are poorly represented and overlap with majority class examples, will be discarded in favor of more general rules whose consequent will be unequivocally associated with the majority class.

Among all types of rule systems, linguistic Fuzzy Rule Based Systems have shown good behavior in the context of classification with imbalanced datasets. Accordingly, we propose a feature weighting approach which aims at analyzing the significance of the problem's variables by weighting the membership degree within the inference process. This is done by applying a different degree of significance to the variables that represent the dataset, enabling to smooth the problem boundaries. These parameters are learnt by means of an optimization process in the framework of evolutionary fuzzy systems. Experimental results using a large number of benchmark problems with different degrees of imbalance and overlapping, show the goodness of our proposal.

1. Introduction

The significance of classification with imbalanced data arises when researchers realize that the datasets they are analyzing hold more instances or examples from one class than that of the remainder, and that they therefore obtain classification models below a desired accuracy threshold for that class. This scenario, known as the problem of classification with imbalanced datasets [41,28], is commonly addressed in a binary context where there is a single minority (positive) class, and a majority (negative) class.

The bias of standard classification algorithms towards the majority class examples [52,27], is the most straightforward consequence derived from the uneven class distribution. Those algorithms which obtain a good behavior in the framework of standard classification do not necessarily achieve the best performance for imbalanced datasets [20]. The imbalanced problem usually appears in combination with several additional data intrinsic characteristics [41]. This imposes further restrictions on the learning stage in terms of it being able to develop a classifier with a high accuracy for the positive and negative classes of the problem.

One of the main drawbacks for the correct identification of the positive class of the problem is related to overlapping between the classes [36,24,13]. Rules with a low confidence and/or coverage, because they are associated with an overlapped boundary area, will be discarded. Therefore, positive class examples belonging to this area are more likely to be misclassified.

* Corresponding author. Tel.: +34 953 213016; fax: +34 953 212472.
E-mail addresses: sshomrani@kau.edu.sa (S. Alshomrani), abawakid@kau.edu.sa (A. Bawakid), seongo@gmail.com (Seong-O Shim), alberto.fernandez@ujaen.es (A. Fernández), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.knosys.2014.09.002
0950-7051/

The representation of a classification problem by means of its variables or features, will determine the way in which the classifier will discriminate between the examples of both classes. It is well known that a large number of features can degrade the discovery of the borderline areas of the problem [39], either because some of these variables might be redundant or because they do not show a good synergy among them. For this reason, some works on the topic have proposed the use of feature selection for imbalanced datasets in order to overcome this problem [54], and to diminish the effect of overlapping [13]. However, the application of feature selection might be too aggressive and therefore some potential features could be discarded. In most cases, every variable of the problem should make at least a small contribution in the learning stage, and the combination of all of them may help to achieve a better separability of the classes.

Linguistic FRBCSs have the advantage of achieving a good performance in the context of classification with imbalanced datasets [19,47]. The use of linguistic fuzzy sets allows the smoothing of the borderline areas in the inference process, which is also a desirable behavior in the scenario of overlapping.

In this paper, we propose the use of a feature weighting approach in the context of Linguistic Fuzzy Rule Based Classification Systems (FRBCSs) [34]. Basically, we propose the consideration of the feature weights as a part of the reasoning model. We modify the computation of the membership functions associated with the fuzzy labels in the antecedents of the rules, in order to take into account the significance of the problem's variables throughout the inference process.

The computation of the optimal parameters for setting the weight of each variable, will be carried out by means of Evolutionary Algorithms [15]. The hybridization of this approach with the previously introduced FRBCSs will lead to the development of an Evolutionary Fuzzy System (EFS) [11,17]. One of the main reasons for the success of this type of techniques is their ability to exploit the information accumulated about and initially unknown search space in order to bias subsequent searches into useful subspaces, i.e. their robustness [11]. For the fuzzy learning classifier, we have considered the use of a robust FRBCS, i.e. the Fuzzy Association Rule-based Classification for High-Dimensional problems (FARC-HD) [1]. The proposed algorithm using feature weighting will receive the acronym FARC-HD-FW, based on the previous name (FARC-HD) and the use of feature weighting.

In order to evaluate the goodness of the feature weighting proposal, we will contrast our results with the standard FARC-HD algorithm and FARC-HD with feature selection. Additionally, we will complement our comparison with the C4.5 decision tree [49] as a standard baseline algorithm, and several EFS approaches developed for both classical and imbalanced classification such as 2-tuples lateral tuning [18], the Hierarchical Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems (GP-COACH-H) [40], and the Interval-Valued Fuzzy Decision Tree (IIVFDT) [50]. The validity of our approach in the scenario of imbalanced and overlapping datasets will be tested using a wide benchmark of 66 different problems commonly used in the topic of classification with imbalanced datasets [41].

This paper is organized as follows. Section 2 briefly introduces the problem of imbalanced data, its relationship with class overlapping and how to address and evaluate this problem. Then, Section 3 contains the central part of the manuscript, in which the proposed methodology for dealing with overlapping in imbalanced data with FRBCSs is described. Next, the details about the experimental framework selected for the validation of our approach are introduced in Section 4. The analysis and discussion of the experimental results is carried out in Section 5. Finally, Section 6 summarizes and concludes the work.

2. Imbalanced datasets in classification

In this section, we present some preliminary concepts regarding classification with imbalanced datasets. This section is divided into the following four parts:

- We will first introduce the problem of imbalanced datasets, describing its features and why is so difficult to learn in this classification scenario (Section 2.1).
- Then, we will focus on the presence of overlapping between the classes, which further complicates the correct identification of the positive instances (Section 2.2).
- In the next section, we will present how to address this problem, focusing on the preprocessing of instances for rebalancing the distribution between the positive and negative classes (Section 2.3).
- Finally, we will discuss how to evaluate the performance of the results in this situation (Section 2.4).

2.1. Basic concepts on classification with imbalanced datasets

The main property of this type of classification problem is that the examples of one class outnumber the examples of the other one [52]. The minority classes are usually the most important concepts to be learnt, since they might be associated with exceptional and significant cases [55] or because the data acquisition of these examples is costly [57]. Since most of the standard learning algorithms consider a balanced training set, this situation may cause suboptimal classification models to be obtained, i.e. a good coverage of the majority examples but a more frequent misclassification of the minority ones [27]. Traditionally, the Imbalance Ratio (IR), i.e. the ratio between the majority and minority class examples [45], is the main clue to identify a set of problems which need to be addressed in a special way.

We must stress the following reasons for this behavior [41]: the use of global performance measures for guiding the search process, such as standard accuracy rate, which may benefit the covering of the majority class examples, and the low coverage of the classification rules for the positive class, which are discarded in favor of more general rules, especially in the case of overlapping [36,13]; small clusters of minority class examples that can be treated as noise and wrongly ignored by the classifier [46,56]; few real noisy examples which may degrade the identification of the minority class, as it has fewer examples to begin with [51]; and dataset shift, i.e. different data distribution between training and test partitions [44]. For an in depth coverage of those data intrinsic characteristics which hinder the classification of imbalanced datasets, the reader may refer to a recent survey carried out in [41].

Finally, regarding the way to overcome the class imbalance problem, we may find a large number of proposed approaches, which can be categorized in three groups [42]:

1. Data level solutions: the objective consists of rebalancing the class distribution by sampling the data space to diminish the effect caused by class imbalance, acting as an external approach [21,25,38].
2. Algorithmic level solutions: these solutions try to adapt several classification algorithms to reinforce the learning towards the positive class. Therefore, they can be defined as internal approaches that create new algorithms or modify existing ones to take the class imbalance problem into consideration [4,58,61].
3. Cost-sensitive solutions: these types of solutions incorporate approaches at the data level, at the algorithmic level, or at both levels jointly. They consider higher costs for the mis-

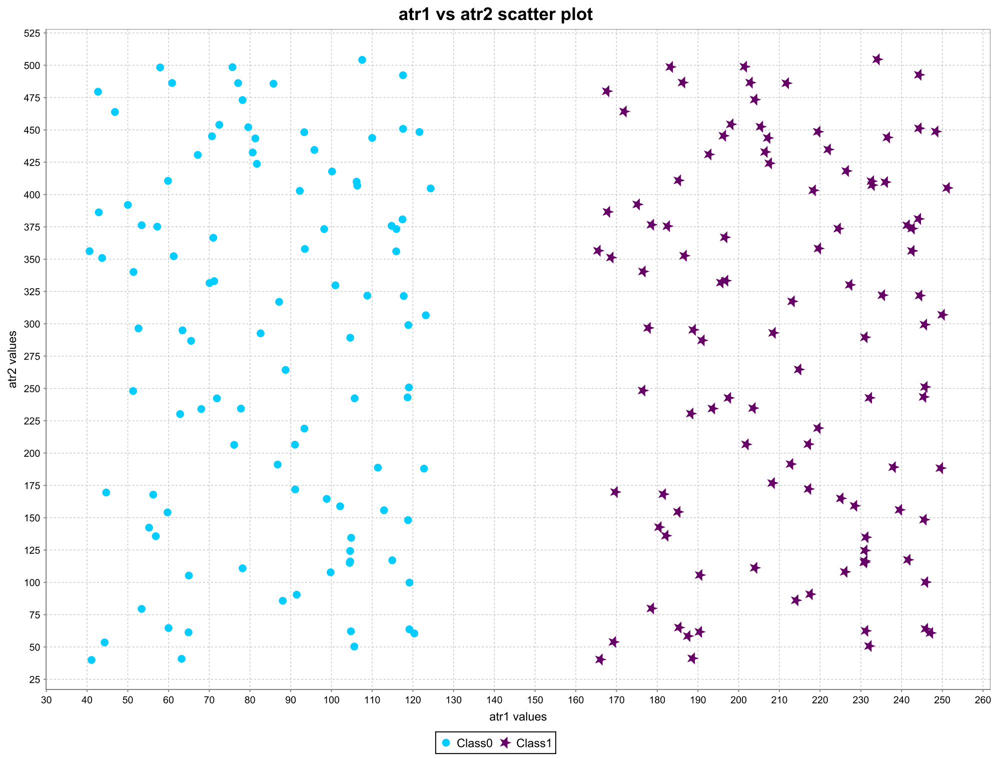


Fig. 1. $F1 = 12.5683$.

classification of examples of the positive class with respect to the negative class, and therefore, trying to minimize higher cost errors [14,53,59].

2.2. The problem of overlapping or class separability

In this paper, we focus on the overlapping between classes. The issue of class separability occurs when a "small" region of the data space is represented by a similar number of training data from both classes; then, the inference mechanism will result from the same a priori probabilities in this area, and the discrimination between the classes will become harder. It is straightforward to observe that any "linearly separable" problem can be solved by a naïve classifier, regardless the class distribution [48].

In one of the latest examples of research into the topic [43], authors have empirically extracted some interesting findings from real world datasets. Specifically, they depicted the performance of the different datasets ordered in accordance with different data complexity measures in order to search for some regions of interesting good or bad behavior. The findings in this work stress that the metrics which measure the overlap between the classes can better characterize the degree of final precision obtained, in contrast to the IR.

The degree of overlap for individual feature values is measured by the so called metric $F1$ or *maximum Fisher's discriminant ratio* [31]. This metric for one feature dimension is defined as:

$$f = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

where $\mu_1, \; \mu_2, \; \sigma_1^2, \; \sigma_2^2$ are the means and variances of the two classes respectively, in that feature dimension. We compute f for each feature and take the maximum as measure F1. For a multidimensional problem, not all features have to contribute to class discrimination. The problem is simple as long as there is one discriminating feature. Therefore, we can just take the maximum f over all feature dimensions when discussing class separability.

Datasets with a small value for the F1 metric will have a high degree of overlapping. Figs. 1–4 show an illustrative example of this behavior, which have been built with synthetic data, using two variables within the range $[0.0; 1.0]$ and two classes.

2.3. Addressing the imbalanced problem: SMOTE preprocessing

Above, we have introduced several approaches to addressing classification with imbalanced datasets. Among them, the advantage of the data level solutions is that they are more versatile, since their use is independent of the classifier selected. Furthermore, we may preprocess all datasets beforehand in order to use them to train different classifiers. In this way, we only need to prepare the data once. Furthermore, previous analysis of preprocessing methods with several classifiers has shown the goodness of the oversampling techniques [5].

The simplest approach, random oversampling, makes exact copies of existing instances, and therefore several authors agree that this method can increase the likelihood of overfitting occurring [5]. Accordingly, more sophisticated methods have been proposed based on the generation of synthetic samples. Among them, the "Synthetic Minority Over-sampling TEchnique" (SMOTE) [9] algorithm, the main idea of which is to form new positive class examples by interpolating between several positive class examples that lie together, has become one of the most significant approaches in this area.

The positive class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. This process is illustrated in Fig. 5, where x_i is the selected point, x_{i1} to x_{i4} are some selected nearest neighbors and r_1 to r_4 the synthetic data points created by the randomized interpolation.

Synthetic samples are generated in the following way: take the difference between the feature vector (sample) under consideration and its nearest neighbor. Multiply this difference by a

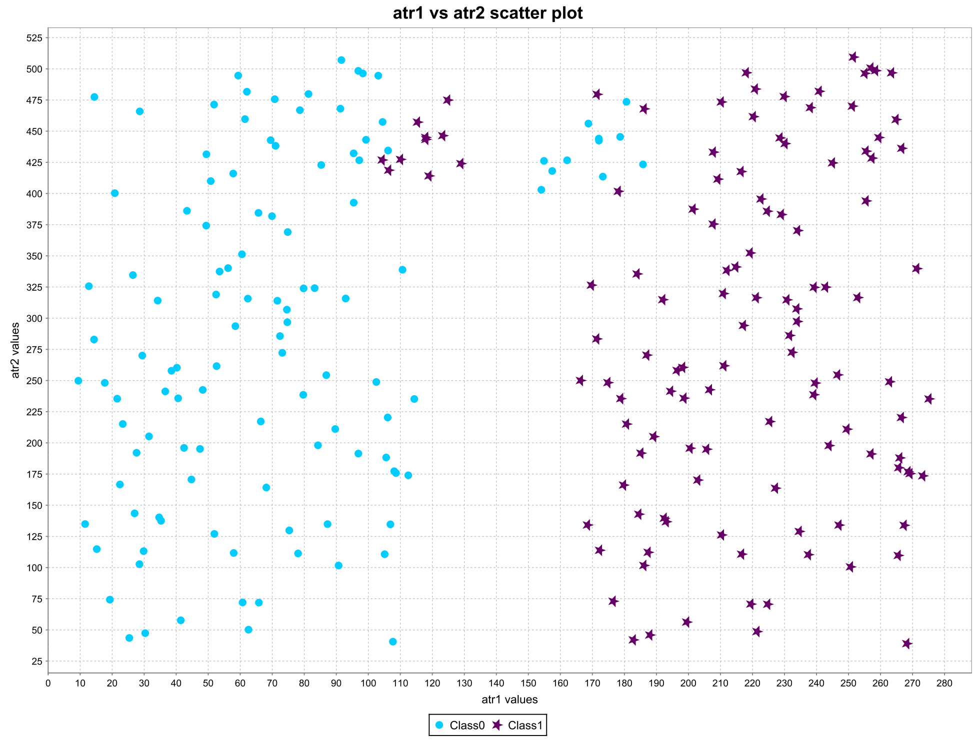


Fig. 2. $F1 = 5.7263$.

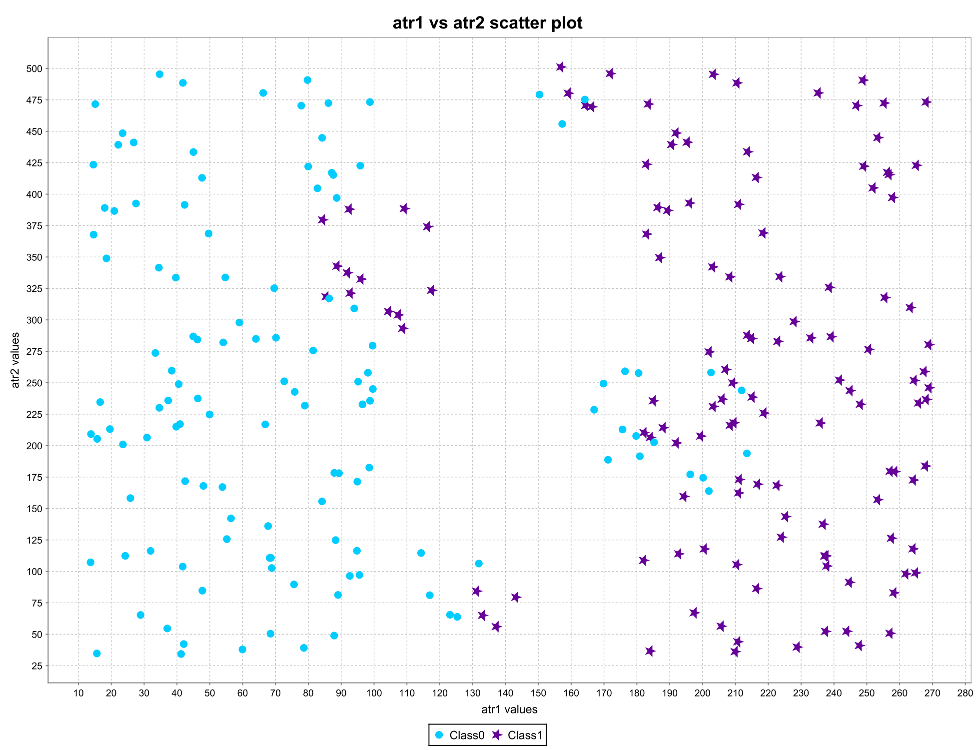


Fig. 3. $F1 = 3.3443$.

random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the positive class to become more general.

2.4. Evaluation in imbalanced domains

The evaluation criterion is a key factor in both assessing the classification performance and guiding the classifier modeling. In a two-class problem, the confusion matrix (shown in Table 1)

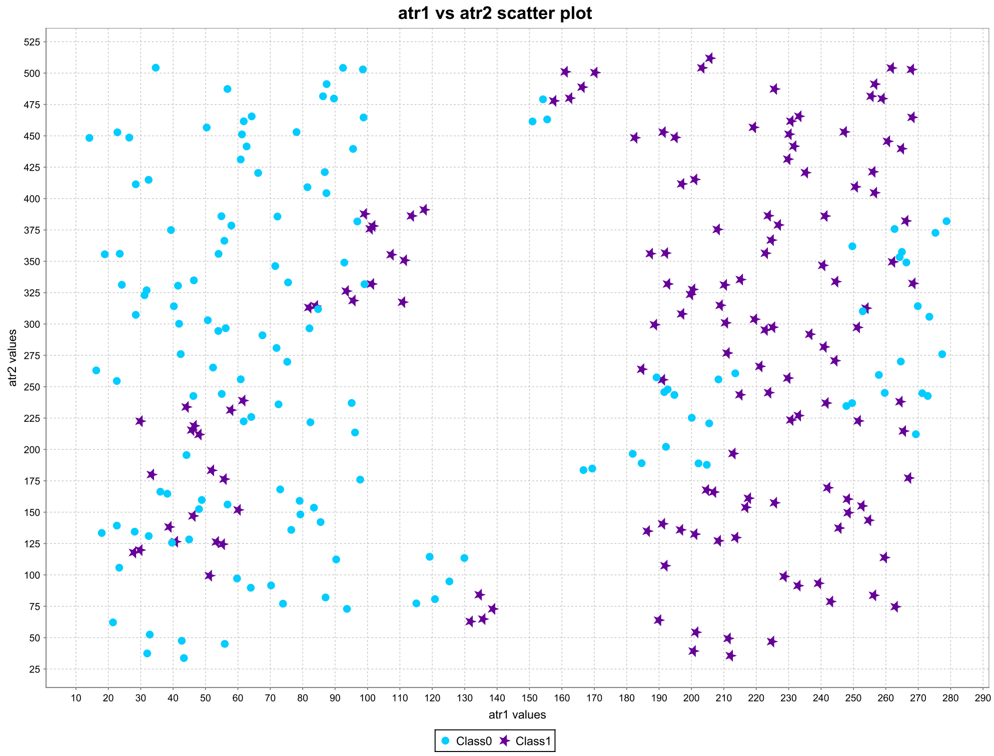


Fig. 4. $F1 = 0.6094$.

records the results of correctly and incorrectly recognized examples of each class.

Traditionally, the accuracy rate (Eq. (1)) has been the most commonly used empirical measure. However, in the framework of imbalanced datasets, accuracy is no longer a proper measure, as it does not distinguish between the numbers of correctly classified examples of different classes. Therefore, it may lead to erroneous conclusions, i.e., a classifier achieving an accuracy of 90% in a dataset with an IR value of 9, is not accurate if it classifies all examples as negatives.

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

In imbalanced domains, the evaluation of the classifiers' performance must be carried out using specific metrics to take into account the class distribution. Specifically, a well-known approach to producing an evaluation criteria in an imbalanced scenario is to use the Receiver Operating Characteristic (ROC) graphic [7]. This graphic allows the trade-off between the benefits (TP_{rate}) and costs (FP_{rate}) to be visualized, and it thus evidences that any classifier is unable to increase the number of true positives without also increasing the false positives.

The Area Under the ROC Curve (AUC) [33] corresponds to the probability of correctly identifying which of the two stimuli is noise and which is signal plus noise. AUC provides a single measure of a classifier's performance for evaluating which model is better on average. Fig. 6 shows how to build the ROC space plotting on a two-dimensional chart the TP_{rate} (Y-axis) against the FP_{rate} (X-axis). Points in $(0,0)$ and $(1,1)$ are trivial classifiers in which the predicted class is always the negative and positive respectively. By contrast, $(0,1)$ point represents the perfect classification. The AUC measure is computed by obtaining the area of the graphic:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (2)$$

3. Addressing overlapping by feature weighting with fuzzy rule based classification systems

In this paper, our main contribution is the development of a model for using feature weighting in combination with FRBCS. This synergy is expected to improve the classification ability in those imbalanced datasets which also suffer from overlapping between their classes. The final aim of this approach is to optimize the feature weights, so that each variable of the problem has a different significance for the final inference in the classification step. Hence, we consider limiting the influence of those features that may hinder the discrimination process between the classes in an imbalanced classification scenario. This process will be carried out by means of a genetic process, leading to an EFS [11,17].

In order to describe our proposal, we will first recall some basic concepts of FRBCS, focusing on the fuzzy inference mechanism, from which we will implement our feature weighting approach (Section 3.1). Then, we will introduce the working procedure of our approach, in which the weights for each variable are optimized to achieve the best results in each context (Section 3.2). We will present the details of the selected FRBCS we will use to test our methodology, namely the FARC-HD algorithm [1] (Section 3.3). Finally, for the sake of showing the goodness of our approach, we will show a graphical example of the working procedure of the feature weighting (Section 3.4).

3.1. Preliminary concepts for FRBCS

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

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

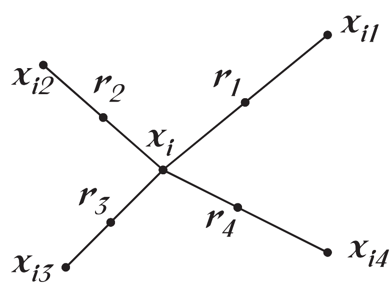


Fig. 5. An illustration of how to create the synthetic data points in the SMOTE algorithm.

Table 1
Confusion matrix for a two-class problem.

	Positive prediction	Negative prediction
Positive class	True Positive (TP)	False Negative (FN)
Negative class	False Positive (FP)	True Negative (TN)

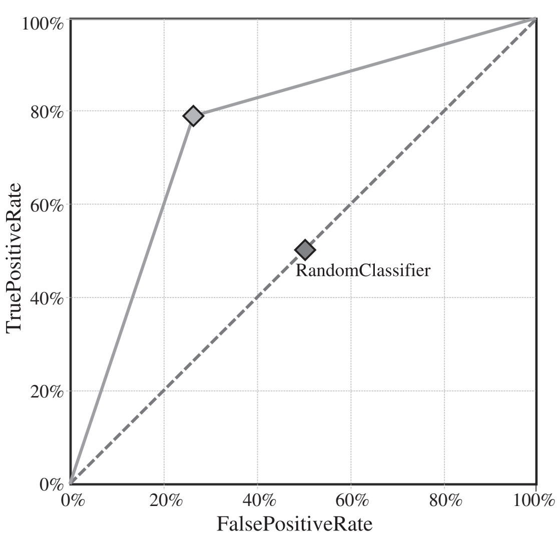


Fig. 6. Example of an ROC plot. Two classifiers' curves are depicted: the dashed line represents a random classifier, whereas the solid line is a classifier which is better than the random classifier.

$$\begin{aligned}&\text{Rule } R_j : \text{If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn}\\&\text{then Class} = C_j \text{ with } RW_j\end{aligned} \tag{3}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label, and RW_j is the rule weight [35]. We use triangular MFs as antecedent fuzzy sets.

When a new pattern x_p is selected for classification, then the steps of the fuzzy reasoning method are as follows:

1. **Matching degree**, that is, the strength of activation of the if-part for all rules in the Rule Base with the pattern x_p. In order to carry out this computation, a conjunction operator γ shall be applied. This operator is used to combine the membership degrees for every variable of the example, which were obtained by means of the μ function. Traditionally, a T-norm is selected for this purpose, although any aggregation operator can be employed:

$$\mu_{A_j}(x_p) = \gamma(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \quad j = 1, \ldots, L \tag{4}$$

2. **Association degree**. To compute the association degree of the pattern x_p with the M classes according to each rule in the Rule Base. To this end, a combination operator h is applied in order to combine the matching degree with the rule weight (RW). In our case, this association degree only refers to the consequent class of the rule (i.e. $k = Class(R_j)$).

$$b_j^k = h\left(\mu_{A_j}(x_p), RW_j^k\right), \quad k = 1, \ldots, M; \qquad j = 1, \ldots, L \tag{5}$$

3. **Pattern classification soundness degree for all classes**. We use an aggregation function f, which combines the positive degrees of association calculated in the previous step.

$$Y_k = f\left(b_j^k, j = 1, \ldots, L \text{ and } b_j^k > 0\right), \quad k = 1, \ldots, M \tag{6}$$

4. **Classification**. We apply a decision function F over the soundness degree of the system for the pattern classification for all classes. This function will determine the class label l corresponding to the maximum value.

$$F(Y_1, \ldots, Y_M) = \arg\max(Y_k), \quad [k = 1, \ldots, M] \tag{7}$$

Where L denotes the number of rules in the Rule Base and M the number of classes of the problem ($M = 2$ in our current case).

3.2. Learning the optimal weights for the problem's variables

Our approach is based on a mechanism which modifies the bias for each variable during the fuzzy inference process, which was previously introduced in Section 3.1.

We will modify Step (1), which was devoted to the computation of the matching degree. In order to take into account the different significance associated with each variable, we will perform an adjustment of function γ. Specifically, we will include a power weighting function Ω so that:

$$\Omega = (\mu_{A_j}(x_p))^w \text{ with } w = [0, 1] \tag{8}$$

so that Eq. (4) will have the following expression:

$$\mu_{A_j}^{\Omega}(x_p) = \gamma(\mu_{A_{j1}}^{w_1}(x_{p1}), \ldots, \mu_{A_{jn}}^{w_n}(x_{pn})), \quad j = 1, \ldots, L; \tag{9}$$

Fig. 7 depicts the influence of the value of w for the different variables in the Ω function, comparing this value with the initial membership function computed with μ. Notice that when $w_j = 1$ the original value obtained by the membership function is not modified at all.

The contrary case occurs when $w_j = 0$, i.e. the current feature will have no influence throughout the fuzzy reasoning method, since $x^0 = 1\ \forall x$. Intermediate values will make the membership degree of the example for that variable to be increased or decreased depending on its closeness to the maximum value 1. In summary, a high value for w_j should be set for those significant variables that truly contribute to the classification, whereas those redundant or noisy features might consider a lower value of w_j.

The estimation of these parameters is not trivial since, as pointed out above, their values directly affect the prediction of the final class. For a proper definition of these weights, an optimization process must be carried out.

Among the different techniques that can be used for this search procedure, genetic algorithms excel due to their ability to perform a good exploration and exploitation of the solution space [15]. Of all the available evolutionary procedures, the CHC technique [16] is recommended as it presents a good trade-off between diversity and convergence, making it a good choice for problems with

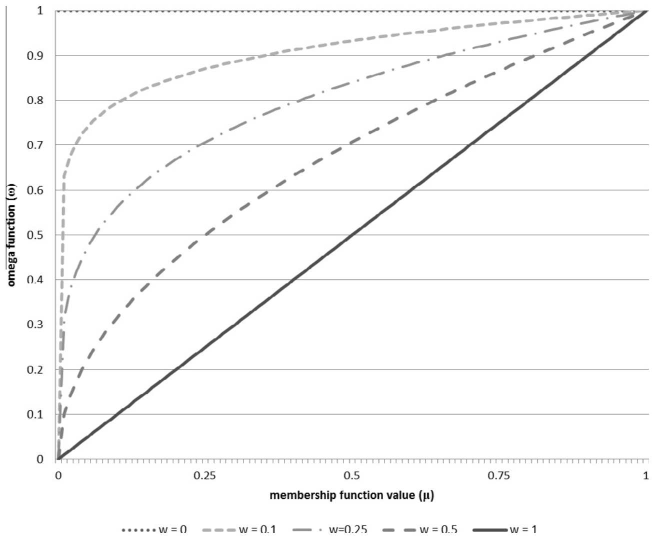


Fig. 7. Graphical representation of the Ω function. Different values of w will alter the final computation of the membership function. When w is closer to 1, the membership function will have a closer value to its original. The membership degree will tend to 1 when w is closer to 0.

complex search spaces. It has also shown good results in the scenario of imbalanced datasets [18]. The features of this evolutionary model are presented below:

- **Coding scheme**: We have a set of real parameters to be optimized (w_j, with $j = 1, \ldots, n$), where the range in which each one varies is [0, 1], and n stands for the number of attributes. Thus, the chromosome will be an array with the following structure: $W_j = (w_1, \ldots, w_n)$
- **Initial Gene Pool**: In the first place, we include an individual with all genes with value 1, so that it represents the standard case. Additionally, we have generated "ad-hoc" as many chromosomes as the problem has variables. Each of these individuals will have a value $w_V = 1$ for its corresponding variable V, with the rest of genes $w_j = 0$. In this way, we hope to include information within the population to reduce the global weight of the variables throughout the crossover operations between the individuals. The remaining chromosomes of the population will be initialized with random values between 0 and 1.
- **Chromosome Evaluation**: First, a fuzzy knowledge base is extracted according to the current values of the weights for the Ω function for each individual. It is expected to obtain a different set of rules for different values of w_j. Then, this knowledge base is used to classify the training set, and the obtained AUC value is selected as fitness for the chromosome. We must point out that the bias of the search towards the sole coverage of the majority class examples is avoided by the use of this metric.
- **Crossover Operator**. The crossover operator is based on the concept of environments (the offspring are generated around their parents). These kinds of operators present a good cooperation when they are introduced within evolutionary models, forcing the convergence by pressure on the offspring (as the case of CHC). Specifically, we consider the PC-BLX-α operator [29], which allows the offspring genes to be around the genes of one parent.
- **Restarting Approach**. To get away from local optima, this algorithm uses a restarting approach since it does not apply mutation during the recombination phase. Therefore, when the threshold value is lower than zero, all the chromosomes are regenerated randomly to introduce new diversity to the search. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm as in the elitist scheme

3.3. Integration of the feature weighting process into FARC-HD algorithm

In this paper we have made use of a robust FRBCS known as Fuzzy Association Rule-based Classification for High-Dimensional problems (FARC-HD) [1]. This algorithm is based on association discovery, a commonly used technique in data mining for extracting interesting knowledge from large datasets [26] by means of finding relationships between the different items in a database [60]. The integration between association discovery and classification leads to precise and interpretable models.

FARC-HD is aimed at obtaining an accurate and compact fuzzy rule-based classifier with a low computational cost. In short, this method is based on the following three stages:

Stage 1 *Fuzzy association rule extraction for classification*: A search tree is employed to list all possible frequent fuzzy item sets and to generate fuzzy association rules for classification, limiting the depth of the branches in order to find a small number of short (i.e., simple) fuzzy rules.

Stage 2 *Candidate rule pre-screening*: After the rule generation, the size of the rule set may be too large to be interpretable by the end user. Therefore, a pre-selection of the most interesting rules is carried out by means of a "subgroup discovery" mechanism based on an improved weighted relative accuracy measure (wWRAcc') [37].

Stage 3 *Genetic rule selection and lateral tuning*: Finally, in order to obtain a compact and accurate set of rules within the context of each problem, an evolutionary process will be carried out in a combination for the selection of the rules with a tuning of the membership function.

In order to integrate the genetic optimization process for feature weighting with the FARC-HD algorithm, we proceed in the following way: once we have set up the values for each chromosome, i.e. the weights for the variables of the problem, a complete rule set will be learnt by means of the association fuzzy rule mining, which corresponds to *Stage 1* and *Stage 2* (rule extraction and pre-screening) of the FARC-HD approach, similar to a wrapper methodology.

We acknowledge that, by merging the fuzzy rule discovery into the fitness computation of the feature weighting approach, we might lose some efficiency in the whole process with respect to just combining our proposal with the genetic algorithm of *Stage 3* of the FARC-HD technique. However, our intention is twofold: in the first place, if we proceed in the latter way the size of the chromosome will become too large to converge into an optimal solution, even if more evaluations are performed. Moreover, there is a clear dependency between the weights of the variables and the learning of the rule set. Hence, obtaining in the first place the weights of the variables, and then discovering the fuzzy rules, we make sure that this constraint is maintained. Furthermore, since we omit the last genetic stage within the evaluation part of our methodology, the number of total evaluations remains admissible for this type of approach.

As a summary, Fig. 8 depicts the complete learning scheme of the EFS developed by the combination of feature weighting and FARC-HD.

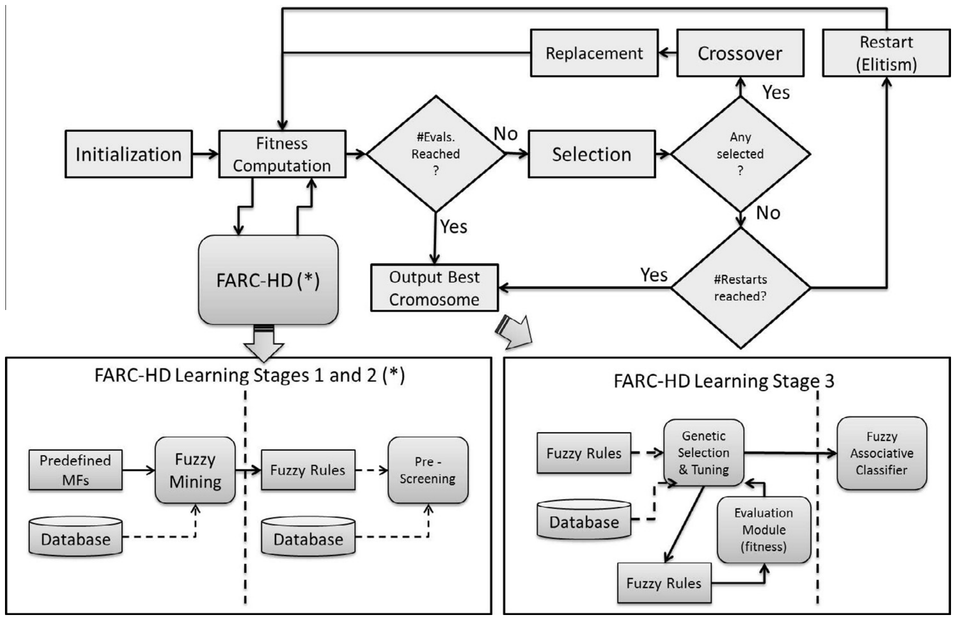
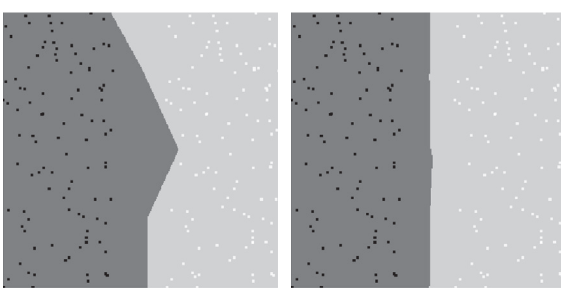
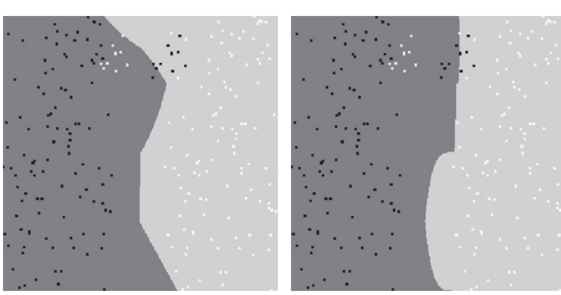


Fig. 8. Scheme of the feature weighting approach with FARC-HD.



(a) FARC-HD classification (b) FARC-HD-FW classification

Fig. 9. Borderline areas obtained by the FRBCSs ($F1 = 12.5683$).



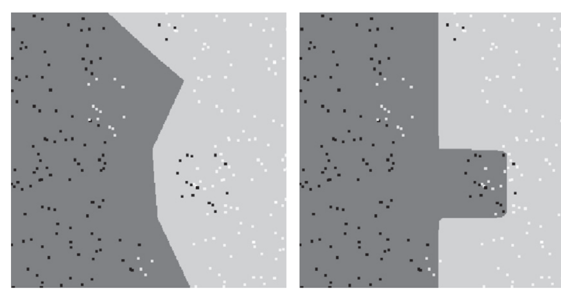
(a) FARC-HD classification (b) FARC-HD-FW classification

Fig. 10. Borderline areas obtained by the FRBCSs ($F1 = 5.7263$).

3.4. Graphical analysis of the feature weighting process on a dataset with overlapping

The goodness of our approach in the scenario of overlapping data will be shown by means of a simple graphical example. We have created a 2-variable synthetic dataset, which corresponds to that shown in Section 2.2, setting up different degrees of overlapping according to the F1 measure.

Then, we have run the standard FARC-HD algorithm, and FARC-HD with Feature Weighting (FARC-HD-FW) with the aim of contrasting the differences both at the performance level (AUC values), and the borderline areas learnt by the classifier. Figs. 9–12 depict the obtained results, whereas Table 2 shows the quality of the methodologies in terms of AUC.



(a) FARC-HD classification (b) FARC-HD-FW classification

Fig. 11. Borderline areas obtained by the FRBCSs ($F1 = 3.3443$).

In accordance with the results previously shown, we may observe that our current approach has two main advantages in this classification scenario:

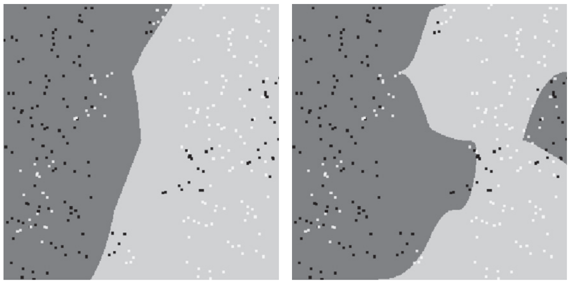
1. It achieves a higher average value for the AUC metric, especially when the degree of overlapping increases.
2. The borderline areas generated by the FRBCSs' rules are simpler and more compact, which favors a better generalization ability.

4. Experimental framework

In this section we first provide details of the real-world binary-class imbalanced problems chosen for the experiments (subSection 4.1). Then, we will describe the learning algorithms selected for this study and their configuration parameters (4.2). Finally, we present the statistical tests applied to compare the results obtained with the different classifiers (subSection 4.3).

4.1. Benchmark data

Table 3 shows the benchmark problems selected for our study, in which the name, number of examples, number of attributes, IR



(a) FARC-HD classification (b) FARC-HD-FW classification

Fig. 12. Borderline areas obtained by the FRBCSs ($F1 = 0.6094$).

Table 2
AUC values for the synthetic problems in FARC-HD and FARC-HD-FW. In boldface the best result in AUC is stressed.

F1-value	FARC-HD	FARC-HD-FW
12.5683	1.0000	**1.0000**
5.7263	.9234	**.9321**
3.3443	.8594	**.8838**
0.6094	.7342	**.7942**
Avg	.8792 ± .0974	**.9025 ± .0749**

and F1 metric (overlapping) is shown. Datasets are ordered with respect to their degree of overlap, from which we can divide them into two disjoint groups with a similar size: datasets with a low degree of overlapping (left column), with their F1 metric higher than 1.5, and high overlapping datasets (right column), identified by having a value below 1.5 for their F1 metric. Additionally, we must draw attention to a subset of 26 problems that share both a high degree of imbalance and overlapping, as these can be defined as the most difficult problems to overcome.

As highlighted throughout this paper, the estimates of the AUC measure are obtained by means of a standard Stratified Cross-Validation. The number of folds selected in both cases is 5. This value is set up with the aim of having enough positive class instances in the different folds, hence avoiding additional problems in the data distribution, especially for highly imbalanced datasets.

We must point out that the original dataset partitions with 5-fold-cross-validation employed in this paper are available for download at the KEEL dataset repository [2] so that any interested researcher can use the same data for comparison. In this case, multi-class problems were modified to obtain two-class imbalanced problems, defining the joint of one or more classes as positive and the joint of one or more classes as negative, as defined in the name of the dataset.

4.2. Algorithms and parameters

In order to validate the robustness of the feature weighting strategy with FRBCS, we will check our experimental results versus the FARC-HD algorithm and FARC-HD with feature selection. In this way, we expect to show the goodness of our approach by enhancing the performance of the baseline classifier. We may also make clear the differences between the use of feature selection and feature weighting in the scenario of overlapping for imbalanced data. We must point out that the components of the feature selection procedure are exactly the same as in the case of the feature weighting, but the chromosome has a binary representation instead of a real one.

Additionally, as a state-of-the-art classifier we have made use of C4.5 [49]. The C4.5 learning algorithm constructs a top-down

Table 3
Summary of imbalanced datasets used.

Name	#Ex.	#Atts.	IR	F1	Name	#Ex.	#Atts.	IR	F1
iris0	150	4	2.00	16.8200	ecoli0146vs5	280	6	13.00	1.3400
shuttle0vs4	1829	9	13.87	12.9700	yeast2vs8	482	8	23.10	1.1420
shuttle2vs4	129	9	20.50	12.1300	ecoli0347vs56	257	7	9.28	1.1300
ecoli0vs1	220	7	1.86	9.7520	vehicle0	846	18	3.23	1.1240
yeast5	1484	8	32.78	4.1980	ecoli01vs235	244	7	9.17	1.1030
new-thyroid2	215	5	4.92	3.5790	yeast05679vs4	528	8	9.35	1.0510
new-thyroid1	215	5	5.14	3.5790	glass06vs5	108	9	11.00	1.0490
wisconsin	683	9	1.86	3.5680	glass5	214	9	22.81	1.0190
glass0123vs456	214	9	3.19	3.3240	ecoli067vs35	222	7	9.09	0.9205
ecoli4	336	7	13.84	3.2470	ecoli0267vs35	244	7	9.18	0.9129
yeast3	1484	8	8.11	2.7510	ecoli0147vs56	332	6	12.28	0.9124
ecoli1	336	7	3.36	2.6500	yeast4	1484	8	28.41	0.7412
vowel0	988	13	10.10	2.4580	yeast0256vs3789	1004	8	9.14	0.6939
glass6	214	9	6.38	2.3910	glass0	214	9	2.06	0.6492
ecoli0137vs26	281	7	39.15	2.3020	abalone918	731	8	16.68	0.6320
yeast6	1484	8	39.15	1.9670	pima	768	8	1.90	0.5760
led7digit02456789vs1	443	7	10.97	1.9570	abalone19	4174	8	128.87	0.5295
glass016vs5	184	9	19.44	1.8510	ecoli0147vs2356	336	7	10.59	0.5275
ecoli2	336	7	5.46	1.8260	page-blocks0	5472	10	8.77	0.5087
segment0	2308	19	6.01	1.7980	glass2	214	9	10.39	0.3952
ecoli067vs5	220	6	10.00	1.6920	vehicle2	846	18	2.52	0.3805
yeast02579vs368	1004	8	9.14	1.6350	yeast1289vs7	947	8	30.56	0.3660
ecoli034vs5	200	7	9.00	1.6320	yeast1vs7	459	8	13.87	0.3534
ecoli0234vs5	202	7	9.10	1.6180	glass0146vs2	205	9	11.06	0.3487
ecoli046vs5	203	6	9.15	1.6030	yeast0359vs78	506	8	9.12	0.3113
ecoli0346vs5	205	7	9.25	1.5950	glass016vs2	192	9	10.29	0.2692
ecoli3	336	7	8.19	1.5790	yeast1	1484	8	2.46	0.2422
yeast2vs4	514	8	9.08	1.5790	glass1	214	9	1.82	0.1897
page-blocks13vs4	472	10	15.85	1.5470	vehicle3	846	18	2.52	0.1855
glass04vs5	92	9	9.22	1.5420	haberman	306	3	2.68	0.1850
glass4	214	9	15.47	1.4690	yeast1458vs7	693	8	22.10	0.1757
ecoli01vs5	240	6	11.00	1.3900	vehicle1	846	18	2.52	0.1691
cleveland0vs4	177	113	12.62	1.3500	glass015vs2	172	9	9.12	0.1375

decision tree by using the normalized information gain (difference in entropy) that results from choosing an attribute to split the data. The attribute with the highest normalized information gain is the one used to make the decision.

Finally, we will complement our study by using several EFS approaches developed for both classical and imbalanced classification. Next we detail the features of these approaches:

- The use of the linguistic 2-tuples representation [30], which allows the lateral displacement of the labels considering only one parameter (slight displacements to the left/right of the original MFs). Specifically, we will consider the Global Tuning of the Semantics (GTS) approach, in which the tuning is applied to the level of linguistic partition. The pair (X_i, label) takes the same tuning value in all the rules where it is considered. As FRBCS we will use Chi et al.'s learning procedure [10] as it has shown a good behavior in the scenario of highly imbalanced datasets [18].
- The GP-COACH-H algorithm [40] is an FRBCS with different granulation levels that integrates rule selection and the 2-tuples tuning approach to improve the performance in imbalanced data-sets. It is based on the standard GP-COACH algorithm [6], a genetic programming scheme for the learning of the fuzzy rules, specifically a genetic cooperative-competitive learning approach in which the whole population represents the Rule Base obtained.
- IIVFDT [50] is composed of the baseline fuzzy ID3 algorithm, from which linguistic labels are modeled with Interval-Valued Fuzzy Sets [8]. To do this, a parametrized construction method of Interval-Valued Fuzzy Sets is defined, whose length represents the ignorance degree. Finally, an evolutionary tuning step is applied to compute the optimal ignorance degree for each Interval-Valued Fuzzy Set.

Next, we detail the parameter values for the different learning algorithms selected in this study, which have been set considering the recommendation of the corresponding authors:

1. **C4.5**
   For C4.5 we have set a confidence level of 0.25, the minimum number of item-sets per leaf was set to 2 and the application of pruning was used to obtain the final tree.
2. **FARC-HD**
   First, we have selected 5 labels per variables for the fuzzy sets, product t-norm as conjunction operator and additive combination for the inference procedure. As specific parameters of the learning stage, we have set up the minimum support to 0.05 and the minimum confidence to 0.8. Finally, we have fixed the maximum depth of the tree to a value of 3, and the k parameter for the pre-screening to 2. For more details about these parameters, please refer to [1].
3. **Chi et. al's algorithm**
   The configuration of this FRBCS consists of 3 labels per variable, product T-norm as conjunction operator, together with the penalized certainty factor approach [35] for the rule weight and fuzzy reasoning method of the winning rule.
4. **GP-COACH-H**
   We use a minimum t-norm and a maximum t-conorm. The rule weight is the standard certainty factor, and as the fuzzy reasoning method we employ the additive combination. The number of fuzzy labels has been set to 5 for low granularity rules and 9 for high granularity rules.
5. **IIVFDT**
   In this case, the number of labels per variable has been set to 3, and the conjunction operator will be the IV product t-norm [50]. Specific parameters of the fuzzy decision tree are the evidence significance level (0.4) and the truth level threshold (0.95).
   The ignorance weighting function for interval-valued fuzzy sets has been defined as $g(x) = 2 \cdot \min(x, 1 - x)$, whereas the optimization parameters δ and γ have been initialized to 0.5. For more details about these parameters, please refer to [50].
6. **CHC optimization process**
   Every EFS algorithm used in the experimental study contains a CHC optimization process. In all cases, the number of individuals has been set to 50 chromosomes. The crossover mechanism is the PC-BLX-α with α equal to 1.0, and 30 bits will be used to translate the real numbers into a binary string for computing the hamming distance in the incest prevention mechanism. However, the number of evaluations vary depending on the algorithm that is being considered:
   - For FARC-HD-FW, a total of 16,000 evaluations will be carried out throughout the genetic process, i.e. 1000 for the weight adjustment of the variables, and another 15,000 for the lateral tuning and rule selection of the final Rule Base (Stage 3 of FARC-HD). The original FARC-HD approach has been set up with exactly the same parameters for Stage 3.
   - GP-COACH-H uses 10,000 evaluations.
   - IIVFDT and GTS consider $5000 \cdot n$ evaluations, with n the number of input variables.

Regarding the SMOTE preprocessing technique, we will consider the *5-nearest neighbors of the positive class* to generate the synthetic samples, and *balancing both classes to the 50% distribution*.

We must also point out that most of these algorithms are available within the KEEL software tool [3], whereas a beta implementation (also in KEEL format) of the FARC-HD-FW is available for download at http://www.keel.es/beta/farchd-fw.zip.

4.3. Statistical tests for performance comparison

In this paper we use the hypothesis testing techniques to provide statistical support for the analysis of the results [22]. Specifically, we will use non-parametric tests, due to the fact that the initial conditions that guarantee the reliability of the parametric tests may not be satisfied, causing the statistical analysis to lose credibility with these types of tests [23,12]. Any interested reader can find additional information on the Website http://sci2s.ugr.es/sicidm/.

First of all, we consider the method of aligned ranks of the algorithms in order to show graphically how good a method is with respect to its partners. In order to compute this ranking, the first step 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 a descending way and, finally, we average the rankings obtained by each algorithm. In this manner, the algorithm which achieves the **lowest average ranking** is the best one.

The Friedman aligned test [22] will be used to check whether there are significant differences among the results, and the Holm post hoc test [32] in order to find which algorithms reject the hypothesis of equality with respect to a selected control method in a $1 \cdot n$ comparison. We will compute the adjusted p-value (APV) associated with each comparison, which represents the lowest level of significance of a hypothesis that results in a rejection. This value differs from the standard p-value in the sense that it determines univocally whether the null hypothesis of equality is rejected at a significance level α. This fact eases the comparison of the algorithms, as it is no longer necessary to contrast it with the $alpha/i$ value of a standard statistical results table.

5. Experimental study

In this section we will show the goodness of our proposal under three different cases of study:

1. First, we will make a global analysis for all benchmark datasets, in order to determine the goodness of our approach in a general case (Section 5.1).
2. Then, we will focus on the main issue of our study, which is related to those cases where datasets have a high degree of overlap between their classes (Section 5.2).
3. Finally, we propose to investigate the most difficult case, i.e. when the classification problem includes an extremely skewed distribution together with a high degree of overlapping (Section 5.3).

As a complementary study, we will contrast the results of FARC-HD-FW versus several EFSs in order to provide greater support to the findings extracted throughout this analysis (Section 5.4).

The complete experimental results for every dataset introduced in the experimental framework, are shown in Table 4. In this table, FARC-HD-FW stands for our proposed methodology, FARC-HD is the standard FARC-HD algorithm, and FARC-HD-FS is the methodology with feature selection. The reader must recall that every single algorithm uses SMOTE preprocessing prior to its application.

5.1. On the behavior of feature weighting for imbalanced classification

By observing Table 4 we note that our feature weighting approach obtains the highest AUC test result in almost half of the problems (29 out of the 66 datasets), and that it only achieves the worst value in 10% of the problems (6 out of 66 datasets) with respect to the algorithms of comparison.

This good behavior is emphasized if we take a closer look at the average values of the performance (Table 5), in which both the AUC results for the test partitions and the APV of the Holm test are included. We might observe that the differences in AUC in favor of our approach are clear in all cases, especially versus the C4.5 algorithm, which presents a higher degree of overfitting. Furthermore, when contrasting the results for the FARC-HD and FARC-HD-FS algorithms, it also shows the highest value for the training partitions, thus stressing the correctness of the learning stage that has been implemented.

In order to obtain well founded conclusions of these results, we will thoroughly analyze these four scenarios by means of a statistical study. This study will allow us to determine whether our proposal outperforms the remaining approaches, and therefore to support the quality of this methodology under the different cases of study proposed.

In first place, Fig. 13 shows the average ranks for our feature weighting approach and the three algorithms for comparison. Within these pictures, the relative differences among the algorithms are depicted, and interesting conclusions can be extracted from them. On the one hand, we can highlight the robustness of feature weighting with respect to the other methodologies, as there is a clear gap for the average ranking in each of the four cases. On the other hand, we must also stress that the distance between FARC-HD-FW and FARC-HD-FS is higher than that between FARC-HD-FS and FARC-HD. This latter fact indicates the strength of our initial premise with regard to the significance of the use of feature weighting rather than just feature selection.

In the last part of our analysis, we will carry out some non-parametric tests that will conclude our study, aiming to find statistical differences between our proposal and the remaining methodologies. The first step is to compute whether significant differences are found among the results. For this purpose, we obtain the p-value for the Friedman aligned test and, if it is below 0.05, then it will confirm the presence of significant differences among the results. In this current case, i.e. when considering all datasets as a whole, this p-value is near zero ($4.3E^{-12}$). Since FARC-HD-FW is the algorithm with the best ranking, it will be used as the control method in the statistical test.

A post hoc test is then carried out to establish which approaches are outperformed by FARC-HD-FW. The *APVs* obtained by a Holm test were previously shown in Table 5.

We can observe that, in all cases, the null hypothesis of equality is rejected. This implies that our proposal statistically outperforms all the algorithms of comparison and is an appropriate tool for solving classification with imbalanced datasets in a general scenario. We must also stress that the APVs associated with the comparisons support our conclusions with a high degree of confidence.

5.2. Analysis of the performance with a high degree of overlapping

We focus now on the central part of our work, which is the case study for those datasets with high overlapping ($F1 < 1.5$). For this purpose, Table 6 shows the average results (together with the standard deviation values) in the test partitions for the four algorithms considered in this study. We must point out that a number of 36 problems has been selected for this case study. Please refer to Table 3 to check the specific datasets that fulfill this condition.

Table 6 shows the goodness of the feature weighting approach, as it achieves the highest AUC value in test when we are dealing with high overlapping. Additionally, we can observe that the standard deviation is small, so that the quality of our proposal is related to a high performance for this group of datasets on average, and not just simply for a few outlier problems.

As was carried out in the previous section, we also depict the average ranking of the algorithms to enable a visual comparison among them. Fig. 14 shows the values of the Friedman aligned test for every algorithm, so that the one with the lowest ranking is the one that obtains the best result on average for most of the datasets. The behavior in this case is identical to the one shown when all datasets were considered. FARC-HD-FW is the best-ranked method with a significant difference with respect to the other methods. Absolute values of this ranking allow us to extract some interesting conclusions: the gap between FARC-FW and FARC-FS is much higher than for FARC-FS versus the standard FARC algorithm. This fact demonstrates our initial hypothesis in favor of the weighting approach rather than a simple selection/removal of the variables.

The p-value associated with the Friedman aligned test indicates the presence of significant differences among the results, as its value is again near to zero ($2.1965E^{-7}$). The results from this test were included in Table 6, and also supports our previous conclusions. First, our feature weighting approach has shown to be statistically better when compared with both FARC-HD and C4.5 algorithms. When we analyze the results of FARC-HD-FW and FARC-HD-FS, we must state that we achieve a "low" APV associated with this comparison, very near to the degree of confidence of the 95% we have set in our experimental framework. This fact, in conjunction with the performance and average ranking indicators, stress the good behavior for our proposal in contraposition to the simple feature selection approach.

5.3. Classification in the hardest scenario: high imbalance and overlapping

Finally, we aim to investigate a very interesting case study, i.e. classification in the presence of both high imbalance and high overlapping. This scenario represents the most difficult case for

Table 4
Complete training and test results (AUC metric). FARC-HD-FW is the proposed methodology (Feature Weighting), FARC-HD is the baseline FARC-HD algorithm, FARC-HD-FS includes Feature Selection, C4.5 is the C4.5 decision tree. GP-COACH-H stands for the GP hierarchical approach, GTS is the Chi et al.'s algorithm with 2-tuples tuning, and IIVFDT is the interval-valued fuzzy decision tree. All datasets are preprocessed using SMOTE. In boldface the best result in AUC is stressed.

Dataset	IR	F1	FARC-HD-FW		FARC-HD		FARC-HD-FS		C4.5		GP-COACH-H		GTS		IIVFDT	
			Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst
iris0	2.00	16.8200	1.000	**1.000**	1.000	**1.000**	1.000	.9950	1.000	.9900	1.000	**1.000**	1.000	**1.000**	1.000	**1.000**
shuttle0vs4	13.87	12.9700	1.000	**1.000**	1.000	1.000	.9999	.9994	.9999	.9997	1.000	1.000	1.000	.9912	1.000	.9996
shuttle2vs4	20.50	12.1300	1.000	**1.000**	1.000	1.000	.9980	.9960	1.000	.9958	1.000	.9920	1.000	.9838	1.000	.8841
ecoli0vs1	1.86	9.7520	.9919	.9691	.9919	.9702	.9886	.9730	.9927	.9761	.9870	**.9867**	.9910	.9524	.9860	.9667
yeast5	32.78	4.1980	.9872	.9149	.9872	.9354	.9666	.9177	.9777	.9337	.9724	.9448	.9803	**.9590**	.9818	.9458
newthyroid2	4.92	3.5790	.9993	.9718	.9986	.9488	.9986	.9603	.9957	.9631	.9993	.9774	.9993	**.9917**	.9979	.9802
newthyroid1	5.14	3.5790	.9993	.9659	.9993	.9373	.9965	.9373	.9922	.9802	.9986	.9774	.9986	.9718	.9965	**.9917**
wisconsin	1.86	3.5680	.9852	.9577	.9864	.9578	.9782	.9530	.9832	.9545	.9831	**.9768**	.9913	.9313	.9851	.9676
glass0123vs456	3.19	3.3240	.9865	**.9254**	.9741	.8710	.9684	.9023	.9916	.8832	.9682	.9001	.9803	.9002	.9762	.9184
ecoli4	13.84	3.2470	.9952	.9358	.9937	.9029	.9869	.9060	.9773	.8044	.9937	**.9389**	.9917	.9262	.9683	.9076
yeast3	8.11	2.7510	.9491	**.9270**	.9481	.9212	.9420	.9240	.9565	.8869	.9421	.9224	.9487	.9163	.9359	.9085
ecoli1	3.36	2.6500	.9443	.8523	.9461	.9060	.9277	.8724	.9517	**.9162**	.9304	.8630	.9463	.8529	.9257	.8715
vowel0	10.10	2.4580	1.000	.9706	1.000	.9600	.9997	.9706	.9967	.9433	.9947	.9489	.9994	.9805	.9976	**.9844**
glass6	6.38	2.3910	.9863	.9225	.9709	**.9365**	.9678	.9009	.9966	.8450	.9739	.8857	.9821	.8577	.9642	.9036
ecoli0137vs26	39.15	2.3020	.9959	.8299	.9845	.8191	.9795	.8118	.9669	.8136	.9322	.8295	.9909	.8190	.9147	**.8930**
yeast6	39.15	1.9670	.9258	.8681	.9158	.8184	.8855	**.8737**	.9242	.8280	.9845	.8116	.9361	.8480	.9117	.8683
led7digit02456789vs1	10.97	1.9570	.9326	.8872	.8989	.8957	.8989	.8957	.9300	.8832	.9162	**.9056**	.9514	.8240	.9618	.8891
glass016vs5	19.44	1.8510	.9950	.8743	.9879	.8686	.9857	.8157	.9914	**.9714**	.9921	.8686	.9929	.8571	.9586	.7800
ecoli2	5.46	1.8260	.9646	.9147	.9672	.9120	.9433	.8745	.9815	.8921	.9681	.9134	.9374	.8753	.9489	**.9369**
segment0	6.01	1.7980	.9994	**.9937**	.9973	.9909	.9937	.9871	.9987	.9927	.9984	.9884	.9978	.9883	.9935	.9851
ecoli067vs5	10.00	1.6920	.9900	.8350	.9663	.7950	.9663	.7950	.9900	.8250	.9850	**.8700**	.9688	.8550	.9506	.8550
yeast02579vs368	9.14	1.6350	.9340	.8927	.9180	.8916	.9180	.8916	.9800	**.9171**	.9304	.9121	.9332	.8957	.9257	.8985
ecoli034vs5	9.00	1.6320	.9965	.9194	.9910	**.9306**	.9910	**.9306**	.9882	.8583	.9833	.8750	.9875	.8611	.9785	.9167
ecoli0234vs5	9.10	1.6180	.9938	**.9254**	.9876	.9030	.9876	.9030	.9918	.8974	.9966	.8613	.9863	.8975	.9725	.9114
ecoli046vs5	9.15	1.6030	.9932	.8811	.9925	.8727	.9925	.8727	.9877	.8729	.9952	.8921	.9904	.9255	.9816	**.9338**
ecoli0346vs5	9.25	1.5950	.9980	.8757	.9932	**.9257**	.9932	**.9257**	.9899	.8703	.9993	.8899	.9939	.8784	.9743	.9088
ecoli3	8.19	1.5790	.9579	.8761	.9464	.8452	.9377	.8655	.9631	.7755	.9615	.8854	.9638	**.9150**	.9398	.8595
yeast2vs4	9.08	1.5790	.9559	.9058	.9505	.9144	.9525	.9181	.9778	.8620	.9652	**.9313**	.9551	.8788	.9120	.8942
pageblocks13vs4	15.85	1.5470	.9983	.9755	.9972	.9710	.9941	.9532	.9975	**.9955**	.9994	.9498	.9992	.9476	.9645	.9518
glass04vs5	9.22	1.5420	.9970	.9702	.9894	**.9882**	.9894	**.9882**	.9910	.9816	.9910	.9452	.9894	.8518	.9638	.8757
glass4	15.47	1.4690	.9994	.8325	.9963	.7825	.9907	.8342	.9844	.8508	.9907	.8342	.9963	**.8992**	.9714	.8776
ecoli01vs5	11.00	1.3900	.9983	.9068	.9869	.8455	.9869	.8455	.9937	**.9147**	.9977	.8977	.9932	.8932	.9869	.9045
cleveland0vs4	12.62	1.3500	.9938	.8437	.9899	.8573	.9852	.8573	.9953	.7210	.9728	**.8844**	.9937	.5992	.9821	.8299
ecoli0146vs5	13.00	1.3400	.9933	**.9462**	.9755	**.9462**	.9755	**.9462**	.9813	.8981	.9952	.9231	.9923	.8981	.9889	.8750
yeast2vs8	23.10	1.1420	.8511	**.8098**	.8824	.7859	.8647	.7848	.9125	.8066	.9938	.7784	.8560	.7609	.8474	.7978
ecoli0347vs56	9.28	1.1300	.9780	**.9119**	.9817	.8984	.9817	.8984	.9772	.8368	.9881	.8812	.9692	.8768	.9510	.8920
vehicle0	3.23	1.1240	.9791	.9456	.9663	.9262	.9645	**.9515**	.9897	.9143	.9582	.9044	.9638	.8745	.9654	.9478
ecoli01vs235	9.17	1.1030	.9780	.8691	.9680	.8450	.9680	.8450	.9685	.8041	.9847	.8518	.9744	.8405	.9574	**.8918**
yeast05679vs4	9.35	1.0510	.8722	**.8103**	.8725	.7891	.8565	.8023	.9504	.7682	.8970	.7163	.8793	.7865	.8467	.8064
glass06vs5	11.00	1.0490	.9987	**.9700**	.9937	.8547	.9937	.8547	.9830	.8227	.9975	.9197	1.000	.9147	.9849	.8300
glass5	22.81	1.0190	.9957	.7878	.9957	.7329	.9933	.7329	.9976	.8829	.9957	**.8854**	.9933	.8256	.9921	.7402
ecoli067vs35	9.09	0.9205	.9775	.8325	.9626	.7925	.9626	.7925	.9672	.8125	.9710	.8250	.9643	.8375	.9400	**.8725**
ecoli0267vs35	9.18	0.9129	.9802	.8505	.9524	.7932	.9524	.7932	.9876	.7704	.9709	**.9054**	.9654	.8152	.9530	.8677
ecoli0147vs56	12.28	0.9124	.9915	.8707	.9767	.8624	.9767	.8624	.9802	.8641	.9853	.8424	.9678	.8907	.9662	**.9123**
yeast4	28.41	0.7412	.8930	.7912	.9011	.7953	.8547	.8067	.9101	.7004	.9012	.8196	.9053	**.8242**	.8746	.8149
yeast0256vs3789	9.14	0.6939	.8398	.7831	.8271	**.8109**	.8271	**.8109**	.9441	.7543	.8396	.8084	.8231	.7895	.8209	.8076
glass0	2.06	0.6492	.9361	**.8558**	.9128	.8023	.8902	.7814	.9451	.7856	.9009	.7881	.8845	.8148	.8446	.7717
abalone918	16.68	0.6320	.8658	.7632	.8451	.7934	.8248	**.7944**	.9529	.6201	.8596	.7588	.8335	.7310	.7948	.6858
pima	1.90	0.5760	.8141	.7364	.8125	.7476	.7996	**.7486**	.8154	.7145	.8106	.7047	.8177	.7282	.8061	.7267
abalone19	128.87	0.5295	.8729	.7124	.8605	**.7129**	.8028	.6802	.8544	.5203	.8577	.6054	.8399	.6632	.7578	.5192
ecoli0147vs2356	10.59	0.5275	.9706	.8475	.9600	.8842	.9600	.8842	.9742	.8461	.9596	.8329	.9459	.8657	.9463	**.8889**

pageblocks0	8.77	0.5087	.9241	.9064	.9220	.9130	.9102	.8883	.9845	**.9484**	.9113	.8924	.9098	.8917	.9048	.8621
glass2	10.39	0.3952	.9505	**.7513**	.9276	.6403	.8785	.7174	.9636	.5424	.9670	.6345	.9064	.7131	.7779	.5672
vehicle2	2.52	0.3805	.9878	.9540	.9909	.9521	.9827	**.9581**	.9895	.9476	.9872	.9440	.9796	.9213	.9819	.9509
yeast1289vs7	30.56	0.3660	.8257	.6666	.8313	.7234	.7927	**.7312**	.9486	.7051	.8860	.7111	.8518	.7007	.7619	.7071
yeast1vs7	13.87	0.3534	.8913	.7137	.8843	.6494	.8406	.7021	.9315	.7064	.9008	.6986	.8791	**.7761**	.7267	.6378
glass0146vs2	11.06	0.3487	.9608	**.8166**	.8690	.7349	.8690	.7349	.9821	.7564	.9330	.7433	.9157	.6923	.7972	.5626
yeast0359vs78	9.12	0.3113	.9629	.6804	.9005	**.8435**	.9005	**.8435**	.9571	.7222	.8925	.7323	.8235	.7466	.7840	.7393
glass016vs2	10.29	0.2692	.9636	**.7250**	.8952	.6102	.8671	.6179	.9623	.6367	.9429	.6838	.8653	.5624	.8257	.5017
yeast1	2.46	0.2422	.7753	.7191	.7747	.7202	.7495	.7178	.8050	.7113	.7679	**.7253**	.7644	.7124	.7245	.7039
glass1	1.82	0.1897	.9184	**.7734**	.8858	.7538	.8505	.7416	.9031	.7577	.9343	.7620	.8575	.6620	.8243	.7223
vehicle3	2.52	0.1855	.8409	**.7521**	.8355	.7363	.7968	.7424	.9547	.7015	.8580	.7415	.8709	.7233	.8236	.7286
haberman	2.68	0.1850	.7595	**.6322**	.7386	.5924	.7102	.6150	.7122	.6309	.7953	.5070	.7349	.5942	.6978	.5936
yeast1458vs7	22.10	0.1757	.8417	.5944	.8123	.6325	.7519	.5789	.9120	.5230	.8995	.6573	.8407	**.6931**	.7072	.6255
vehicle1	2.52	0.1691	.8445	.7365	.8409	.7327	.8179	.7456	.9483	**.7468**	.8596	.7136	.8613	.7455	.7939	.7428
glass015vs2	9.12	0.1375	.8596	.7365	.7932	.7021	.7932	.7021	.9646	**.7444**	.9516	.6556	.8829	.6653	.8024	.6011

any learning methodology, as two of the properties which could hinder the performance conflux in the same problem.

Our objective is to analyze whether our feature weighting approach for FRBCS is also able to excel under these conditions. Table 7 shows the average performance for the FARC-HD-FW and the three algorithms for comparison in 26 selected problems. We can observe that the trend is similar to the previous case studies, in which the feature weighting methodology achieves the highest value for the AUC metric in the test partitions.

Fig. 15 depicts a graphical comparison of the ranks (computed with the Friedman aligned test). This allows us to better understand the superior quality of our approach, as the differences between the former and the remaining methods stand out. The scheme is identical to the one shown in the two previous case studies, thus indicating a very robust behavior for our proposed methodology.

The *p*-value associated with the Friedman aligned test is equal to $1.0482E^{-5}$, so that statistical differences are found among the results. Accordingly, Table 7 presents the Holm test in which the null hypothesis of equality is rejected for both the standard FARC-HD algorithm and C4.5.

Regarding the comparison with the feature selection mechanism, we observe that a low APV is also obtained, very close to 0.1. Since this value may be considered as a high threshold for determining that a given approach is statistically superior, we observe that the statistical differences are significant. Therefore, we may state that the higher performance of the feature weighting approach, supported by the average ranking result, make it a better suited methodology for this classification scenario in contrast with the simple feature selection.

In accordance with the above, we must also affirm in this case the good properties of the feature weighting approach, since it has been shown to be the most robust one even with the most difficult intrinsic data characteristics.

5.4. Analysis of the feature weighting approach versus EFS

In this last section of the experimental study, we analyze the performance of our feature weighting approach versus several selected EFSs. This is done to provide additional support to the good behavior shown for this proposal with respect to other related approaches that have been used in classification with imbalanced datasets. With this consideration in mind, we have selected the global 2-tuples tuning (GTS) [18], the GP-COACH-H algorithm [40], and IIVDFT [50], as a recent EFS decision tree.

As a summary, we show in Table 8 both the AUC test results and the APVs obtained by a Holm test in all three case studies, since the Friedman aligned *p*-values were lower than $1E^{-}7$. This table is divided into all datasets, datasets with high overlapping, and datasets with high overlapping and imbalance.

Observing these experimental results, we can draw similar conclusions to those in the previous part of our study. Specifically, our approach achieves the highest performance in all three scenarios, and is shown to be statistically superior for all datasets and for those problems with high overlapping. In the last analysis, i.e. high overlapping and imbalance, it outperforms IIVFDT but no differences are found with respect to GP-COACH-H and GTS, in spite of the fact the AUC value and the ranking are clearly higher for FARC-HD-FW.

5.5. On the time efficiency of FARC-HD-FW

For the sake of studying the efficiency of the algorithms, we show in Table 9 the elapsed time for all the approaches considered in the experimental study. These experiments were performed on a cluster-based computer with Intel(R) Core(TM) i7 CPU 930 microprocessors (4 cores/8 threads, 2.8 GHz, 8 MB Cache), 24 GB DDR2

Table 5
Average results and adjusted *p*-values (Holm test) for **all imbalanced datasets** (AUC metric). FARC-HD-FW is the proposed methodology (Feature Weighting), FARC-HD is the baseline FARC-HD algorithm, FARC-HD-FS includes Feature Selection, and C4.5 is simply the C4.5 decision tree. All datasets are preprocessed using SMOTE.

Algorithm	AUC test	Adjusted *p*-Value (Holm test)
FARC-HD-FW	**.8572 ± .0981**	********
FARC-HD	.8452 ± .1037	0.004510
FARC-HD-FS	.8463 ± .1007	0.007973
C4.5	.8288 ± .1192	0.000000

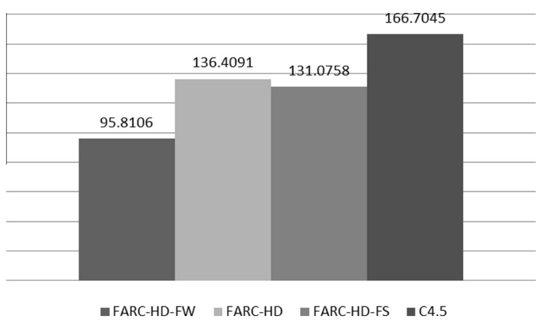


Fig. 13. Average ranking (computed by Friedman aligned test) for all datasets.

Table 6
Average results and adjusted *p*-values (Holm test) for **imbalanced datasets with high overlapping** ($F1 < 1.5$): *36 problems* are selected.

Algorithm	AUC test	Adjusted *p*-value (Holm test)
FARC-HD-FW	**.8010 ± .0929**	********
FARC-HD	.7832 ± .0932	0.00763
FARC-HD-FS	.7873 ± .0919	0.033857
C4.5	.7608 ± .1108	0.000009

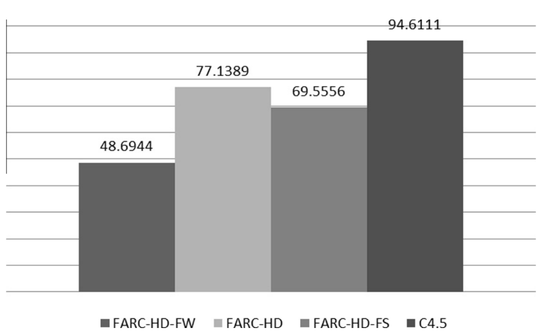


Fig. 14. Average ranking (computed by Friedman aligned test) for datasets with a high degree of overlap ($F1 < 1.5$): 36 problems.

memory and SATA 1TB 3Gb/s hard-drives, running on CentOS 6.4. We implemented all algorithms with Java(TM) programming language. From these results we can extract the following conclusions:

1. The baseline FARC-HD algorithm is the fastest method among the EFS (not comparable with C4.5). Therefore, this made it very suitable for its inclusion within the global feature weighting methodology.
2. Regarding the comparison between FARC-HD-FW and FARC-HD-FS, we observe that both of them are practically identical in efficiency when the number of variables is low, such as in ecoli, glass, and some yeast problems. However, when the number of attributes increases, the elapsed time of the feature weighting approach grows in a high rate. This is due to the way the membership function is computed, i.e. by means of the "power function", which implies a considerable delay in a part of the approach that is called often.
3. Finally, we must stress that with respect to the selected EFSs, FARC-HD-FW is the second fastest algorithm, only behind GTS. GP-COACH-H and IIVFDT have a lower efficiency and also worse results than our proposal.

Table 7
Average results and adjusted *p*-values (Holm test) **classification datasets with high imbalanced and high overlapping** (**$IR > 9$ and $F1 < 1.5$**): *26 problems* are selected.

Algorithm	AUC test	Adjusted *p*-value (Holm test)
FARC-HD-FW	**.8009 ± .0887**	********
FARC-HD	.7815 ± .08722	0.054687
FARC-HD-FS	.7867 ± .0865	0.107625
C4.5	.7512 ± .1110	0.000086

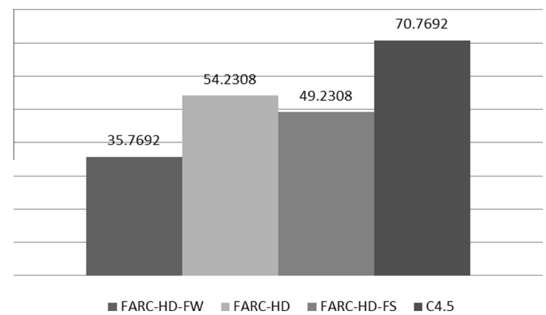


Fig. 15. Average ranking (computed by Friedman aligned test) for datasets with both high imbalance and high overlapping ($IR > 9$ and $F1 < 1.5$): 26 problems.

Table 8
Average results, ranks for the Friedman aligned test, and adjusted *p*-values (Holm test) for FARC-HD-FW and EFSs. All three case studies are shown.

Case study	Algorithm	AUC test	Rank	Adjusted *p*-value (Holm test)
All data	FARC-HD-FW	**.8556 ± .0998**	109.4545	********
	GP-COACH-H	.8456 ± .1089	133.9242	0.067966
	GTS	.8373 ± .1058	148.9848	0.008816
	IIVFDT	.8348 ± .1254	137.6364	0.067966
$F1 < 1.5$	FARC-HD-FW	**.7992 ± .0946**	55.7778	********
	GP-COACH-H	.7825 ± .1037	75.4722	0.06724
	GTS	.7758 ± .0985	76.6667	0.06724
	IIVFDT	.7641 ± .1254	82.0833	0.022384
$F1 < 1.5$ and $IR > 9$	FARC-HD-FW	**.8004 ± .0906**	41.6923	********
	GP-COACH-H	.7879 ± .0960	53.0000	0.353053
	GTS	.7793 ± .0974	52.8077	0.353053
	IIVFDT	.7599 ± .1317	62.5000	0.038648

In summary, we may highlight the good properties of our feature weighting methodology as it has achieved the best performance results while maintaining an acceptable execution time. The hitch here is only related to the "power function" which is known to be time restrictive. Nevertheless, we must stress that the number of required evaluations for convergence is only set to 1000, which is a low number for these types of approaches.

6. Concluding remarks

In the scenario of classification with imbalanced datasets, researchers must adapt the classification procedure in order to

Table 9
Average running time for all algorithms in the experimental study. FARC-HD-FW is the proposed methodology (Feature Weighting), FARC-HD is the baseline FARC-HD algorithm, FARC-HD-FS includes Feature Selection, C4.5 is the C4.5 decision tree. GP-COACH-H stands for the GP hierarchical approach, GTS is the Chi et al.'s algorithm with 2-tuples tuning, and IIVFDT is the interval-valued fuzzy decision tree. All datasets are preprocessed using SMOTE.

Dataset	IR	F1	FARC-HD-FW	FARC-HD	FARC-HD-FS	C4.5	GP-COACH-H	GTS	IIVFDT
iris0	2.00	16.8200	0:02:21.0	0:00:03.0	0:04:41.0	0:00:02.0	0:07:11.1	0:01:18.0	0:02:01.0
shuttle0vs4	13.87	12.9700	2:04:18.0	0:00:14.0	0:53:50.0	0:00:08.0	1:30:24.0	0:01:22.0	1:42:25.0
shuttle2vs4	20.50	12.1300	0:57:46.0	0:00:02.0	0:48:37.0	0:00:02.0	0:03:50.0	0:00:10.0	0:07:11.0
ecoli0vs1	1.86	9.7520	1:14:08.0	0:00:05.0	0:27:33.0	0:00:02.0	0:12:37.2	0:00:42.0	0:05:53.0
yeast5	32.78	4.1980	1:16:57.0	0:00:32.0	0:38:53.0	0:00:07.0	11:01:11.5	4:30:36.0	1:03:23.0
newthyroid2	4.92	3.5790	0:13:09.0	0:00:02.0	0:15:14.0	0:00:04.0	0:06:56.0	0:01:47.0	0:04:29.0
newthyroid1	5.14	3.5790	0:10:47.0	0:00:05.0	0:10:30.0	0:00:05.0	0:07:04.6	0:01:25.0	0:04:21.0
wisconsin	1.86	3.5680	0:21:53.0	0:00:10.0	0:19:12.0	0:00:06.0	5:04:43.8	0:08:20.0	0:54:22.0
glass0123vs456	3.19	3.3240	2:05:17.0	0:00:05.0	0:49:03.0	0:00:02.0	0:22:18.1	0:04:04.0	0:12:00.0
ecoli4	13.84	3.2470	0:29:20.0	0:00:08.0	0:25:37.0	0:00:02.0	0:34:33.8	0:17:47.0	0:11:20.0
yeast3	8.11	2.7510	3:17:53.0	0:00:34.0	2:01:17.0	0:00:10.0	9:52:16.5	2:47:02.0	0:59:04.0
ecoli1	3.36	2.6500	0:26:23.0	0:00:07.0	0:23:42.0	0:00:02.0	0:39:43.4	0:09:40.0	0:10:42.0
vowel0	10.10	2.4580	13:44:59.0	0:00:38.0	4:32:27.0	0:00:09.0	17:26:04.2	0:14:43.0	14:12:13.0
glass6	6.38	2.3910	0:44:57.0	0:00:05.0	0:31:25.0	0:00:04.0	0:23:09.6	0:02:33.0	0:14:52.0
ecoli0137vs26	39.15	2.3020	0:29:21.0	0:00:03.0	0:20:44.0	0:00:02.0	7:42:32.7	0:09:24.0	0:55:10.0
yeast6	39.15	1.9670	1:32:13.0	0:00:44.0	0:33:56.0	0:00:06.0	0:45:57.9	4:14:37.0	6:20:53.0
led7digit02456789vs1	10.97	1.9570	0:19:02.0	0:00:05.0	0:20:10.0	0:00:12.0	1:16:31.8	0:06:49.0	0:13:59.0
glass016vs5	19.44	1.8510	1:04:21.0	0:00:06.0	0:42:25.0	0:00:02.0	0:19:32.4	0:02:43.0	0:39:35.0
ecoli2	5.46	1.8260	0:29:19.0	0:00:15.0	0:26:36.0	0:00:02.0	0:43:09.8	0:09:23.0	0:10:23.0
segment0	6.01	1.7980	51:11:49.0	0:01:14.0	8:22:20.0	0:00:07.0	76:20:01.4	0:05:27.0	83:10:53.0
ecoli067vs5	10.00	1.6920	0:33:18.0	0:00:02.0	0:20:23.0	0:00:08.0	0:29:31.0	0:02:43.0	7:29:29.0
yeast02579vs368	9.14	1.6350	1:32:49.0	0:00:08.0	0:55:10.0	0:00:17.0	3:57:29.2	1:44:33.0	2:18:04.0
ecoli034vs5	9.00	1.6320	0:22:58.0	0:00:04.0	0:21:07.0	0:00:10.0	0:23:25.6	0:01:59.0	0:07:58.0
ecoli0234vs5	9.10	1.6180	0:18:18.0	0:00:02.0	0:28:51.0	0:00:09.0	0:19:17.1	0:01:33.0	0:08:01.0
ecoli046vs5	9.15	1.6030	0:20:34.0	0:00:02.0	0:19:17.0	0:00:08.0	0:27:11.1	0:01:43.0	1:54:22.0
ecoli0346vs5	9.25	1.5950	0:50:50.0	0:00:04.0	0:37:31.0	0:00:14.0	0:20:48.2	0:04:00.0	0:08:10.0
ecoli3	8.19	1.5790	0:28:59.0	0:00:12.0	0:27:09.0	0:00:02.0	0:42:48.8	0:08:27.0	0:13:08.0
yeast2vs4	9.08	1.5790	1:23:42.0	0:00:08.0	1:09:31.0	0:00:08.0	0:58:15.2	0:04:46.0	2:08:20.0
pageblocks13vs4	15.85	1.5470	16:36:05.0	0:02:17.0	2:52:53.0	0:00:10.0	1:08:58.9	0:11:26.0	0:49:15.0
glass04vs5	9.22	1.5420	1:11:28.0	0:00:04.0	1:09:31.0	0:00:08.0	0:05:04.0	0:00:55.0	3:13:43.0
glass4	15.47	1.4690	1:09:22.0	0:00:05.0	0:42:02.0	0:00:04.0	0:23:29.0	0:01:48.0	0:16:46.0
ecoli01vs5	11.00	1.3900	0:55:11.0	0:00:04.0	0:51:57.0	0:00:06.0	0:07:31.2	0:03:39.0	2:27:15.0
cleveland0vs4	12.62	1.3500	2:52:16.0	0:00:04.0	1:22:33.0	0:00:04.0	1:13:08.8	0:01:26.0	1:02:06.0
ecoli0146vs5	13.00	1.3400	0:43:36.0	0:00:04.0	0:29:15.0	0:00:08.0	0:34:52.9	0:03:58.0	0:07:51.0
yeast2vs8	23.10	1.1420	3:44:28.0	0:00:56.0	1:32:22.0	0:00:15.0	0:35:07.1	2:02:54.0	0:19:49.0
ecoli0347vs56	9.28	1.1300	0:25:57.0	0:00:03.0	0:28:14.0	0:00:10.0	0:35:43.7	0:04:40.0	2:48:15.0
vehicle0	3.23	1.1240	34:19:42.0	0:01:04.0	7:08:39.0	0:00:07.0	25:38:00.2	0:58:11.0	35:32:21.0
ecoli01vs235	9.17	1.1030	0:24:39.0	0:00:02.0	0:28:18.0	0:00:07.0	0:45:49.9	0:01:56.0	0:10:03.0
yeast05679vs4	9.35	1.0510	1:22:09.0	0:00:18.0	0:47:05.0	0:00:08.0	1:20:05.1	0:00:37.0	1:57:01.0
glass06vs5	11.00	1.0490	0:58:14.0	0:00:02.0	0:40:19.0	0:00:07.0	0:24:59.0	0:00:35.0	0:07:40.0
glass5	22.81	1.0190	0:56:45.0	0:00:05.0	0:31:41.0	0:00:04.0	0:18:31.0	0:03:04.0	0:11:14.0
ecoli067vs35	9.09	0.9205	0:22:14.0	0:00:02.0	0:23:20.0	0:00:08.0	0:26:02.8	0:01:33.0	0:07:47.0
ecoli0267vs35	9.18	0.9129	0:37:36.0	0:00:02.0	0:40:41.0	0:00:07.0	0:27:34.8	0:02:13.0	5:36:08.0
ecoli0147vs56	12.28	0.9124	0:23:08.0	0:00:03.0	0:13:02.0	0:00:05.0	0:48:27.6	0:03:40.0	2:27:47.0
yeast4	28.41	0.7412	3:19:31.0	0:01:06.0	1:00:06.0	0:00:07.0	11:02:13.5	2:45:25.0	1:07:19.0
yeast0256vs3789	9.14	0.6939	1:52:53.0	0:00:11.0	0:54:20.0	0:00:14.0	5:00:13.0	1:39:48.0	2:16:48.0
glass0	2.06	0.6492	1:14:58.0	0:00:05.0	0:29:46.0	0:00:02.0	0:13:06.6	0:06:17.0	0:14:17.0
abalone918	16.68	0.6320	2:45:38.0	0:00:34.0	1:07:38.0	0:00:02.0	1:23:10.6	1:07:26.0	0:27:16.0
pima	1.90	0.5760	3:01:08.0	0:00:44.0	1:26:17.0	0:00:05.0	6:34:51.3	0:48:15.0	7:03:43.0
abalone19	128.87	0.5295	8:31:17.0	0:03:03.0	2:32:13.0	0:00:03.0	24:09:49.6	10:28:12.0	2:25:16.0
ecoli0147vs2356	10.59	0.5275	0:58:06.0	0:00:07.0	0:50:19.0	0:00:10.0	1:11:20.5	0:04:08.0	0:13:35.0
pageblocks0	8.77	0.5087	1:32:53.0	0:00:10.0	0:42:35.0	0:00:07.0	49:05:29.4	0:52:02.0	28:01:21.0
glass2	10.39	0.3952	1:40:41.0	0:00:07.0	1:06:03.0	0:00:02.0	0:08:33.9	0:10:03.0	0:11:31.0
vehicle2	2.52	0.3805	39:10:50.0	0:00:48.0	8:16:47.0	0:00:08.0	38:02:17.9	4:08:40.0	41:00:32.0
yeast1289vs7	30.56	0.3660	1:24:01.0	0:00:14.0	0:50:03.0	0:00:10.0	4:00:03.8	0:31:32.0	11:18:20.0
yeast1vs7	13.87	0.3534	1:58:16.0	0:00:26.0	1:09:09.0	0:00:08.0	2:01:31.8	0:47:24.0	0:18:29.0
glass0146vs2	11.06	0.3487	1:39:43.0	0:00:04.0	0:47:18.0	0:00:04.0	0:06:02.5	0:05:23.0	2:08:00.0
yeast0359vs78	9.12	0.3113	2:11:38.0	0:00:10.0	1:10:16.0	0:00:15.0	0:10:17.9	0:19:49.0	5:51:37.0
glass016vs2	10.29	0.2692	0:58:13.0	0:00:05.0	0:39:10.0	0:00:02.0	0:08:48.3	0:05:05.0	1:11:36.0
yeast1	2.46	0.2422	1:58:43.0	0:00:42.0	1:08:27.0	0:00:07.0	9:39:02.6	1:22:15.0	0:51:27.0
glass1	1.82	0.1897	1:32:14.0	0:00:06.0	1:03:30.0	0:00:02.0	0:20:53.0	0:03:34.0	0:12:09.0
vehicle3	2.52	0.1855	75:30:32.0	0:02:08.0	11:17:21.0	0:00:09.0	30:26:55.2	2:38:54.0	108:33:57.0
haberman	2.68	0.1850	0:11:38.0	0:00:07.0	0:21:12.0	0:00:07.0	0:24:58.1	0:02:32.0	0:03:05.0
yeast1458vs7	22.10	0.1757	0:45:56.0	0:00:09.0	0:47:03.0	0:00:07.0	2:05:36.6	0:34:01.0	4:26:20.0
vehicle1	2.52	0.1691	66:57:29.0	0:01:35.0	11:35:40.0	0:00:09.0	31:24:58.1	2:18:21.0	121:38:32.0
glass015vs2	9.12	0.1375	1:37:55.0	0:00:04.0	1:12:06.0	0:00:05.0	1:17:35.8	0:04:07.0	0:11:46.0
Average	-	-	5:41:53.4	0:00:21.8	1:29:38.5	0:00:06.6	6:00:09.0	0:45:32.2	7:55:31.5

avoid the bias towards the majority class examples, and therefore to achieve a similar accuracy for both concepts of the problem. However, for those problems where the skewed class distribution occurs in combination with overlapping in the borderline areas of the dataset, standard solutions for imbalanced classification no longer obtain the expected performance.

In this work we have focused on these data intrinsic characteristics, and we have proposed a simple but effective feature weighting approach in synergy with FRBCS. Our intention was to minimize the contribution of those variables of the problem for which the examples of both classes are the most "entwined". We rejected the simple removal of these variables, as in the feature selection scheme, as we are aware that these variables might have a small but positive support in the final decision of the fuzzy models for classification purposes. In order to show the behavior of our proposal, we have made use of the FARC-HD algorithm to develop a complete EFS, defining the whole approach as FARC-HD-FW.

From our analysis, we must emphasize that feature weighting in FRBCSs has been shown to be robust in those problems which share a high degree of imbalance and overlapping, achieving a higher performance and average ranking than all algorithms used for comparison, and showing statistical differences with respect to the baseline FARC-HD approach and the C4.5 decision tree. We must stress the significance of these results as this type of problems might be identified as the most difficult from the point of view of the optimal identification of the two classes, and therefore we have proposed an appropriate solution to overcome it.

Acknowledgements

This paper was funded by King Abdulaziz University, under Grant No. (3-611-1434-HiCi). The authors, therefore, acknowledge technical and financial support of KAU. This work was also supported by the Spanish Ministry of Science and Technology under Projects TIN-2011-28488 and TIN-2012-33856, and the Andalusian Research Plans P11-TIC-7765 and P10-TIC-6858.

References

[1] J. Alcalá-Fdez, R. Alcalá, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[2] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Multi-Valued Logic Soft Comput. 17 (2–3) (2011) 255–287.
[3] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera, KEEL: a software tool to assess evolutionary algorithms for data mining problems, Soft. Comput. 13 (2009) 307–318.
[4] R. Barandela, J.S. Sánchez, V. García, E. Rangel, Strategies for learning in class imbalance problems, Pattern Recogn. 36 (3) (2003) 849–851.
[5] G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, A study of the behaviour of several methods for balancing machine learning training data, SIGKDD Explor. 6 (1) (2004) 20–29.
[6] F.J. Berlanga, A.J.R. Rivas, M.J. del Jesús, F. Herrera, GP-COACH: genetic programming-based learning of compact and accurate fuzzy rule-based classification systems for high-dimensional problems, Inform. Sci. 180 (8) (2010) 1183–1200.
[7] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (7) (1997) 1145–1159.
[8] H. Bustince, M. Pagola, E. Barrenechea, J. Fernández, P. Melo-Pinto, P. Couto, H.R. Tizhoosh, J. Montero, Ignorance functions. an application to the calculation of the threshold in prostate ultrasound images, Fuzzy Sets Syst. 161 (1) (2010) 20–36.
[9] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[10] Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific, 1996.
[11] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends, Fuzzy Sets Syst. 141 (1) (2004) 5–31.
[12] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[13] M. Denil, T. Trappenberg, Overlap versus imbalance, in: Proceedings of the 23rd Canadian conference on Advances in artificial intelligence (CCAI'10), Lecture Notes on Artificial Intelligence, vol. 6085, 2010, pp. 220–231.
[14] P. Domingos, Metacost: a general method for making classifiers cost-sensitive, in: Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD'99), 1999, pp. 155–164.
[15] A. Eiben, J. Smith, Introduction to evolutionary computation, in: Natural Computing Series, Springer, 2003.
[16] L.J. Eshelman, Foundations of genetic algorithms, in: The CHC Adaptive Search Algorithm: How to have Safe Search When Engaging in Nontraditional Genetic Recombination, Morgan Kaufman, 1991.
[17] M. Fazzolari, R. Alcalà, Y. Nojima, H. Ishibuchi, F. Herrera, A review of the application of multi-objective evolutionary systems: current status and further directions, IEEE Trans. Fuzzy Syst. 21 (1) (2013) 45–65.
[18] A. Fernández, M.J. del Jesus, F. Herrera, On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets, Inform. Sci. 180 (8) (2010) 1268–1291.
[19] A. Fernández, S. García, M.J. del Jesus, F. Herrera, A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets, Fuzzy Sets Syst. 159 (18) (2008) 2378–2398.
[20] A. Fernandez, S. García, J. Luengo, E. Bernadó-Mansilla, F. Herrera, Genetics-based machine learning for rule induction: state of the art, taxonomy and comparative study, IEEE Trans. Evol. Comput. 14 (6) (2010) 913–941.
[21] S. García, J. Derrac, I. Triguero, C.J. Carmona, F. Herrera, Evolutionary-based selection of generalized instances for imbalanced classification, Knowl. Based Syst. 25 (1) (2012) 3–12.
[22] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power, Inform. Sci. 180 (10) (2010) 2044–2064.
[23] S. García, F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, J. Mach. Learn. Res. 9 (2008) 2607–2624.
[24] V. García, R. Mollineda, J.S. Sánchez, On the k-NN performance in a challenging scenario of imbalance and overlapping, Pattern Anal. Appl. 11 (3–4) (2008) 269–280.
[25] V. García, J. Sánchez, R. Mollineda, On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, Knowl. Based Syst. 25 (1) (2012) 13–21.
[26] J. Han, M. Kamber, Data Mining. Concepts and Techniques, second ed., Morgan Kaufman, 2006.
[27] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[28] H. He, Y. Ma (Eds.), Imbalanced Learning: Foundations, Algorithms, and Applications, first ed., Wiley, 2013.
[29] F. Herrera, M. Lozano, A.M. Sánchez, A taxonomy for the crossover operator for real-coded genetic algorithms: an experimental study, Int. J. Intell. Syst. 18 (2003) 309–338.
[30] F. Herrera, L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, IEEE Trans. Fuzzy Syst. 8 (6) (2000) 746–752.
[31] T. Ho, M. Basu, Complexity measures of supervised classification problems, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 289–300.
[32] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[33] J. Huang, C.X. Ling, Using AUC and accuracy in evaluating learning algorithms, IEEE Trans. Knowl. Data Eng. 17 (3) (2005) 299–310.
[34] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer-Verlag, 2004.
[35] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 13 (2005) 428–435.
[36] N. Japkowicz, S. Stephen, The class imbalance problem: a systematic study, Intell. Data Anal. J. 6 (5) (2002) 429–450.
[37] B. Kavsek, N. Lavrac, Apriori-sd: adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (7) (2006) 543–583.
[38] M. Lin, K. Tang, X. Yao, Dynamic sampling approach to training neural networks for multiclass imbalance classification, IEEE Trans. Neural Netw. Learn. Syst. 24 (4) (2013) 647–660.
[39] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 491–502.
[40] V. López, A. Fernández, M.J. del Jesus, F. Herrera, A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline data-sets, Knowl. Based Syst. 38 (2013) 85–104.
[41] V. López, A. Fernández, S. García, V. Palade, F. Herrera, An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics, Inform. Sci. 250 (20) (2013) 113–141.
[42] V. López, A. Fernández, J.G. Moreno-Torres, F. Herrera, Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics, Expert Syst. Appl. 39 (7) (2012) 6585–6608.
[43] J. Luengo, A. Fernández, S. García, F. Herrera, Addressing data complexity for imbalanced data sets: analysis of SMOTE-based oversampling and evolutionary undersampling, Soft. Comput. 15 (10) (2011) 1909–1936.
[44] J.G. Moreno-Torres, F. Herrera, A preliminary study on overlapping and data fracture in imbalanced domains by means of genetic programming-based feature extraction, in: Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA'10), 2010, pp. 501–506.
[45] A. Orriols-Puig, E. Bernadó-Mansilla, Evolutionary rule-based systems for imbalanced datasets, Soft. Comput. 13 (3) (2009) 213–225.
[46] A. Orriols-Puig, E. Bernadó-Mansilla, D.E. Goldberg, K. Sastry, P.L. Lanzi, Facetwise analysis of XCS for problems with class imbalances, IEEE Trans. Evol. Comput. 13 (2009) 260–283.

[47] A. Palacios, L. Sanchez, I. Couso, Combining adaboost with preprocessing algorithms for extracting fuzzy rules from low quality data in possibly imbalanced problems, Int. J. Uncertain. Fuzz. Knowl. Based Syst. 20 (SUPPL. 2) (2012) 51–71.
[48] R.C. Prati, G.E.A.P.A. Batista, Class imbalances versus class overlapping: an analysis of a learning system behavior, in: Proceedings of the 2004 Mexican International Conference on Artificial Intelligence (MICAI'04), 2004, pp. 312–321.
[49] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kauffman, 1993.
[50] J.A. Sanz, A. Fernández, H. Bustince, F. Herrera, IIVFDT: ignorance functions based interval-valued fuzzy decision tree with genetic tuning, Int. J. Uncertain., Fuzz. Knowl.-Based Syst. 20 (SUPPL. 2) (2012) 1–30.
[51] C. Seiffert, T.M. Khoshgoftaar, J.V. Hulse, A. Folleco, An empirical study of the classification performance of learners on imbalanced and noisy software quality data, Inform. Sci. 259 (2014) 571–595.
[52] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review, Int. J. Pattern Recogn. Artif. Intell. 23 (4) (2009) 687–719.
[53] K.M. Ting, An instance-weighting method to induce cost-sensitive trees, IEEE Trans. Knowl. Data Eng. 14 (3) (2002) 659–665.
[54] P. Villar, A. Fernández, R.A. Carrasco, F. Herrera, Feature selection and granularity learning in genetic fuzzy rule-based classification systems for highly imbalanced data-sets, Int. J. Uncertain., Fuzz. Knowl.-Based Syst. 20 (3) (2012) 369–397.
[55] G.M. Weiss, Mining with rarity: a unifying framework, SIGKDD Explor. 6 (1) (2004) 7–19.
[56] G.M. Weiss, The impact of small disjuncts on classifier learning, in: R. Stahlbock, S.F. Crone, S. Lessmann (Eds.), Data Mining, Annals of Information Systems, vol. 8, Springer, 2010, pp. 193–226.
[57] G.M. Weiss, Y. Tian, Maximizing classifier utility when there are data acquisition and modeling costs, Data Min. Knowl. Discov. 17 (2) (2008) 253–282.
[58] B. Zadrozny, C. Elkan, Learning and making decisions when costs and probabilities are both unknown, in: Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01), 2001, pp. 204–213.
[59] B. Zadrozny, J. Langford, N. Abe, Cost-sensitive learning by cost-proportionate example weighting, in: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03), 2003, pp. 435–442.
[60] C. Zhang, S. Zhang, Association rule mining, models and algorithms, Lecture Notes in Computer Science, vol. 2307, Springer, 2002.
[61] W. Zong, G.-B. Huang, Y. Chen, Weighted extreme learning machine for imbalance learning, Neurocomputing 101 (2013) 229–242.