IFROWANN: Imbalanced Fuzzy-Rough Ordered Weighted Average Nearest Neighbor Classification

Enislay Ramentol, Sarah Vluymans, Nele Verbiest, Yailé Caballero, Rafael Bello, Chris Cornelis, and Francisco Herrera, *Member, IEEE*

***Abstract*—Imbalanced classification deals with learning from data with a disproportional number of samples in its classes. Traditional classifiers exhibit poor behavior when facing this kind of data because they do not take into account the imbalanced class distribution. Four main kinds of solutions exist to solve this problem: modifying the data distribution, modifying the learning algorithm for considering the imbalance representation, including the use of costs for data samples, and ensemble methods. In this paper, we adopt the second type of solution, and introduce a classification algorithm for imbalanced data that uses fuzzy rough set theory and ordered weighted average aggregation. The proposal considers different strategies to build a weight vector to take into account data imbalance. Our methods are validated by an extensive experimental study, showing statistically better results than thirteen other state-of-the-art methods.**

***Index Terms*—machine learning, imbalanced classification, fuzzy rough sets, ordered weighted average.**

I. Introduction

LEARNING from imbalanced data is a challenging task that has gained attention over the last few years [28], [35], [41]. In contrast to traditional classification, it deals with data sets where one or more classes are under-represented. In this paper, we consider the two-class case where one class (the majority or negative class) is over-represented and the other class (the minority or positive class) is under-represented. This characteristic is very common in real-world applications, such as anomaly detection [33], medical applications [34], microarray data [49], database marketing [15], etc., and has opened up a whole new field of research to develop new techniques to overcome the imbalance problem.

Classical machine learning algorithms often obtain high accuracy over the majority class, while for the minority class the opposite occurs. This happens because the classifier focuses on global measures that do not take into account the class data distribution [28], [35], [41]. Nevertheless the most interesting information is often found within the minority class.

Many techniques for dealing with class imbalance have emerged. These techniques can be grouped into four main categories: those that modify the data distribution by preprocessing techniques (data level solutions), those at the level of the learning algorithm which adapt a base classifier to deal with class imbalance (algorithm level solutions), those that apply different costs to misclassification of positive and negative samples (cost-sensitive solutions) and ensemble based solutions that combine the previous solutions by means of an ensemble.

In this paper, we present a new algorithm level solution to classify imbalanced data that is based on the Fuzzy Rough Nearest Neighbor (FRNN) classifier introduced in [32]. In order to predict the class of a new test instance, the FRNN algorithm computes the sum of the memberships of the instance to the fuzzy-rough lower and upper approximation of each class. The lower approximation membership expresses the degree to which similar elements of the opposite class do not exist, while the upper approximation membership tells us to which extent similar elements of the same class exist. Finally, FRNN assigns the instance to the class with the higher sum.

However, this algorithm has some important weaknesses. On one hand, its classifications are completely determined by the closest samples in either class, thus making it very sensitive to noise [44]. On the other hand, FRNN treats the positive and negative class in a symmetric way and hence makes no provisions for the class imbalance. Therefore, in this paper, we have designed a new classifier called the Imbalanced Fuzzy Rough Ordered Weighted Average Nearest Neighbor (IFROWANN) algorithm; it computes the approximations taking into account not only the closest samples of the opposite class, but all of them, assigning them decreasing weights proportionate to their similarity with the test sample x, following two steps:

1) we consider different weight vectors for the majority and the minority class, taking into account the fact that the former contains much fewer elements than the latter.
2) we aggregate training samples' contributions by means of the ordered weighted average (OWA) fuzzy rough set model from [11].

Using this approach, our proposed algorithm can better address

This work was partially supported by the Spanish Ministry of Science and Technology under the project TIN2011-28488 and the Andalusian Research Plans P11-TIC-7765 and P10-TIC-6858, and by project PYR-2014-8 of the Genil Program of CEI BioTic GRANADA. The authors would like to thank Nitesh Chawla for making the code of the Hellinger Distance Decision Tree method described in [10] and referred to in this paper as HDDT+bagging available.

E. Ramentol and Y. Caballero are with the Department of Computer Science, University of Camagüey, Cuba, e-mail: enislay@gmail.com, yailec@yahoo.com.

S. Vluymans, N. Verbiest and C. Cornelis are with the Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Gent, Belgium, email: {Sarah.Vluymans,Nele.Verbiest,Chris.Cornelis}@UGent.be

R. Bello is with the Department of Computer Science, Central University of Las Villas, Cuba, email: rbellop@uclv.edu.cu

C. Cornelis and F. Herrera are with the Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain, e-mail: {chris.cornelis,herrera}@decsai.ugr.es

F. Herrera is with the Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, Saudi Arabia

the imbalanced data distributions.

To evaluate the quality of our model, we have carried out an extensive experimental analysis on a collection of 102 imbalanced data sets with different imbalance ratios (IR), originating from the UCI repository. In the experiments, we have compared our algorithm with the original FRNN proposal to show that it is better positioned to deal with the class imbalance. In order to demonstrate the importance of differentiating the weight vectors for the positive and negative class, we have considered a version of IFROWANN in which equal weight vectors are assigned to each class. This has shown to seriously weaken the performance of the algorithm. Finally, we have compared IFROWANN with a set of thirteen state-of-the-art methods specifically designed for imbalanced classification. To assess the classification performance, we have used the well-known Area Under the Curve (AUC) metric, and the significance of the results has been supported by the proper statistical analysis.

The remainder of this paper is organized as follows. In Section II, we provide an introduction to the imbalanced classification problem, including an overview of the state-of-the-art methods for solving it, and a discussion of its evaluation. In Section III, we recall the standard FRNN algorithm. In Section IV, we introduce the IFROWANN algorithm, and outline the proposed weighting strategies to deal with imbalanced data. In Section V, we discuss the setup of the experimental study, including a description of the benchmark data sets, the algorithms used for comparison along with their parameters, and the statistical tests used for performance comparison. In Section VI, we present and discuss the results. In Section VII, we draw some conclusions about the study and outline future work.

II. Imbalanced Classification Problems

A. Two-Class Imbalanced Classification: Models and Evaluation

The class imbalance problem is growing in importance and has been identified as one of the 10 main challenges of Data Mining [48]. The two-class version of this problem is formally described below.

We consider a set of data samples U, characterized by their values for the set $\mathscr{A} = \{a_1, \ldots, a_m\}$ of attributes. Moreover, $U = P \cup N$, where P represents the positive class and N the negative class. We denote $p = |P|$, $n = |N|$ and $t = |U| = p + n$. The imbalance ratio is then defined as $IR = \frac{n}{p}$.

The imbalanced classification problem can be tackled using four main types of solutions:

1) **Sampling (solutions at the data level)** [4], [7], [8], [22]: this kind of solution consists of balancing the class distribution by means of a preprocessing strategy. Techniques at data level are divided in 3 groups:
   - **Undersampling methods**: create a subset of the original data set by eliminating some of the examples of the majority class.
   - **Oversampling methods:** create a superset of the original data set by replicating some of the examples of the minority class or creating new minority instances, for example by interpolation of original instances.
   - **Hybrid methods**: combine the two previous methods by reducing the size of the majority class and increasing the number of minority elements.

   An important advantage of the data level approaches is that their use is independent from the classifier selected [38].
2) **Design of specific algorithms (solutions at the algorithmic level)** [5], [31], [10] : in this case, a traditional classifier is adapted to deal directly with the imbalance between the classes, for example, modifying the cost per class [26] or adjusting the probability estimation in the leaves of a decision tree to favor the positive class [46].
3) **Cost-sensitive solutions** [14], [42], [50], [51]: these kind of methods incorporate solutions at data level, at algorithmic level, or at both levels together, that try to minimize higher cost errors. Let $C(+,-)$ denote the cost of misclassifying a positive (minority class) instance as a negative (majority class) instance and $C(-,+)$ the cost of the inverse case. We impose $C(+,-) > C(-,+)$, i.e., the cost of misclassifying a positive instance should be higher than the cost of misclassifying a negative one.
4) **Ensemble solutions** [20]: Ensemble techniques for imbalanced classification usually consist of a combination of an ensemble learning algorithm and one of the techniques above, specifically, data level and cost-sensitive. Through the addition of a data level approach to the ensemble learning algorithm, the new hybrid method usually preprocesses the data before training each classifier. On the other hand, instead of modifying the base classifier in order to accept costs in the learning process, cost-sensitive ensembles guide the cost minimization via the ensemble learning algorithm.

Below, we review some high-quality proposals that will be used in our experimental study.

- **Synthetic Minority Oversampling Technique (SMOTE)** [7]. An oversampling method that creates new minority class examples by interpolating between minority class examples and their nearest neighbors.
- **SMOTE-ENN** [4]. This hybrid method applies the Edited Nearest Neighbor (ENN) technique to remove examples from both classes after SMOTE has been applied. In particular, any example that is misclassified by its three nearest neighbors is removed from the training set.
- **SMOTE-RSB$_*$** [38]. This is another hybrid data level method. It first applies SMOTE to introduce new synthetic minority class instances to the training set, and then removes synthetic instances that do not belong to the lower approximation of its class, computed using rough set theory [36]. This process is repeated until the training set is balanced.
- **Hellinger Distance Decision Trees (HDDT)** [10]. This algorithm level method is a decision tree technique that uses the Hellinger distance as the splitting criterion. It yields very good results for imbalanced data when used in a bagging (ensemble) configuration, which is the setup

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2014.2371472, IEEE Transactions on Fuzzy Systems

considered in this paper.

- **Cost-sensitive C4.5 decision tree (CS-C4.5)** [42]. This method builds decision trees that try to minimize the number of high cost errors and, as a consequence, leads to the minimization of the total misclassification costs in most cases. The method changes the class distribution such that the induced tree is in favor of the class with high weight/cost and is less likely to commit errors with high cost.
- **Cost-sensitive Support Vector Machine (CS-SVM)** [45]. This method is a modification of the soft-margin support vector machine [43]. It biases SVM in a way that will push the boundary away from the positive instances using different error costs for the positive and negative classes.
- **EUSBOOST** [21]. An ensemble method that uses Evolutionary UnderSampling (EUS, [25]) guided boosting. EUS arises from the application of evolutionary prototype selection algorithms to imbalanced domains. In EUS, each chromosome is a binary vector representing the presence or absence of instances in the data set. This method reduces the search space by considering only the majority class instances; hence, all the minority class instances are always introduced in the new data set. The fitness function tries to balance between the minority class and majority class instances, and includes a diversity mechanism among classifiers.

Next, we will discuss the evaluation of machine learning algorithms in imbalanced domains. Consider a two-class problem. For any given classifier, a correctly classified positive instance is called a *true positive (TP)*. Similarly, a *true negative (TN)* is a negative instance that was correctly classified as negative. In the remaining cases, a positive instance was either misclassified as negative, a *false negative (FN)*, or a negative instance was wrongly predicted as positive, a *false positive (FP)*. The *confusion matrix*, shown in Table I, presents a numerical summary of this information, showing the number of instances in each case.

Actual/Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table I: Confusion matrix obtained after classification of a two-class dataset.

For classical domains, the performance is typically evaluated using predictive accuracy (*acc*), defined by

$$acc = \frac{TP+TN}{TN+TN+FP+FN}.$$

However, this is not appropriate when the data are imbalanced or when the costs of different errors vary markedly [9]. Indeed, accuracy can take on misleadingly high values. As an example, assume that the IR of the data set is 9, meaning that 90% of the elements belong to the negative class. When we classify all instances as negative, we obtain a predictive accuracy of 90%. Even though this is a high value, the classifier has still misclassified the entire positive class, which renders it quite useless.

A more appropriate way to measure the performance of classification over imbalanced data sets are the Receiver Operating Characteristic (ROC) graphs [6]. These graphs visualize the tradeoff between the True Positive Rate (TPR) and False Positive Rate (FPR), defined as

$$TPR = \frac{TP}{TP+FN} \quad \text{and} \quad FPR = \frac{FP}{FP+TN},$$

when the classifier is treated as a probabilistic classifier, that is, one which calculates the probability that the element under consideration belongs to the given class. By varying the threshold for belonging to the positive class, different points of the ROC curve are generated.

The Area Under the ROC Curve (AUC) [30] then provides a single-number summary for the performance of learning algorithms. The AUC can be interpreted as the probability that the classifier assigns a lower probability of belonging to the positive class to a randomly chosen negative instance than to a randomly chosen positive instance [16].There are many ways to compute the AUC. In this paper, we use the definition given by Fawcett [16], who proposed an algorithm that, instead of collecting ROC points, adds successive areas of trapezoids to the computed AUC value.

III. Fuzzy-Rough Nearest Neighbor Algorithm (FRNN)

In this section, we recall the FRNN classification algorithm proposed in [32]. We apply it directly to the specific case of two-class imbalanced data. In order to predict the class of a new test instance x, the FRNN algorithm computes the sum of the memberships of x to the fuzzy-rough lower and upper approximation of each class, and assigns the instance to the class for which this sum is higher. More precisely, let $\mathscr{I}$ be an implicator[1], $\mathscr{T}$ a t-norm and R a fuzzy relation that represents approximate indiscernibility between instances. The membership degrees $\underline{P}(x)$ and $\underline{N}(x)$ of x to the lower approximation of P and N are defined by, respectively,

$$\underline{P}(x) = \min_{y \in U} \mathscr{I}(R(x,y),P(y)) \quad (1)$$

$$\underline{N}(x) = \min_{y \in U} \mathscr{I}(R(x,y),N(y)) \quad (2)$$

The value $\underline{P}(x)$ can be interpreted as the degree to which objects outside P (thus, in N) which are approximately indiscernible from x do not exist. A similar interpretation can be given to the value $\underline{N}(x)$.

On the other hand, the membership degrees $\overline{P}(x)$ and $\overline{N}(x)$ of x to the upper approximation of P and N under R are defined by, respectively,

$$\overline{P}(x) = \max_{y \in U} \mathscr{T}(R(x,y),P(y)) \quad (3)$$

$$\overline{N}(x) = \max_{y \in U} \mathscr{T}(R(x,y),N(y)) \quad (4)$$

[1] An implicator $\mathscr{I}$ is a $[0,1]^2 \rightarrow [0,1]$ mapping that is decreasing in its first argument and increasing in its second argument, and that satisfies $\mathscr{I}(0,0) = \mathscr{I}(0,1) = \mathscr{I}(1,1) = 1$ and $\mathscr{I}(1,0) = 0$.

$\overline{P}(x)$ can be interpreted as the degree to which another element in P close to x exists, and similarly for $\overline{N}(x)$.

In this paper, we consider $\mathscr{I}$ and $\mathscr{T}$ defined by $\mathscr{I}(a,b) = \max(1-a,b)$ and $\mathscr{T}(a,b) = \min(a,b)$, for a,b in $[0,1]$. It can be verified that in this case, Eqs. (1)–(4) can be simplified to

$$\begin{aligned} \underline{P}(x) &= \min_{y\in N} 1 - R(x,y) && (5)\\ \underline{N}(x) &= \min_{y\in P} 1 - R(x,y) && (6)\\ \overline{P}(x) &= \max_{y\in P} R(x,y) && (7)\\ \overline{N}(x) &= \max_{y\in N} R(x,y) && (8) \end{aligned}$$

In other words, $\underline{P}(x)$ is determined by the similarity to the closest negative (majority) sample, and $\underline{N}(x)$ is determined by the similarity to the closest positive (minority) sample. On the other hand, to obtain $\overline{P}(x)$ and $\overline{N}(x)$, we look for the most similar element to x belonging to the positive, resp., negative class. Also, the lower and upper approximations are clearly related: $\overline{P}(x) = 1 - \underline{N}(x)$ and $\overline{N}(x) = 1 - \underline{P}(x)$. The FRNN algorithm then determines the classification of the test instance x as follows. We compute

$$\begin{aligned} \mu_P(x) &= \frac{\underline{P}(x) + \overline{P}(x)}{2} = \frac{\underline{P}(x) + 1 - \underline{N}(x)}{2} && (9)\\ \mu_N(x) &= \frac{\underline{N}(x) + \overline{N}(x)}{2} = \frac{\underline{N}(x) + 1 - \underline{P}(x)}{2} && (10) \end{aligned}$$

x is classified to the positive class if $\mu_P(x) \geq \mu_N(x)$, otherwise it is classified to the negative class.

The main drawback of this method for imbalanced classification is that it treats all classes symmetrically, not making a distinction between majority and minority instances. The next section introduces a new strategy to deal with imbalanced data based on FRNN.

IV. Fuzzy-Rough Ordered Weighted Average Approach to Imbalanced Classification

As discussed in the previous section, FRNN treats the positive and negative class in a completely symmetric way and hence makes no provisions for the class imbalance. On the other hand, the classifications of the FRNN algorithm are completely determined by the closest samples in either class, which may be too naive a strategy, especially if noise is present in the data.

To deal with these problems, in this section we introduce the imbalanced fuzzy-rough ordered weighted average nearest neighbor (IFROWANN) classifier. Its general format is introduced in Section IV-A, while in Section IV-B, we propose different weighting strategies for the positive and the negative class and in Section IV-C, we consider different strategies to model the indiscernibility relation.

A. Imbalanced Fuzzy-Rough Ordered Weighted Average Nearest Neighbor Algorithm (IFROWANN)

In order to take into account not just the closest samples for a test instance, we rely on ordered weighted average (OWA) operators [47], which are recalled first. Given a sequence A of t real values $A = \langle a_1, \ldots, a_t \rangle$, and a weight vector $W = \langle w_1, \ldots, w_t \rangle$ such that $w_i \in [0,1]$ and $\sum_{i=1}^{t} w_i = 1$, the OWA aggregation of A by W is given by

$$OWA_W(A) = \sum_{i=1}^{t} w_i b_i$$

where $b_i = a_j$ if a_j is the i^{th} largest value in A. For instance, if $A = \langle 0.3, 0.1, 0.2 \rangle$, and $W = \langle 0.3, 0.2, 0.5 \rangle$, then

$$OWA_W(A) = 0.3 * 0.3 + 0.2 * 0.2 + 0.5 * 0.1 = 0.18$$

The OWA operator has the minimum and the maximum operator as a special case. Indeed, if $W = \langle 0, 0, \ldots, 1 \rangle$, then $OWA_W(A)$ will return the minimum value in A, while $W = \langle 1, 0, \ldots, \rangle$ will cause $OWA_W(A)$ to be the maximum of A. Furthermore, we can consider OWA weight vectors to model a wide variety of aggregation strategies different from min and max, and apply them in Eqs. (1)–(4).

In general, given OWA weight vectors W_P^l and W_N^l of length $t = |U|$, an implicator $\mathscr{I}$ and a fuzzy relation R, we can define the membership of a test instance x to the W_P^l-lower approximation of P, and to the W_N^l-lower approximation of N by

$$\begin{aligned} \underline{P}_{W_P^l}(x) &= \underset{y\in U}{OWA_{W_P^l}} \langle \mathscr{I}(R(x,y), P(y)) \rangle && (11)\\ \underline{N}_{W_N^l}(x) &= \underset{y\in U}{OWA_{W_N^l}} \langle \mathscr{I}(R(x,y), N(y)) \rangle, && (12) \end{aligned}$$

On the other hand, given OWA weight vectors W_P^u and W_N^u of length $t = |U|$ and a t-norm $\mathscr{T}$, we can define the membership of x to the W_P^u-upper approximation of P, and to the W_N^u-upper approximation of N by

$$\begin{aligned} \overline{P}_{W_P^u}(x) &= \underset{y\in U}{OWA_{W_P^u}} \langle \mathscr{T}(R(x,y), P(y)) \rangle && (13)\\ \overline{N}_{W_N^u}(x) &= \underset{y\in U}{OWA_{W_N^u}} \langle \mathscr{T}(R(x,y), N(y)) \rangle && (14) \end{aligned}$$

The following proposition shows that, similar to Section III, a relationship between the lower and upper approximation can be established when specific conditions are imposed on the logical connectives and weight vectors.

Proposition 1. *Let $\mathscr{I}$ and $\mathscr{T}$ be defined by $\mathscr{I}(a,b) = \max(1-a,b)$ and $\mathscr{T}(a,b) = \min(a,b)$, for a,b in $[0,1]$. Additionally, we impose the conditions $(W_P^u)_i = (W_N^l)_{t-i+1}$ and $(W_N^u)_i = (W_P^l)_{t-i+1}$, for $i = 1, \ldots, t$. Under these restrictions, $\overline{P}_{W_P^u}(x) = 1 - \underline{N}_{W_N^l}(x)$ and $\overline{N}_{W_N^u}(x) = 1 - \underline{P}_{W_P^l}(x)$, for any x in U.*

Proof. We rename the elements of U such that $U = \{y_1, \ldots, y_t\}$, where

$$\min(R(x,y_i), P(y_i)) \geq \min(R(x,y_j), P(y_j))$$

for $i \geq j$. Let $x \in U$, it holds that

$$\begin{aligned}
\overline{P}_{W_P^u}(x) &= OWA_{W_P^u}\underset{y\in U}{\langle}\mathscr{T}(R(x,y),P(y))\rangle \\
&= \sum_{i=1}^{t}(W_P^u)_i \min(R(x,y_i),P(y_i)) \\
&= \sum_{i=1}^{t}(W_N^l)_{t-i+1} \min(R(x,y_i),1-N(y_i)) \\
&= \sum_{i=1}^{t}(W_N^l)_{t-i+1}(1-\max(1-R(x,y_i),N(y_i))) \\
&= 1-\sum_{i=1}^{t}(W_N^l)_i \mathscr{I}(R(x,y_{t-i+1}),N(y_{t-i+1})) \\
&= 1-\underline{N}_{W_N^l}(x)
\end{aligned}$$

Analogously, we can establish that $\overline{N}_{W_N^u}(x)=1-\underline{P}_{W_P^l}(x)$. □

Assuming the conditions of Proposition 1, the IFROWANN algorithm then determines the classification of the test instance x by computing

$$\mu_P(x) = \frac{\underline{P}_{W_P^l}(x)+\overline{P}_{W_P^u}(x)}{2} = \frac{\underline{P}_{W_P^l}(x)+1-\underline{N}_{W_N^l}(x)}{2} \quad (15)$$

$$\mu_N(x) = \frac{\underline{N}_{W_N^l}(x)+\overline{N}_{W_N^u}(x)}{2} = \frac{\underline{N}_{W_N^l}(x)+1-\underline{P}_{W_P^l}(x)}{2} \quad (16)$$

Similarly as in FRNN, x is classified to the positive class if $\mu_P(x)\geq\mu_N(x)$, otherwise it is classified to the negative class.

B. OWA Weight Vectors for Imbalanced Classification

A crucial factor in the application of IFROWANN is the choice of the OWA weight vectors in Eqs. (11)–(14). Because of the relationship we assume between the lower and upper approximation, in this section we only focus on the former. In particular, we design weight vectors that provide flexible generalizations of the minimum operator, and at the same time take into account the imbalance present in the data.

First note that, under our assumptions, $\mathscr{I}(R(x,y),P(y))=1$ as soon as $P(y)=1$, in other words, when sample y is positive. Similarly, $\mathscr{I}(R(x,y),N(y))=1$ always holds when y is negative. It can be argued that these values should not be taken into account when computing the lower approximation; indeed, the commonsense interpretation of rough sets [40] states that an instance x belongs to the lower approximation of a class to the extent that it can be discerned (separated) from instances belonging to different classes; thus, instances from x's own class should not influence the instance's membership to the lower approximation.

In the context of the IFROWANN approach, we can implement this idea by assigning a weight of 0 to the corresponding positions in the OWA weight vectors. In particular, the first p positions in W_P^l can be put to 0, taking into account that they correspond to the highest values of $\mathscr{I}(R(x,y),P(y))$, and thus to the p positive samples in the training data.

The remaining n positions in the weight vector W_P^l correspond to the instances in N. For these instances, the implication values equal $\mathscr{I}(R(x,y),P(y))=\max(1-R(x,y),0)=1-R(x,y)$. We consider two alternative strategies to construct the weight vectors, both of which assign higher weights to the smaller implication values.

$$\begin{aligned}
W_P^{l_1} &= \left\langle 0,\ldots,0,\frac{2}{n(n+1)},\frac{4}{n(n+1)},\ldots,\right. \\
&\qquad \left.\frac{2(n-1)}{n(n+1)},\frac{2}{n+1}\right\rangle && (17) \\
W_P^{l_2} &= \left\langle 0,\ldots,0,\frac{1}{2^n-1},\frac{2}{2^n-1},\ldots,\right. \\
&\qquad \left.\frac{2^{n-2}}{2^n-1},\frac{2^{n-1}}{2^n-1}\right\rangle && (18)
\end{aligned}$$

The main difference between both vectors is that in the first case, weights decrease less rapidly than in the second case and are distributed more evenly among the instances. For instance, if $n=5$, then

$$W_P^{l_1} = \left\langle 0,\ldots,0,\frac{1}{15},\frac{2}{15},\frac{3}{15},\frac{4}{15},\frac{5}{15}\right\rangle \quad (19)$$

$$W_P^{l_2} = \left\langle 0,\ldots,0,\frac{1}{31},\frac{2}{31},\frac{4}{31},\frac{8}{31},\frac{16}{31}\right\rangle \quad (20)$$

In a completely analogous way, we can obtain two versions of the weight vectors $W_N^{l_1}$, where the first n positions are given a value of 0.

$$\begin{aligned}
W_N^{l_1} &= \left\langle 0,\ldots,0,\frac{2}{p(p+1)},\frac{4}{p(p+1)},\ldots,\right. \\
&\qquad \left.\frac{2(p-1)}{p(p+1)},\frac{2}{p+1}\right\rangle && (21) \\
W_N^{l_2} &= \left\langle 0,\ldots,0,\frac{1}{2^p-1},\frac{2}{2^p-1},\ldots,\right. \\
&\qquad \left.\frac{2^{p-2}}{2^p-1},\frac{2^{p-1}}{2^p-1}\right\rangle && (22)
\end{aligned}$$

Since typically p is a lot smaller than n, the obtained weight vectors for the positive and the negative classes will be quite different. However, for the second weighting strategy, we need to take into account that in practice, even for fairly small values of n and p, $W_P^{l_2}$ and $W_N^{l_2}$ soon approximate the fixed weight vector

$$W = \left\langle \ldots,\frac{1}{32},\frac{1}{16},\frac{1}{8},\frac{1}{4},\frac{1}{2}\right\rangle \quad (23)$$

For this reason, in our experiments we will also consider mixed approaches, where e.g. $W_P^{l_1}$ and $W_N^{l_2}$ are used in combination.

On the other hand, when n gets large, all the weights in $W_P^{l_1}$ become very small. Consequently, a similar phenomenon occurs as for the kNN (k Nearest Neighbor) classifier [12] when the number k of considered neighbors gets very high, i.e., the individual impact of instances gets diluted and the classification performance drops sharply. In order to mitigate this effect, we consider the following variant of $W_P^{l_1}$. Given $0\leq\gamma\leq 1$,

$$\begin{aligned}
W_P^{l_1,\gamma} &= \left\langle 0,\ldots,0,\frac{2}{r(r+1)},\frac{4}{r(r+1)},\ldots,\right. \\
&\qquad \left.\frac{2(r-1)}{r(r+1)},\frac{2}{r+1}\right\rangle && (24)
\end{aligned}$$

where $r=\lceil p+\gamma(n-p)\rceil$, and the first $t-r$ values of the vector are equal to 0. Clearly, $W_P^{l_1,0} = W_N^{l_1}$ and $W_P^{l_1,1} = W_P^{l_1}$. Hence, the number r of non-zero weights in $W_P^{l_1,\gamma}$ will always be between p and n. In our experiments, we will use a small value of γ, e.g. $\gamma = 0.1$, to limit the number of instances which receive strictly positive weights.

C. Indiscernibility Relation

Apart from the OWA weight vectors, we also need to make a choice for the fuzzy relation R. In order to determine the approximate indiscernibility between two instances x and y based on the set $\mathscr{A}$ of attributes, in this paper we assume the following definitions. Given a quantitative (i.e., real) attribute a,

$$R_a(x,y) = 1 - \frac{|a(x) - a(y)|}{range(a)} \tag{25}$$

while for a nominal attribute a,

$$R_a(x,y) = \begin{cases} 1 & \text{if } a(x) = a(y) \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

We establish the range of a feature based on the training data. In case a test sample has a value for a feature that lies outside this range, we dynamically change the range to take into account the extreme value.

We then consider the three following alternatives for defining the fuzzy relation R:

$$\begin{aligned} R_{\mathscr{T}_L}(x,y) &= \mathscr{T}_L(R_{a_1}(x,y),\ldots,R_{a_m}(x,y)) && (27)\\ R_{Min}(x,y) &= \min(R_{a_1}(x,y),\ldots,R_{a_m}(x,y)) && (28)\\ R_{Av}(x,y) &= \frac{R_{a_1}(x,y)+\ldots+R_{a_m}(x,y)}{m} && (29) \end{aligned}$$

where the Łukasiewicz t-norm $\mathscr{T}_L$ is defined by, for $u_1,u_2,\ldots,u_m$ in $[0,1]$,

$$\mathscr{T}_L(u_1,u_2,\ldots,u_m) = \max(u_1+u_2+\ldots+u_m-m,0). \tag{30}$$

It can be easily checked that $R_{\mathscr{T}_L}(x,y) \le R_{Min}(x,y) \le R_{Av}(x,y)$ always holds. In other words, $R_{\mathscr{T}_L}$ provides a comparatively more conservative (lower) estimate for the similarity between x and y, while R_{Av} provides a more liberal (higher) one, and R_{Min} is in between the two. In the next sections, we will evaluate the impact of this choice on the results of our experiments.

V. Experimental Setup

In this section, we describe the experimental framework used to validate our proposal, including the benchmark data sets, the particular configurations considered for IFROWANN and for the baseline and state-of-the-art methods, and the statistical tests used in order to carry out the performance comparison.

A. Data sets

We consider 102 data sets with different imbalance ratios (between 1.82 and 129.44) to evaluate our proposal. They originate from the UCI repository [3] and were obtained by modifying multiple class data sets into two-class imbalanced problems. To create a new two-class data set, we take one or more small classes versus one or more of the remaining classes. The name of the resulting data set references the original classes used in the construction, for instance: in ecoli-0-1-3-7vs2-6 the first class consists of class0, class1, class3 and class7 from the original ecoli data set, while the second is composed of class2 and class6. The characteristics of these data sets can be found in Table II, showing the imbalance ratio (IR), the number of instances (Inst) and the number of attributes (Attr) for each of them.

Apart from considering the data set collection as a whole, in our experimental study we have also considered three subsets of the collection based on their IR. The purpose of this division is to evaluate the behavior of the algorithms at different imbalance levels.

1) $IR < 9$ (low imbalance): This group contains 22 data sets, all with IR lower than 9.
2) $IR \geq 9$ (high imbalance): This group contains 80 data sets, all with IR at least 9.
3) $IR \geq 33$ (very high imbalance): This group contains 31 data sets, all with IR at least 33. This is a subset of the collection considered in the second case.

Furthermore, each data set is partitioned in order to perform a five fold cross-validation (5FCV). The partitions were built in such a way that the quantity of elements in each class remains uniform [17]. The data sets are available online[2] as part of the KEEL data set repository [1], [2].

B. Algorithms Analyzed in the Experimental Study

1) IFROWANN: based on the proposals in Section IV-B, we consider the following six configurations for the IFROWANN weight vectors:

1) $\mathscr{W}_1 = \langle W_P^{l_1}, W_N^{l_1} \rangle$
2) $\mathscr{W}_2 = \langle W_P^{l_1}, W_N^{l_2} \rangle$
3) $\mathscr{W}_3 = \langle W_P^{l_2}, W_N^{l_1} \rangle$
4) $\mathscr{W}_4 = \langle W_P^{l_2}, W_N^{l_2} \rangle$
5) $\mathscr{W}_5 = \langle W_P^{l_1,\gamma}, W_N^{l_1} \rangle$ with $\gamma = 0.1$
6) $\mathscr{W}_6 = \langle W_P^{l_1,\gamma}, W_N^{l_2} \rangle$ with $\gamma = 0.1$

In order to check the robustness of the parameter γ in the last two configurations, we will also perform a sensitivity analysis with γ taking values between 0 and 1.

Each of these weight vectors will be combined with the three indiscernibility relations considered in Section IV-C. The resulting 18 combinations will be denoted TL-$\mathscr{W}_i$, MIN-$\mathscr{W}_i$ and AV-$\mathscr{W}_i$, with $i = 1,\ldots,6$.

2) Baseline Methods—FRNN and IFROWANN using Equal Weight Vectors: Apart from comparing IFROWANN with the original FRNN algorithm, we also want to demonstrate the importance of using different weight vectors for the positive and the negative class. For this reason, we will consider a particular configuration of IFROWANN, denoted $\mathscr{W}_7 = \langle W^l, W^l \rangle$, in which equal weight vectors are used for both classes:

$$\begin{aligned} W^l \;=\; \Big\langle & \frac{2}{(n+p)(n+p+1)}, \frac{4}{(n+p)(n+p+1)}, \ldots, \\ & \frac{2(n+p-1)}{(n+p)(n+p+1)}, \frac{2}{n+p+1} \Big\rangle \end{aligned} \tag{31}$$

[2] See http://www.keel.es/datasets.php.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2014.2371472, IEEE Transactions on Fuzzy Systems

Table II: Description of the data sets used in the experimental evaluation.

Dataset	IR	Inst	Attr	Dataset	IR	Inst	Attr
glass1	1.82	214	9	ecoli4	15.8	336	7
ecoli-0vs1	1.86	220	9	page-blocks-1-3vs4	15.86	472	10
wisconsinImb	1.86	683	7	abalone9-18	16.4	731	8
iris0	2	150	4	glass-0-1-6vs5	19.44	184	9
glass0	2.06	214	9	shuttle-c2-vs-c4	20.5	129	9
yeast1	2.46	1484	8	cleveland-4	21.85	297	13
habermanImb	2.78	306	3	shuttle-6vs2-3	22	230	9
vehicle2	2.88	846	18	yeast-1-4-5-8vs7	22.1	693	8
vehicle1	2.9	846	18	ionosphere-bredvsg	22.5	235	33
vehicle3	2.99	846	18	glass5	22.78	214	9
glass-0-1-2-3vs4-5-6	3.2	214	9	yeast-2vs8	23.1	482	8
vehicle0	3.25	846	18	wdbc-MredBvsB	23.8	372	30
ecoli1	3.36	336	7	texture-2redvs3-4	23.81	1042	40
appendicitisImb	4.05	106	7	yeast4	28.1	1484	8
new-thyroid1	5.14	215	5	winequalityred-4	29.17	1599	11
new-thyroid2	5.14	215	5	kddcup-guess-passwdvssatan	29.98	1642	41
ecoli2	5.46	336	7	yeast-1-2-8-9vs7	30.57	947	8
segment0	6.02	2308	19	abalone-3vs11	32.47	502	8
glass6	6.38	214	9	winequalitywhite-9vs4	32.6	168	77
yeast3	8.1	1484	8	yeast5	32.73	1484	8
ecoli3	8.6	336	7	winequalityred-8vs6	35.44	656	11
page-blocks0	8.79	5472	10	ionosphere-bredBvsg	37.5	231	33
ecoli-0-3-4vs5	9	200	7	ecoli-0-1-3-7vs2-6	39.14	281	7
ecoli-0-6-7vs3-5	9.09	222	7	abalone-17vs7-8-9-10	39.31	2338	8
yeast-2vs4	9.1	515	7	abalone-21vs8	40.5	581	8
ecoli-0-2-3-4vs5	9.1	202	7	yeast6	41.4	1484	8
glass-0-1-5vs2	9.12	172	9	segment-7redvs2-4-5-6	42.58	1351	19
yeast-0-3-5-9vs7-8	9.12	506	8	winequalitywhite-3vs7	44	900	11
yeast-0-2-5-6vs3-7-8-9	9.14	1004	8	wdbc-MredvsB	44.63	365	30
yeast-0-2-5-7-9vs3-6-8	9.14	1004	8	segment-5redvs1-2-3	45	1012	19
ecoli-0-4-6vs5	9.15	203	6	winequalityred-8vs6-7	46.5	855	11
ecoli-0-1vs2-3-5	9.17	244	7	phoneme-1redvs0red	46.98	2543	5
ecoli-0-2-6-7vs3-5	9.18	224	7	texture-6redvs7-8	47.62	1021	40
glass-0-4vs5	9.22	92	9	kddcup-landvsportsweep	49.52	1061	41
ecoli-0-3-4-6vs5	9.25	205	7	abalone-19vs10-11-12	49.69	1622	8
ecoli-0-3-4-7vs5-6	9.28	257	7	magic-hredvsgred	54.1	2645	10
yeast-0-5-6-7-9vs4	9.35	528	8	winequalitywhite-3-9vs5	58.28	1482	11
ecoli-0-6-7vs5	10	220	6	shuttle-2vs5	66.67	3316	9
glass-0-1-6vs2	10.29	192	9	winequalityred-3vs5	68.1	691	11
ecoli-0-1-4-7vs2-3-5-6	10.59	336	7	phoneme-1redBvs0redB	69.7	2333	5
ecoli-0-1vs5	11	240	6	texture-12redvs13-14	71.43	1014	40
glass-0-6vs5	11	108	9	abalone-20vs8-9-10	72.69	1916	8
glass-0-1-4-6vs2	11.06	205	9	kddcup-bufferoverowvsback	73.43	2233	41
glass2	11.59	214	9	kddcup-landvssatan	75.67	1610	41
ecoli-0-1-4-7vs5-6	12.28	332	7	shuttle-2vs1red	81.63	4049	9
cleveland-0vs4	12.31	173	13	segment-6redvs3-4-5	82.5	1002	19
ecoli-0-1-4-6vs5	13	280	6	shuttle-6-7vs1red	86.96	2023	9
movement-libras-1	13	336	90	magic-hredBvsgredB	88	2403	10
shuttle-c0-vs-c4	13.87	1829	9	texture-7redvs2-3-4-6	95.24	2021	40
yeast-1vs7	14.3	459	7	kddcup-rootkit-imapvsback	100.14	2225	41
glass4	15.46	214	9	abalone19	129.44	4174	8

Each of the above baseline configurations will be combined with the same indiscernibility relations considered above. We denote the resulting methods TL-FRNN, MIN-FRNN, AV-FRNN, TL-$\mathscr{W}_7$, MIN-$\mathscr{W}_7$ and AV-$\mathscr{W}_7$.

3) State-of-the-Art Methods: as discussed in Section II-A, we will consider the following imbalanced learning methods to compare our method with:

- SMOTE
- SMOTE-ENN
- SMOTE-RSB$_*$
- CS-C4.5
- CS-SVM
- EUSBOOST
- HDDT+Bagging

The first three methods are preprocessing techniques, so they need to be combined with a base classifier. We chose three well known classifiers, representing lazy learners, decision tree-based methods and support vector machines, respectively:

- kNN [12]
- C4.5 [37]
- SVM [43]

The parameters of all the resulting thirteen proposals which were used in our experimentation are described in Table III. For detailed explanation of these parameters, we refer to the corresponding articles. For the kNN method, in order to set the number of neighbors optimally, we used the best value of k for each data set, obtained by trying all values between 1 and the total number of training instances, with 100 equidistant steps. Figure 1 shows this analysis, averaged over all data sets.

C. Statistical tests for performance comparison

In order to compare the different algorithms appropriately, we will conduct a statistical analysis using non-parametric tests as suggested in the literature [13], [23], [24].

We first use Friedman's aligned-ranks test [19] to detect statistical differences among a set of algorithms. The Friedman test computes the average aligned-ranks of each algorithm, obtained by computing the difference between the performance of the algorithm and the mean performance of all algorithms for each data set. The lower the average rank, the better the corresponding algorithm.

Then, if significant differences are found by the Friedman test, we check if the control algorithm (the one obtaining the smallest rank) is significantly better than the others using Holm's post hoc test [29]. The post hoc procedure allows us to decide whether a hypothesis of comparison can be rejected at a

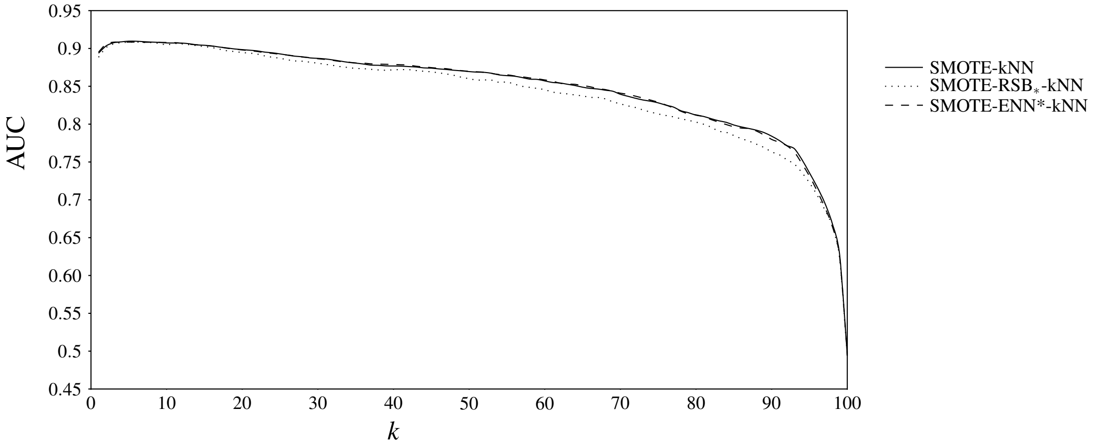
Table III: Parameters of the state-of-the-art methods for the experimental study.

Algorithm	Parameters
SMOTE	Number of Neighbors = 5, Type of SMOTE = both, Balancing = YES Quantity of generated examples = 1, Distance Function = HVDM, Type of Interpolation = standard
SMOTE-ENN	Number of Neighbors ENN = 3, Number of Neighbors SMOTE = 5, Type of SMOTE = both, Balancing = YES Quantity of generated examples = 1, Distance Function (SMOTE) = HVDM, Distance Function (ENN) = Euclidean
SMOTE-RSB$_*$	Number of Neighbors = 5, Type of neighbors = Both, Balance = Yes, Smoting = 1 Type of Interpolation = standard, Cutoffini = 0.6, Cutoffinal = 0.9
kNN	Distance Function = Euclidean
C4.5	pruned = TRUE, confidence = 0.25, instancesPerLeaf = 2

SVM	c = 1.0, Tolerance Parameter = 0.001, epsilon = 1.0E-12, Kernel Type = polynomial
	Normalized PolyKernel exponent = 1.0, Normalized PolyKernel use Lower Order = False
	FitLogisticModels = TRUE, ConvertNominalAttributesToBinary = True, PreprocessType = Normalize

EUSBOOST	pruned = TRUE, confidence = 0.25, instancesPerLeaf = 2, Number of Classifiers = 10, Algorithm = ERUSBOOST Train Method = NORESAMPLING, Quantity of balancing SMOTE = 50, IS Method = HammingEUB_M_GM
C4.5-CS	pruned = TRUE, confidence = 0.25, instancesPerLeaf = 2, minimumExpectedCost = TRUE
SVM-CS	Kernel Type = polynomial, C = 100.0, eps = 0.001 degree = 1, gamma = 0.01, coef0 = 0.0, nu = 0.1, p = 1.0, shrinking = 1
HDDT+Bagging	For Bagging: bagSizePercent = 100, calcOutOfBag = false, numIterations = 100 For HDDT: binarySplits = true, collapse = false, confidenceFactor = 0.25, minNumObj = 2, reducedErrorPruning = false saveInstanceData = false, subtreeRaising = true, unpruned = false, useLaplace = false

Figure 1: Tuning of the k parameter for kNN with SMOTE, SMOTE-RSB$_*$ and SMOTE-ENN.



specified level of significance α. In this paper, we set $\alpha = 0.05$. In practice, it is very interesting to compute the adjusted p-value, which represents the lowest level of significance of a hypothesis that results in a rejection. In this manner, we can find out whether two algorithms are significantly different and how different they are.

VI. Experimental Results

In this section, we present the results of our experimental analysis[3]. In Section VI-A, we first compare the 18 variants of IFROWANN over the entire collection of 102 data sets. Next, in Section VI-B, we provide a detailed analysis for different IR levels (low IR, high IR and very high IR). Section VI-C compares our proposal with the baseline methods FRNN and $\mathscr{W}_7$. Furthermore, in Section VI-D we compare the algorithms that perform best in the first analysis with the state-of-the-art methods for imbalanced classification. Finally, Section VI-E provides a graphical analysis.

A. Comparative Analysis between IFROWANN Variants over All Data Sets

Table IV shows the mean AUC obtained for 18 variants of IFROWANN. We can see that AV-$\mathscr{W}_6$ obtains the highest average AUC. There are also some quite noticeable differences between the results obtained with each indiscernibility relation (TL, AV, MIN). The best general results are obtained with AV, while there are no great performance difference between TL and MIN.

Table IV: Mean AUC for IFROWANN variants over all data sets. The values marked in light blue (values higher than 0.91) are taken into account in the statistical analysis.

Algorithm	**AUC**	**Algorithm**	**AUC**	**Algorithm**	**AUC**
TL-$\mathscr{W}_1$	0.8943	AV-$\mathscr{W}_1$	0.9098	MIN-$\mathscr{W}_1$	0.8908
TL-$\mathscr{W}_2$	0.8802	AV-$\mathscr{W}_2$	0.9094	MIN-$\mathscr{W}_2$	0.8908
TL-$\mathscr{W}_3$	0.8893	AV-$\mathscr{W}_3$	0.8990	MIN-$\mathscr{W}_3$	0.8813
TL-$\mathscr{W}_4$	0.8998	AV-$\mathscr{W}_4$	0.9181	MIN-$\mathscr{W}_4$	0.9030
TL-$\mathscr{W}_5$	0.8928	AV-$\mathscr{W}_5$	0.9122	MIN-$\mathscr{W}_5$	0.8955
TL-$\mathscr{W}_6$	0.9054	AV-$\mathscr{W}_6$	**0.9256**	MIN-$\mathscr{W}_6$	0.9071

Next, we can also notice several differences between the weighting strategies, which are summarized below:

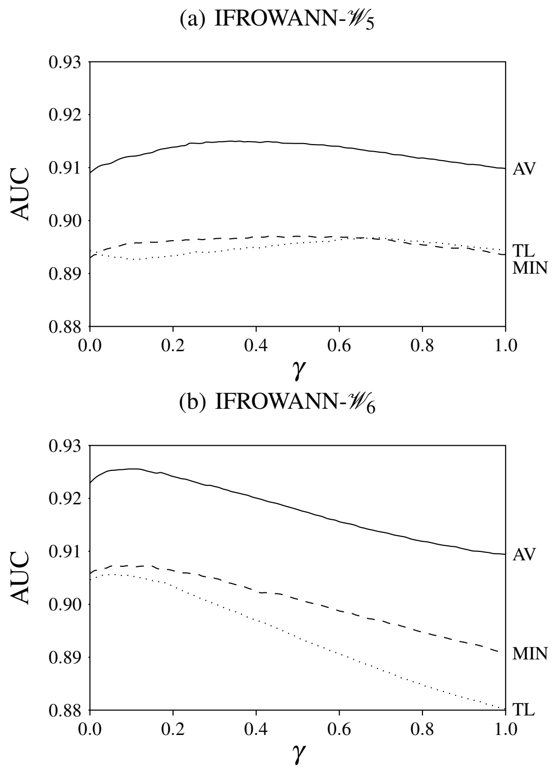
- Exponentially decreasing weights ($\mathscr{W}_4$) outperform linearly decreasing weights ($\mathscr{W}_1$).
- Mixing different weighting strategies ($\mathscr{W}_2$ and $\mathscr{W}_3$) generally lowers the results compared to $\mathscr{W}_1$, and thus also compared to $\mathscr{W}_4$.
- Varying $\mathscr{W}_1$ to only weigh a fraction of the negative instances ($\mathscr{W}_5$) improves the results when using AV and MIN; yet, they remain inferior to those of $\mathscr{W}_4$. On the

[3] The detailed results, per method and per data set, are available online at the website associated to this paper, http://sci2s.ugr.es/frowa-imbalanced/

other hand, this strategy slightly lowers the results when TL is used. Figure 2a shows the sensitivity analysis for γ, obtained over all data sets when this parameter moves between 0 and 1. As can be seen, the results are overall very stable. The TL curve shows a slight performance drop for small values of γ, which might explain why results worse than $\mathscr{W}_1$ are obtained in this case.

- Varying $\mathscr{W}_4$ to only a fraction of the negative instances ($\mathscr{W}_6$) benefits the classification for high IR data sets, but slightly deteriorates it for low IR data sets. The sensitivity analysis in Figure 2b shows that the best results are obtained for low values of γ, which justifies our choice of $\gamma = 0.1$.

Figure 2: Sensitivity analysis for the parameter γ in the weighting strategies $\mathscr{W}_5$ and $\mathscr{W}_6$, evaluated over all data sets.



(a) IFROWANN-$\mathscr{W}_5$

(b) IFROWANN-$\mathscr{W}_6$

We proceed with the statistical analysis of our results. In order to reduce the number of variants considered in the test, and thus increase its discriminatory power, we have selected only the highest scoring proposals (AUC higher than 0.91). Such values are marked in light blue in Table IV.

The average ranks of the algorithms and the adjusted p-values obtained by Holm's post-hoc procedure are shown in Table V. The p-value computed by Friedman test is 0.003426, which indicates that the hypothesis of equivalence can be rejected with high confidence.

Table V: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for all data sets, using AV-$\mathscr{W}_6$ as the control algorithm.

Algorithm	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_6$	1.7549	-
AV-$\mathscr{W}_4$	2.0196	0.058707
AV-$\mathscr{W}_5$	2.2255	0.001555

As we can observe, AV-$\mathscr{W}_6$ obtains the lowest ranking of the algorithms used which turns it into the control method. The adjusted p-values are low enough to reject the null hypothesis with a high confidence level for AV-$\mathscr{W}_4$ and for AV-$\mathscr{W}_5$. This confirms that AV-$\mathscr{W}_6$ is indeed the best overall IFROWANN configuration.

B. Comparative Analysis between IFROWANN Variants for Different Levels of Data Imbalance

Table VI shows the mean AUC obtained for each method and each block of data sets. Every row represents one variant of IFROWANN, and the columns represent the data set groups based on IR. For every column, the highest AUC value is marked in bold.

Table VI: Mean AUC for IFROWANN variants for different IR levels. The values marked in light blue (values higher than 0.91) are taken into account in the statistical analysis.

Method	**<9**	**≥9**	**≥33**
TL-$\mathscr{W}_1$	0.9186	0.8877	0.8845
TL-$\mathscr{W}_2$	0.9076	0.8727	0.8424
TL-$\mathscr{W}_3$	0.8935	0.8882	0.8941
TL-$\mathscr{W}_4$	0.9180	0.8948	0.8959
TL-$\mathscr{W}_5$	0.9163	0.8863	0.8925
TL-$\mathscr{W}_6$	0.9148	0.9028	0.8989
AV-$\mathscr{W}_1$	0.9014	0.9121	0.9023
AV-$\mathscr{W}_2$	0.9029	0.9112	0.8938
AV-$\mathscr{W}_3$	0.8900	0.9014	0.8938
AV-$\mathscr{W}_4$	**0.9232**	0.9167	0.9073
AV-$\mathscr{W}_5$	0.9068	0.9136	0.9030
AV-$\mathscr{W}_6$	0.9139	**0.9288**	0.9166
MIN-$\mathscr{W}_1$	0.8844	0.8961	0.9062
MIN-$\mathscr{W}_2$	0.8809	0.8935	0.8990
MIN-$\mathscr{W}_3$	0.8713	0.8841	0.8975
MIN-$\mathscr{W}_4$	0.9101	0.9010	0.9156
MIN-$\mathscr{W}_5$	0.8877	0.8977	0.9085
MIN-$\mathscr{W}_6$	0.8925	0.9111	**0.9230**

It can be noticed that for the high imbalance data sets ($IR \geq 9$), AV-$\mathscr{W}_6$ still obtains the highest average AUC. However, for low imbalance data sets ($IR < 9$), AV-$\mathscr{W}_4$ reaches the highest value, and for very high imbalance data sets ($IR \geq 33$), the best variant is MIN-$\mathscr{W}_6$.

Below, we carry out a statistical analysis of our results for each block of data sets. As before, we consider only the proposals obtaining a mean AUC higher than 0.91. Such values are marked in light blue in Table VI.

1) Statistical analysis for low IR data sets: For the low IR data sets, seven proposals are selected. Table VII shows the average ranking obtained by the Friedman test. As we can observe, the best ranking is obtained by AV-$\mathscr{W}_4$. The p-value computed by the Friedman Test is 0.082069, which is low enough to conclude that there are significant differences among the algorithms.

Based on the adjusted p-values, the Holm post hoc test allows to conclude that the control method AV-$\mathscr{W}_4$ is significantly better than MIN-$\mathscr{W}_4$. The fairly low adjusted p-values for AV-$\mathscr{W}_6$ and TL-$\mathscr{W}_6$ also suggest that in this case $\mathscr{W}_4$ is indeed a better weighting strategy than $\mathscr{W}_6$.

Table VII: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for low imbalance data sets, using AV-$\mathscr{W}_4$ as the control algorithm.

Algorithm	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_4$	2.9545	-
TL-$\mathscr{W}_4$	3.3864	0.507350
AV-$\mathscr{W}_6$	4.0455	0.193227
TL-$\mathscr{W}_6$	4.1591	0.193227
TL-$\mathscr{W}_1$	4.2500	0.186845
TL-$\mathscr{W}_5$	4.3864	0.139650
MIN-$\mathscr{W}_4$	4.8182	0.025319

2) Statistical analysis for high IR data sets: Table VIII shows the average ranking obtained by the Friedman test for the five proposals selected in this case. The p-value computed by the Friedman test is approximately 0, which indicates that the hypothesis of equivalence can be rejected with high confidence. As we can observe, the best ranking is obtained by AV-$\mathscr{W}_6$ which is used as the control algorithm. The adjusted p-values are all very low, indicating that the method AV-$\mathscr{W}_6$ significantly outperforms the remaining methods when high IR data sets are considered.

Table VIII: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for high imbalance data sets, using AV-$\mathscr{W}_6$ as the control algorithm.

Algorithm	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_6$	2.1000	-
AV-$\mathscr{W}_5$	3.0187	0.000238
AV-$\mathscr{W}_4$	3.0875	0.000156
AV-$\mathscr{W}_2$	3.3312	0.000003
AV-$\mathscr{W}_1$	3.4625	0.000000

3) Statistical analysis for very high IR data sets: Table IX shows the average ranking obtained by the Friedman test for the three selected proposals. The Friedman p-value in this case is 0.706965, which indicates that the hypothesis of equivalence of the five considered methods can be accepted. As we can observe, the best ranking is obtained by AV-$\mathscr{W}_6$.

Table IX: Average Friedman rankings for very high imbalance data sets. The Friedman test does not discover significant differences, so Holm's test is not performed.

Algorithm	Average Friedman ranking
AV-$\mathscr{W}_6$	1.8871
MIN-$\mathscr{W}_6$	2.0161
MIN-$\mathscr{W}_4$	2.0968

C. Comparative Analysis of IFROWANN and Baseline Methods

Table X shows the results over all 102 data sets of the basic FRNN algorithm and the IFROWANN baseline configuration $\mathscr{W}_7$ employing equal weight vectors for both classes, combined with the three indiscernibility relations TL, AV and MIN. The table also shows the results obtained with the best three IFROWANN variants AV-$\mathscr{W}_4$, AV-$\mathscr{W}_5$ and AV-$\mathscr{W}_6$.

Table X: Mean AUC for baseline methods and best IFROWANN variants over all data sets.

Method	AUC
TL-$\mathscr{W}_7$	0.8288
AV-$\mathscr{W}_7$	0.7798
MIN-$\mathscr{W}_7$	0.7754
TL-FRNN	0.8905
AV-FRNN	0.9083
MIN-FRNN	0.8925
AV-$\mathscr{W}_4$	0.9181
AV-$\mathscr{W}_5$	0.9122
AV-$\mathscr{W}_6$	0.9256

As can be seen in Table X, considering equal weight vectors affects the results adversely, causing a drop in AUC of over 10%. This clearly shows the advantage of using different weight vectors for the positive and negative class. On the other hand, when the basic FRNN algorithm is used, we obtain fairly good results. However, these results rank below those obtained with the best IFROWANN variants.

We support the comparison with a statistical analysis in order to demonstrate the superiority of our proposal. The average ranks of the algorithms and the adjusted p-values obtained by Holm's post-hoc procedure are shown in Table XI. The p-value computed by the Friedman test is approximately 0, which indicates that the hypothesis of equivalence can be rejected with high confidence. From Table XI, we can conclude that the control algorithm AV-$\mathscr{W}_6$ obtains significantly better results than all baseline methods.

Table XI: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for all data sets, using AV-$\mathscr{W}_6$ as the control algorithm.

Method	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_6$	2.6667	-
AV-$\mathscr{W}_4$	3.0147	0.364103
AV-$\mathscr{W}_5$	3.7157	0.012457
AV-FRNN	4.1765	0.000247
TL-FRNN	4.6618	0.000001
MIN-FRNN	5.1127	<0.000001
TL-$\mathscr{W}_7$	6.6324	<0.000001
AV-$\mathscr{W}_7$	7.3529	<0.000001
MIN-$\mathscr{W}_7$	7.6667	<0.000001

D. Comparative analysis with the state-of-the-art methods

The experimental study carried out in Section VI-A and VI-B shows that the best two proposals are $\mathscr{W}_4$ in the case of low IR data sets and $\mathscr{W}_6$ in the remaining cases. This section compares these two methods with the state-of-the-art methods. The mean AUC results for the different blocks are shown in Table XII.

Table XII: Mean AUC for state-of-the-art methods and the best IFROWANN variants. The values marked in light blue (values higher than 0.91) are taken into account in the statistical analysis.

Method	**all**	**<9**	**>9**	**>33**
SMOTE-kNN	0.9096	0.9143	0.9083	0.8987
SMOTE-C4.5	0.8315	0.8604	0.8235	0.8050
SMOTE-SVM	0.9000	0.9051	0.8986	0.9133
SMOTE-ENN-kNN	0.8839	0.9093	0.8769	0.8320
SMOTE-ENN-C4.5	0.8412	0.8714	0.8329	0.8218
SMOTE-ENN-SVM	0.9005	0.9046	0.8994	0.9130
C4.5-CS	0.8263	0.8691	0.8146	0.8083
SVM-CS	0.8952	0.9137	0.8901	0.9032
EUSBOOST	0.9094	0.9263	0.9048	0.8977
SMOTE-RSB$_*$-kNN	0.9085	0.9119	0.9076	0.8975
SMOTE-RSB$_*$-C4.5	0.8266	0.8681	0.8152	0.8021
SMOTE-RSB$_*$-SVM	0.9001	0.9036	0.8991	0.9130
HDDT+Bagging	0.9158	**0.9281**	0.9124	0.9019
AV-$\mathscr{W}_4$	0.9181	0.9232	0.9167	0.9073
AV-$\mathscr{W}_6$	**0.9256**	0.9139	**0.9288**	**0.9166**

From these results, we can observe that $\mathscr{W}_6$ obtains the highest AUC value in all blocks, except for low IR data sets for which HDDT+Bagging gets the highest score. Again, we will subject these results to a thorough statistical analysis. In this case, per block we take into account the methods which obtain a mean AUC of at least 0.9. These methods are marked in light blue in Table XII.

1) Statistical analysis for all data sets: Table XIII shows the average ranking obtained by the Friedman test. The p-value computed by the Friedman test is 0.000011, which indicates that the hypothesis of equivalence can be rejected with high confidence. As we can observe, the best ranking is obtained by AV-$\mathscr{W}_6$. Moreover, the adjusted p-values are all very low, so we may conclude that AV-$\mathscr{W}_6$ statistically outperforms all of them.

Table XIII: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for all data sets, using AV-$\mathscr{W}_6$ as the control algorithm.

Algorithm	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_6$	3.299	-
HDDT+Bagging	4.2941	0.003718
SMOTE-RSB$_*$-kNN	4.5147	0.000787
EUSBOOST	4.5833	0.000543
SMOTE-kNN	4.6275	0.000430
SMOTE-RSB$_*$-SVM	4.7304	0.000150
SMOTE-SVM	4.8186	0.000056
SMOTE-ENN-SVM	5.1324	0.000001

2) Statistical analysis for low IR data sets: In Table XIV, the results of applying the Friedman test are shown. In this case, the associated p-value is 0.204917, which is not low enough to reject the hypothesis of equivalence and which leads us to conclude that there are no statistically significant differences among the compared methods. Note that while EUSBOOST obtains the highest AUC mean for this block, the lowest Friedman rank is obtained by AV-$\mathscr{W}_4$.

Table XIV: Average Friedman rankings for low imbalance data sets. The Friedman test does not discover significant differences, so Holm's test is not performed.

Algorithm	Average Friedman ranking
AV-$\mathscr{W}_4$	3.9091
HDDT+Bagging	4.8636
SVM-CS	5.0227
EUSBOOST	5.2955
SMOTE-kNN	5.6136
SB-kNN	5.7500
SMOTE-SVM	5.8409
SB-SVM	6.0682
SMOTE-ENN-kNN	6.2727
SMOTE-ENN-SVM	6.3636

3) Statistical analysis for high IR data sets: the results, shown in Table XV, are concordant with those obtained for all data sets. The p-value computed by the Friedman test is smaller than 0.000001. AV-$\mathscr{W}_6$ obtains the best ranking and significantly outperforms all the remaining methods.

Table XV: Average Friedman rankings and adjusted p-values using Holm's post-hoc procedure for high imbalance data sets, using AV-$\mathscr{W}_6$ as the control algorithm.

Algorithm	Average Friedman ranking	Adjusted p-value
AV-$\mathscr{W}_6$	2.0125	-
HDDT+Bagging	2.9812	0.000107
SMOTE-RSB$_*$-kNN	3.25	0.000001
EUSBOOST	3.2938	0.000001
SMOTE-kNN	3.4625	<0.000001

4) Statistical analysis for very high IR: in this case, the p-value computed by the Friedman test is 0.574894, which indicates that the hypothesis of equivalence between the five considered methods can be accepted. In Table XVI, the Friedman aligned ranks are shown. It is interesting to note that in this case, SMOTE-RSB$_*$-SVM gets the best rank, while AV-$\mathscr{W}_6$ obtains the highest mean AUC.

Table XVI: Average Friedman rankings for very high imbalance data sets. The Friedman test does not discover significant differences, so Holm's test is not performed.

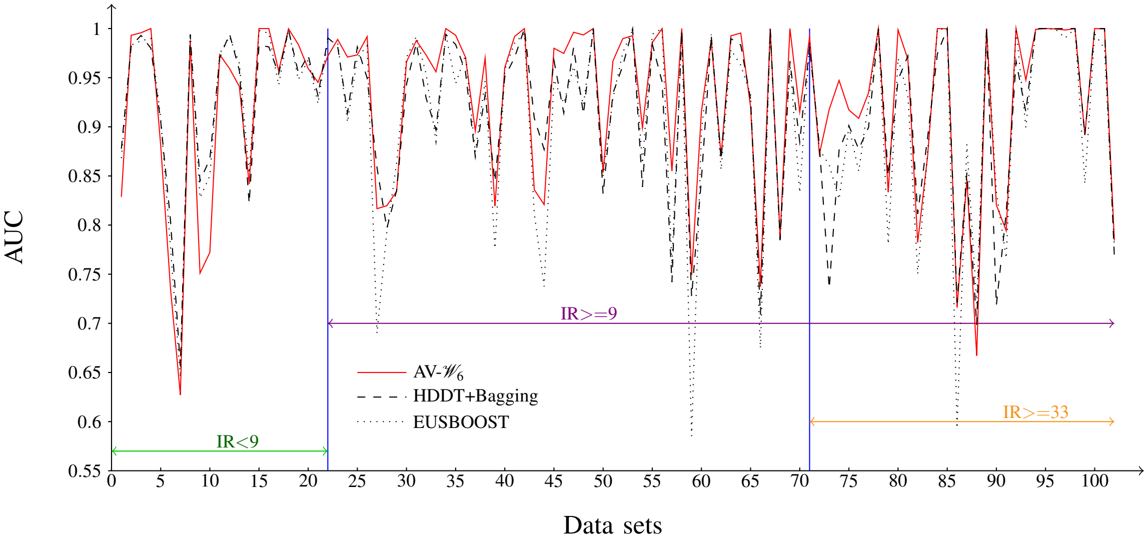
Algorithm	Average Friedman ranking
SMOTE-RSB$_*$-SVM	3.0968
SMOTE-SVM	3.2258
SMOTE-ENN-SVM	3.4516
SVM-CS	3.6613
HDDT+Bagging	3.7258
AV-$\mathscr{W}_6$	3.8387

E. Graphical analysis

To complement the statistical study from the previous section, we have also provided a graphical analysis that compares the behavior of our best two proposals (AV-$\mathscr{W}_6$ and AV-$\mathscr{W}_4$) to its closest competitors among the state-of-the-art methods. To this end, Figure 3 plots the considered method's AUC (Y axis) for all data sets, which are ordered on the X axis according to their IR. Similarly, in Figure 4, we show a more fine-grained analysis depicting the results for each of the experiment blocks, considering in each case the best performing algorithms.

In both figures, we can see that for both low and very high IR data sets, the compared methods behave more or less similarly, and that the most noticeable differences are in the middle section (IR between 9 and 33), where our method AV-$\mathscr{W}_6$ clearly shows the best performance.

Figure 3: AUC for all data sets, ordered according to their IR, for our best proposal (AV-$\mathscr{W}_6$) and the best algorithms from the state-of-the-art (SMOTE-RSB$_*$-kNN and EUSBOOST).



VII. Concluding remarks

In this paper, we have presented the Imbalanced Fuzzy-Rough Ordered Weighted Average Nearest Neighbor (IFROWANN) method, a new algorithm level solution to two-class imbalanced classification problems that is based on the Fuzzy-Rough Nearest Neighbor (FRNN) method and on Ordered Weighted Average (OWA) aggregation. In particular, we considered six weighting strategies, combined with three different indiscernibility relations.

Our experimental results and statistical analysis have shown that IFROWANN can outperform not only the classical FRNN algorithm over a large collection of imbalanced data sets with varying IR degrees, but also a selection of state-of-the-art representative algorithms that cover algorithm level, cost-sensitive and ensemble solutions specifically designed for imbalanced learning.

For future work, we will consider the integration of IFROWANN within ensemble methods, where it can be combined with data level (preprocessing) techniques to further optimize the classification performance. Another possible refinement of the approach concerns the automated extraction of OWA weight vectors and indiscernibility relations from the training data, using either a wrapper method, or basing ourselves on data characteristics, such as the imbalance ratio or other data complexity measures.

Finally, our third idea for future work is to extend IFROWANN to handle multi-class problems. One solution is to transform a multi-class problem into a two-class problem using binarization techniques such as the One-vs-One Approach (OVO) introduced by Hastie and Tibshirani [27], and the One-vs-All Approach (OVA) of Rifkin and Klautau [39]. In [18], the authors presented a complete experimental study for the classification of multi-class imbalanced data sets, which concluded that the OVO strategy is a better option than OVA. This allows us to design a new method for multi-class problems combining OVO and IFROWANN. Another solution will be to modify the IFROWANN itself to directly operate with multi-class problems.

References

[1] J. Alcalá, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17:255–287, 2010.

[2] J. Alcalá, L. Sánchez, S. García, M.J. del Jesús, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, and F. Herrera. KEEL: A software tool to assess evolutionary algorithms to data mining problems. *Soft Computing*, 13:3:307–318, 2009.

[3] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.

[4] G. E. A. P. A. Batista, R. C. Prati, and M.C. Monard. A study of the behaviour of several methods for balancing machine learning training data. *SIGKDD Explorations*, 6(1):20–29, 2004.

[5] E. Bernadó-Mansilla and J.M. Garrell-Guiu. Accuracy-based learning classifier systems: Models, analysis and applications to classification tasks. *Evolutionary Computation*, 11(3):209–238, 2003.

[6] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.

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

[8] N.V. Chawla, D.A. Cieslak, L.O. Hall, and A. Joshi. Automatically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery*, pages 225–252, 2008.

[9] N.V. Chawla, N. Japkowicz, and A. Kolcz. Editorial: special issue on learning from imbalanced data sets. *SIGKDD Explorations*, 6(1):1–6, 2004.

[10] D.A. Cieslak, T.R Hoens, N.V. Chawla, and W.P. Kegelmeyer. Hellinger distance decision trees are robust and skew-insensitive. *Data Mining Knowl Disc*, 24:136–158, 2012.

[11] C. Cornelis, N. Verbiest, and R. Jensen. Ordered weighted average based fuzzy rough sets. In *Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology*, pages 78–85, 2010.

[12] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27, 1967.

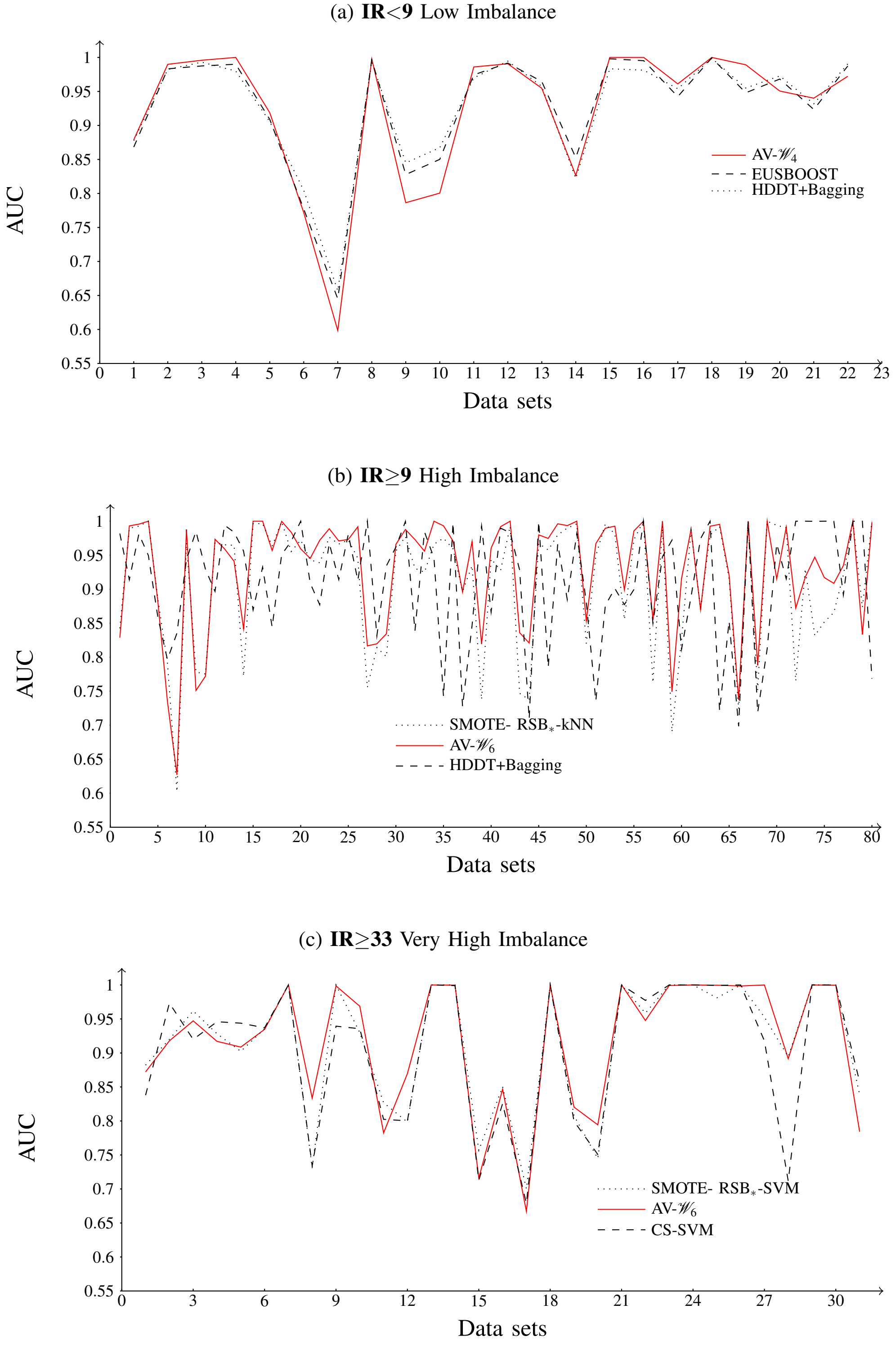
[13] J. Demsâr. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[14] P. Domingos. MetaCost: a general method for making classifiers cost sensitive. In *Proceedings of Fifth International Conference on Knowledge Discovery and Data Mining (KDD99)*, pages 155–164, 1999.

[15] E. Duman, Y. Ekinci, and A. Tanrıverdi. Comparing alternative classifiers for database marketing: The case of imbalanced datasets. *Expert Systems with Applications*, 39(1):48–53, 2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2014.2371472, IEEE Transactions on Fuzzy Systems

Figure 4: AUC for all blocks of data sets. Data sets are ordered according to their IR.



(a) **IR<9** Low Imbalance

(b) **IR≥9** High Imbalance

(c) **IR≥33** Very High Imbalance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2014.2371472, IEEE Transactions on Fuzzy Systems

[16] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

[17] A. Fernández, S. García, M.J. del Jesús, and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.

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

[19] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701, 1937.

[20] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 42 (4):463–484, 2012.

[21] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera. EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition*, 46:3460–3471, 2013.

[22] S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability. *Soft Computing*, 13(10):959–977, 2009.

[23] S. García, A. Fernández, J. Luengo, and F. Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power. *Information Sciences*, 180:2044–2064, 2010.

[24] S. García and F. Herrera. An extension on ”statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2677–2694, 2008.

[25] S. García and F. Herrera. Evolutionary undersampling for classification with imbalanced datasets: proposals and taxonomy. *Evolutionary Computation*, 17:275–306, 2009.

[26] J.W. Grzymala-Busse, J. Stefanowski, and S. Wilk. A comparison of two approaches to data mining from imbalanced data. *Journal of Intelligent Manufacturing*, 16(6):565–573, 2005.

[27] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *Ann. Statist*, 26(2):451–471, 1998.

[28] H. He and E.A. García. Learning from imbalanced data. *IEEE Transactions On Knowledge And Data Engineering*, 21(9):1263–1284, 2009.

[29] S. Holm. A simple sequentially rejective multiple test procedure, scandinavian. *Journal of Statistics*, 6:65–70, 1979.

[30] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):299–310, 2005.

[31] Y.M. Huang, C.M. Hung, and H.C. Jiau. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. *Nonlinear Analysis: Real World Applications*, 7(4):720–747, 2006.

[32] R. Jensen and C. Cornelis. Fuzzy rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42):5871–5884, 2011.

[33] W. Khreich, E. Granger, A. Miri, and R. Sabourin. Iterative boolean combination of classifiers in the ROC space: An application to anomaly detection with HMMs. *Pattern Recognition*, 43:2732–2752, 2010.

[34] Y.H. Lee, P.J.H. Hu, T.H. Cheng, T.C. Huang, and W.Y. Chuang. A preclustering-based ensemble learning technique for acute appendicitis diagnoses. *Artificial Intelligence in Medicine*, 58(2):115–124, 2013.

[35] V. López, A. Fernández, S. García, V. Palade, and F. Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, 2013.

[36] Z. Pawlak. Rough sets. *International journal of Computer and Information Sciences*, 11:145–172, 1982.

[37] J.R Quinlan. C4.5 programs for machine learning. *Morgan Kaufmann, CA*, 1993.

[38] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera. SMOTE-RSB$_*$: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. *International Journal of Knowledge and Information Systems*, 33:245–265, 2012.

[39] R. Rifkin and A. Klautau. In defense of one-vs-all classification. *Machine Learning Research*, 5:101–145, 2004.

[40] Z. Pawlak S.K.M., Wong S.K.M., and W. Ziarko. Rough sets: probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, 29:81–95, 1988.

[41] Y. Sun, A. K. C. Wong, and M. S. Kamel. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(4):687–719, 2009.

[42] K.M. Ting. An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering*, 14 (3):659–665, 2002.

[43] V. Vapnik. *The nature of statistical learning*. Springer, 1995.

[44] N. Verbiest, C. Cornelis, and R. Jensen. Fuzzy rough positive region-based nearest neighbour classification. In *Proceedings of the 20th International Conference on Fuzzy Systems (FUZZ-IEEE2012)*, pages 1961–1967, 2012.

[45] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of the international joint conference on AI*, pages 55–60, 1999.

[46] G. Weiss and F. Provost. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19:315–354, 2003.

[47] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, 18:183–190, 1988.

[48] Q. Yang and X. Wu. 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making*, 5(4):597–604, 2006.

[49] H. Yu, J. Ni, and J. Zhao. ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced dna microarray data. *Neurocomputing*, 101:309–318, 2013.

[50] B. Zadrozny, J. Langford, and N. Abe. Cost-sensitive learning by cost-proportionate example weighting. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 03)*, pages 435–442, 2003.

[51] Z.H. Zhou and X.Y. Liu. On multi-class cost-sensitive learning. *Computational Intelligence*, 26(3):232–257, 2010.

Enislay Ramentol received the M.Sc. degree in Informatics in 2008 from the University of Camagüey, Cuba. In 2010, she received the M.Sc. degree in Soft Computing, and in 2014 the Ph.D. degree in informatics, both from the University of Granada, Spain. She is currently an Assistant Professor in the Department of Computer Science at University of Camagüey, Cuba. Her research interests include data mining, imbalanced learning, instance selection and fuzzy rough set theory.

Sarah Vluymans obtained her M.Sc. in Mathematical Informatics from Ghent University (Belgium) in 2014. Currently, she is working as a Ph.D. student at Ghent University and at the Inflammation Research Centre, part of the Flemish Institute of Biotechnology. Her research is funded by the Special Research Fund (BOF) by Ghent University. Her work focuses on the integration of concepts from fuzzy rough theory in a wide variety machine learning techniques.

Nele Verbiest holds a master's degree in Mathematical Informatics and a Ph.D. in Computer Science, both from Ghent University. Her research interests include classification, evolutionary algorithms, instance selection, feature selection and fuzzy rough set theory. Currently, she is an analyst at Python Predictions, a Brussels-based service provider specialized in predictive analytics in the field of marketing, risk and operations.

Yailé Caballero received the M.Sc. degree in Cybernetic in 2001 and the Ph.D. degree in 2007, both from the Universidad Central ÒMarta AbreuÓ de Las Villas (UCLV), Cuba. She has published more than 90 papers in journals and proceedings of international congresses. Her current research interests include rough set theory, machine learning, metaheuristics and solution of problems of classification and prediction.

Rafael Bello received his Bachelor degree in Mathematics and Computer Science (1982) at Universidad Central de Las Villas (UCLV), Santa Clara, Cuba and his Ph.D. in Mathematics at UCLV in 1988. He is a Full Professor at Computer Science Department, UCLV, Cuba, and has developed academic exchange with universities in Latin America and Europe. He has taught more than 60 undergraduate and graduate courses in those academic centers. He has authored/edited 9 books, and published over 180 papers in conference proceedings and scientific journals. He has a been reviewer for different scientific journals. His research interests include metaheuristics, soft computing (rough and fuzzy set theories), knowledge discovery and decision making. He is a Member of the Cuban Science Academy.

Chris Cornelis holds an M.Sc. and a Ph.D. degree in Computer Science from Ghent University (Belgium). Currently, he is a postdoctoral fellow at the University of Granada supported by the Ramón y Cajal programme of the Spanish government, as well as a guest professor at Ghent University. He has co-supervised 7 Ph.D. theses and has authored over 130 papers in international journals, edited volumes and conference proceedings. He serves as an executive board member of the International Rough Set Society (IRSS). His current research interests include fuzzy sets, rough sets and machine learning.

Francisco Herrera received his M.Sc. in Mathematics in 1988 and Ph.D. in Mathematics in 1991, both from the University of Granada, Spain. He is currently a full professor at the Department of Computer Science and Artificial Intelligence at the University of Granada. He has been the supervisor of 36 Ph.D. students, and has published more than 290 papers in international journals. He currently acts as Editor in Chief of the international journals ”Information Fusion” (Elsevier) and ÒProgress in Artificial Intelligence (Springer). He acts as editorial board member of a dozen of journals, among others: International Journal of Computational Intelligence Systems, IEEE Transactions on Fuzzy Systems, Information Sciences, Knowledge and Information Systems, Fuzzy Sets and Systems, Applied Intelligence, Knowledge-Based Systems, Memetic Computation, and Swarm and Evolutionary Computation. He received the following honors and awards: ECCAI Fellow 2009, IFSA Fellow 2013, 2010 Spanish National Award on Computer Science ARITMEL to the ”Spanish Engineer on Computer Science”, International Cajastur ”Mamdani” Prize for Soft Computing (Fourth Edition, 2010), IEEE Transactions on Fuzzy Systems Outstanding 2008 and 2012 Paper Award (bestowed in 2011 and 2015 respectively), 2011 Lotfi A. Zadeh Prize Best paper Award of the International Fuzzy Systems Association, 2013 AEPIA Award to a scientific career in Artificial Intelligence (September 2013). His current research interests include bibliometrics, computing with words in decision making, information fusion, evolutionary algorithms, evolutionary fuzzy systems, biometrics, data preprocessing, data mining, cloud computing and big data.