Enhancing multi-class classification in FARC-HD fuzzy classifier: On the synergy between n-dimensional overlap functions and decomposition strategies

Mikel Elkano, Mikel Galar, Jose Sanz, Alberto Fernández, Edurne Barrenechea, Francisco Herrera, *Member, IEEE*, and Humberto Bustine, *Member, IEEE*

Abstract—There are many real-world classification problems involving multiple classes, e.g., in bioinformatics, computer vision or medicine. These problems are generally more difficult than their binary counterparts. In this scenario, decomposition strategies usually improve the performance of classifiers. Hence, in this paper we aim to improve the behaviour of FARC-HD fuzzy classifier in multi-class classification problems using decomposition strategies, and more specifically *One-vs-One* (OVO) and *One-vs-All* (OVA) strategies. However, when these strategies are applied on FARC-HD a problem emerges due to the low confidence values provided by the fuzzy reasoning method. This undesirable condition comes from the application of the product t-norm when computing the matching and association degrees, obtaining low values, which are also dependent on the number of antecedents of the fuzzy rules. As a result, robust aggregation strategies in OVO such as the weighted voting obtain poor results with this fuzzy classifier.

In order to solve these problems, we propose to adapt the inference system of FARC-HD replacing the product t-norm with *overlap functions*. To do so, we define *n-dimensional overlap functions*. The usage of these new functions allows one to obtain more adequate outputs from the base classifiers for the subsequent aggregation in OVO and OVA schemes. Furthermore, we propose a new aggregation strategy for OVO to deal with the problem of the weighted voting derived from the inappropriate confidences provided by FARC-HD for this aggregation method.

The quality of our new approach is analyzed using twenty datasets and the conclusions are supported by a proper statistical analysis. In order to check the usefulness of our proposal, we carry out a comparison against some of the state-of-the-art fuzzy classifiers. Experimental results show the competitiveness of our method.

Index Terms—Multi-classification, one-vs-one, fuzzy rule-based classification systems, aggregations, overlaps.

Mikel Elkano, Mikel Galar, Jose Sanz, Edurne Barrenechea, and Humberto Bustine are with the Departamento de Automática y Computación, Universidad Pública de Navarra, Navarra, 31006 Spain (e-mail: elkano.mikel@gmail.com, {mikel.galar, joseantonio.sanz, edurne.barrenechea, bustince}@unavarra.es)

Alberto Fernández is with the Department of Computer Science, University of Jaén, Jaén, 23071 Spain (e-mail: alberto.fernandez@ujaen.es).

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

I. INTRODUCTION

FUZZY Rule-Based Classification Systems (FRBCSs) are well-known and widely used tools in the field of pattern recognition and classification problems. They provide an interpretable model by using linguistic labels in the antecedents of their rules [1]. FRBCSs have been applied in multiple real-world problems, including domotics [2], anomaly intrusion detection [3], image processing [4], and medical problems [5], among others.

In classification, two types of problems can be differentiated depending on the number of classes that compose the output of the problem: binary (two classes) and multi-class problems (more than two classes). Usually, it is more difficult to build a classifier in the latter case due to the overlapping among the examples of the different classes of the problem, which makes the definition of decision boundaries more complex. Even so, multi-class problems are present in several applications domains such as fingerprints recognition [6], handwritten digits [7], microarrays [8] or cardiovascular disease classification [5]. A commonly used solution to deal with multi-class classification problems is to use decomposition techniques [9], [10], which try to divide the original multi-class problem into easier to solve binary classification problems, which are faced by independent binary classifiers named *base classifiers*.

Different decomposition strategies have been proposed in the specialized literature [10]. Two of the most well-known and used ones are *One-vs-One* (OVO) and *One-vs-All* (OVA) [9], which can be included within the wider Error Correcting Output Codes (ECOC) framework [11]. OVO scheme divides the original problem into as many sub-problems as possible pairs of classes, whereas in OVA the division results in as many sub-problems as classes in the original one. In both strategies each binary problem is addressed by an independent base classifier. When classifying a new example, the outputs of all the base classifiers are combined to make the final decision (aggregation phase).

In this paper we aim to improve the performance of FARC-HD (*Fuzzy Association Rule-based Classification model for High-Dimensional problems*) [12] in multi-class problems using decomposition strategies. Previous works have shown that although the base classifier can be capable of solving multi-class problems, such as FARC-HD, these strategies usually

work better than addressing the problem directly [9], [13]–[16]. However, two main difficulties emerge when carrying out the proposed hybridization:

1) The aggregation of the classifiers in OVA and OVO schemes directly depends on the confidences provided by the base classifiers. In the case of FARC-HD, due to the usage of *additive combination* [17] as fuzzy reasoning method, we consider as confidence the sum of the association degrees obtained for each class, which are computed by multiplying the matching degrees (of the example with the antecedents of the rules using the product t-norm to model the conjunction) and the rule weight. For this reason, when combining FARC-HD and decomposition strategies the confidences obtained when carrying out the inference process of FARC-HD are not suitable for the subsequent aggregation. From our point of view, this is due to the usage of the product in the inference of FARC-HD, which produces small confidences with low variations for each pair of classes and penalizes the rules with the largest number of antecedents.

2) Aggregation strategies that usually have a robust and accurate performance in OVO, such as the weighted voting (WV) [18], [9], [19], do not obtain good results when using FARC-HD as base classifier (while others do not present this problem). Our hypothesis is that the confidence estimation of the non-predicted class provided by FARC-HD distorts the combination in OVO when using this aggregation strategy.

In order to address the former problem, we propose to adapt the inference process of FARC-HD in such a way that the confidences obtained allow decomposition strategies to produce more accurate aggregations and consequently, can lead to improve the classification in OVO and OVA models. To do so, we will use *overlap functions* [20], which satisfy similar properties to those of the product, in the inference of FARC-HD. These functions allow us to obtain values with a higher variation than those provided by the product, in such a way that the confidences used in OVO and OVA (stored in the score-matrix or vector) are better modeled, and hence greater knowledge is acquired for the posterior aggregation.

Since overlaps functions are originally defined for 2-dimensional problems, in this paper we introduce the concept of n-dimensional overlap functions to be able to compute the overlap among n input values. More specifically, according to the problem related to the number of antecedents, we propose the usage of n-dimensional overlap functions that do not decrease the results as the number of input values (n) increases.

In addition, aiming at facing the latter problem, we propose an alternative to the usage of the WV in the aggregation phase of OVO strategy. To do so, we propose an aggregation strategy named WinWV, which follows the idea of the WV, in which we do not consider the confidences obtained by non-predicted classes, since its usage is not appropriate for the classification in the case of FARC-HD (we will show that OVA and other OVO aggregations are not affected by this problem).

In order to assess the quality of the methods, we use twenty numerical datasets from the KEEL dataset repository [21] and we contrast the results obtained using non-parametric statistical tests, as suggested in the specialized literature [22], [23]. In these experiments, we will study the goodness of the usage n-dimensional overlap functions and we will also analyze whether the usage of WinWV allows the performance of the WV to be enhanced. Moreover, we will show the validity of our proposal to improve the performance of FARC-HD in multi-class problems, comparing it against the original FARC-HD algorithm and some of the best performing fuzzy methods, i.e., FURIA algorithm [24], IVTURS$_{\text{FARC-HD}}$ [25], and PTTD [26].

The rest of this paper is organized as follows. In Section II, we briefly introduce FARC-HD and decomposition strategies and we describe some of the aggregations for OVO that we use in this paper. Section III contains a detailed description of our proposals to use FARC-HD with decomposition strategies and puts forward the definition of the n-dimensional overlap functions. The set-up of the experimental framework is given in Section IV and the analysis of the results obtained is presented in Section V. Finally, Section VI concludes this paper.

II. PRELIMINARIES

In this section, we first recall some concepts about FRBCSs and we briefly explain the FARC-HD algorithm [12] (Section II-A). Then we describe OVO and OVA decomposition strategies and some of the OVO aggregation methods studied in the literature (Section II-B). Finally, we review the related works in Section II-C.

A. Fuzzy Rule-Based Classification Systems and FARC-HD

A classification problem consists in learning a mapping function called *classifier* from a set of training examples, named *training set*, that allows one to classify previously unknown examples. Let $x_p = (x_{p1}, \ldots, x_{pn})$ be the pth example of the training set which is composed of P examples, where x_{pi} is the value of the i-th attribute ($i = 1, 2, \ldots, n$) of the p-th training example. Each example belongs to a class $y_p \in \mathbb{C} = \{C_1, C_2, ..., C_m\}$, where m is the number of classes of the problem.

We find multiple techniques used to cope with classification problems. Among them, FRBCSs are widely used because they provide an interpretable model by means of the use of linguistic labels in their rules [1].

The two main components of FRBCSs are the following ones:

1) *Knowledge base*: It is composed of both the rule base (RB) and the database, where the rules and the membership functions are stored, respectively.

2) *Fuzzy reasoning method*: This is the mechanism that classifies examples using the information stored in the knowledge base.

In this work we focus on a fuzzy rule learning algorithm known as FARC-HD [12], since it is currently one of the

most accurate and interpretable FRBCSs in the literature. This algorithm makes use of the following rule structure:

$$\text{Rule } R_j : \text{If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn}$$
$$\text{then Class} = C_j \text{ with } RW_j \quad (1)$$

where R_j is the label of the j-th rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector that represents the example, A_{ji} is a fuzzy set, $C_j \in \mathbb{C}$ is the class label and RW_j is the rule weight, which is computed using the most common specification, i.e., the fuzzy confidence value or certainty factor defined in [27]:

$$RW_j = CF_j = \frac{\sum\limits_{x_p \in ClassC_j} \mu_{A_j}(x_p)}{\sum\limits_{p=1}^{P} \mu_{A_j}(x_p)} \quad (2)$$

where $\mu_{A_j}(x_p)$ is the matching degree of the example x_p with the antecedent part of the fuzzy rule R_j, which is computed using the Eq. (3), shown further in this subsection. In the case of FARC-HD, linguistic labels are modeled using uniformly distributed triangular membership functions, which form a strong partition (Fig. (1)).
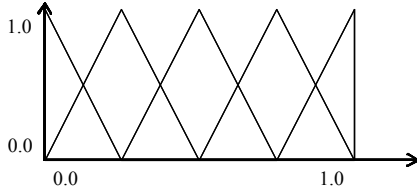

Fig. 1. Linguistic labels partitioning in FARC-HD.

In order to generate the rule base, FARC-HD applies a learning process composed of three steps:

1) *Fuzzy association rule extraction for classification*: With the aim of obtaining the fuzzy rule base, a search tree [28] is constructed for each class. To do so, the frequent itemsets (an item is a linguistic label) are computed using the support and confidence. Finally, the fuzzy rules are generated from the obtained frequent itemsets. The number of linguistic terms in the antecedents of the rules is limited by the maximum depth of the tree.
2) *Candidate rule prescreening*: This phase makes use of subgroup discovery to preselect the most interesting rules from the rule base obtained in the previous stage by means of a pattern weighting scheme [29]. The weights of the examples are based on the coverage of the fuzzy rules.
3) *Genetic rule selection and lateral tuning*: An evolutionary algorithm is used both to perform a lateral tuning of the fuzzy sets [30] and to select the most accurate rules from the rule base generated in the previous steps.

Let $x_p = (x_{p1}, ..., x_{pn})$ be a new example to be classified, FARC-HD applies a fuzzy reasoning method called *additive combination* [17], computed in four steps.

1) *Matching degree*. In this step the strength of activation of the *if*-part for all rules in the RB with the pattern x_p

is computed.

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_j n_j}(x_{pn_j})) \quad (3)$$

where $\mu_{A_{ji}}(x_{pi})$ is the matching degree of the example with the i-th antecedent of the rule R_j, T is a t-norm (in the case of FARC-HD the product), and n_j is the number of antecedents of the rule.

2) *Association degree*. The association degree of the pattern x_p with each rule in the RB is computed.

$$b_j(x_p) = \mu_{A_j}(x_p) \cdot RW_j \quad (4)$$

3) *Confidence degree*. In this stage the confidence degree for each class is computed. To obtain the confidence degree of a class, the association degrees of the rules of that class, i.e., those whose consequent class is the class we are considering, are summed.

$$conf_l(x_p) = \sum_{R_j \in RB;\ C_j = l} b_j(x_p), \qquad l = 1, 2, \ldots, m \quad (5)$$

4) *Classification*. The class that obtain the highest confidence degree is the predicted one.

$$Class = arg \max_{l=1,\ldots,m} (conf_l(x_p)) \quad (6)$$

As we can observe in the rule structure and the fuzzy reasoning method, FARC-HD is capable of solving multi-class classification problems directly. However, previous works have shown that decomposition strategies usually work better than addressing the problem directly [9], [10], [13]–[16]. Therefore, we propose to use the decomposition strategies with the aim of improving the performance of FARC-HD when facing multi-class problems. However, as we have stated in the introduction, we will show that there are some issues we need to address when combining FARC-HD and these strategies.

B. Decomposition strategies

Decomposition strategies [10] divide the original multi-class problem into simpler binary problems that are faced by independent binary classifiers, which are referred as *base classifiers*. These strategies can be used both with classifiers that are able to face only two-class classification problems and those having an inherent multi-class support. In this paper we consider two of the most used decomposition strategies in the literature: *One-Versus-One* (OVO) and *One-Versus-All* (OVA) [9] strategies.

1) One-Versus-One (OVO): OVO decomposition divides a problem of m classes into $m(m-1)/2$ binary sub-problems (all possible pairs of classes). Each problem is faced by a binary classifier, which is responsible for distinguishing a pair of classes $\{C_i, C_j\}$. When classifying a new example, each base classifier will return a pair of confidence degrees $r_{ij}, r_{ji} \in [0,1]$ in favour of classes C_i, C_j, respectively (if only a confidence degree is given for the predicted class, the other is usually computed as $r_{ji} = 1 - r_{ij}$, being C_i the predicted one). The outputs (confidence degrees) provided by all the base classifiers are stored in the score-matrix R as

follows:

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix} \quad (7)$$

Since each binary sub-problem is faced by an independent classifier, we normalize the score-matrix such that the range of the confidences provided by all classifiers is the same. This normalization is very important when using FARC-HD because it does not return confidences in [0,1] that can be interpreted as probabilities. The score-matrix is normalized as follows.

$$r_{ij} = \begin{cases} \dfrac{r_{ij}}{r_{ij} + r_{ji}} & \text{if } r_{ij} \neq 0 \text{ or } r_{ji} \neq 0 \\ 0.5 & \text{if } r_{ij} = r_{ji} = 0 \end{cases} \quad (8)$$

Finally, the outputs of the base classifiers are aggregated and the predicted class is obtained. This aggregation step is a key factor for the classification process [9]. In this paper, we consider four well-known OVO aggregation methods.

- *Voting strategy (VOTE) [31].* Each base classifier votes for the predicted class and the class having the largest number of votes is given as output:

$$Class = arg \max_{i=1,\ldots,m} \sum_{1 \leq j \neq i \leq m} s_{ij} \quad (9)$$

  where s_{ij} is 1 if $r_{ij} > r_{ji}$ and 0 otherwise.
- *Weighted Voting (WV) [18].* Each base classifier votes for both classes based on the confidences obtained for them. The class having the largest value is given as output:

$$Class = arg \max_{i=1,\ldots,m} \sum_{1 \leq j \neq i \leq m} r_{ij} \quad (10)$$

- *Non-Dominance Criteria (ND) [32].* The score-matrix is considered as a fuzzy preference relation. Then class with the highest non-dominance degree is the predicted one:

$$Class = arg \max_{i=1,\ldots,m} \left\{ 1 - \max_{j=1,\ldots,m} r'_{ji} \right\} \quad (11)$$

  where R' is the strict score-matrix.
- *Learning valued preference for classification (LVPC) [33], [34].* This aggregation strategy, as in ND, considers the score-matrix as a fuzzy preference relation. In this manner, the original relation is decomposed into three new relations with different meanings: the strict preference, the conflict, and the ignorance. In order to obtain the output class, a decision rule based on voting strategy is proposed:

$$Class = arg \max_{i=1,\ldots,m} \sum_{1 \leq j \neq i \leq m} P_{ij} + \frac{1}{2} C_{ij} + \frac{N_i}{N_i + N_j} I_{ij} \quad (12)$$

  where N_i is the number of examples from the class i in the training data, C_{ij} is the degree of conflict (the degree to which both classes are supported), I_{ij} is the degree of ignorance (the degree to which none of the classes are supported), and P_{ij} and P_{ji} are the strict preference for i and j, respectively. These variables are computed as

follows:

$$C_{ij} = \min \{r_{ij}, r_{ji}\}$$
$$P_{ij} = r_{ij} - C_{ij}$$
$$P_{ji} = r_{ji} - C_{ij}$$
$$I_{ij} = 1 - \max \{r_{ij}, r_{ji}\}$$

Owing to the way in which the multi-class problem is divided in OVO scheme, there is an issue inherent to this decomposition method: the non-competent classifiers [19]. The learning process of each base classifier is performed using only the examples belonging to the two classes that this classifiers will classify and consequently, it ignores the examples belonging to other classes. Therefore, the remainder classes are unknown for these classifiers and their outputs are irrelevant when classifying examples of those classes. However, these outputs are aggregated in the same way as the relevant ones, possibly misleading the correct labeling of the example. Although this is an interesting line, it is out of the scope of this paper and we leave it as a future research line.

2) One-Versus-All (OVA): OVA decomposition divides a problem of m classes into m binary problems, which are addressed by independent binary classifiers. Each base classifier distinguishes one of the classes from all other classes. The learning process of these classifiers is performed using the whole training data, considering the examples from the single class as positives and the rest of examples as negatives. When classifying a new example, each base classifier will return a confidence degree $r_i \in [0, 1]$ in favour of the class C_i, which will be stored in a *score-vector* R:

$$R = (r_1, \ldots, r_i, \ldots, r_m) \quad (13)$$

As in OVO, we need to normalize the score-vector such that the range of the confidences provided by all classifiers is the same. In order to do so, we also need the score-vector in which the confidences obtained by each classifier for the negative class are stored ($\hat{R}$). With both vectors, the normalization of the score-vector R is performed as follows.

$$r_i = \frac{r_i}{r_i + \hat{r}_i} \quad (14)$$

Finally, the most commonly used aggregation in OVA consider the usage of the maximum value in the score-vector, and thus the class with the highest confidence will be predicted.

C. Related works

Decomposition strategies can be included in the broader categories of ensembles and multiple classifier systems (MCSs) [35], [36]. These types of systems aim to improve the classification performance by the combination of several classifiers. In fact, ensembles and MCSs are usually referred to those methods where the base classifiers are able to predict any of the classes of the problem; however, decomposition techniques are also formed of sets of classifiers, but there is a major difference: each base classifier is not able to predict all the classes in the problem (only two of them or two combinations of different classes are predicted). Such an important differ-

ence has produced rather different approaches for each type of method.

Traditional ensembles aim to produce diverse base classifiers [37], whose differences in their predictions allow them to increase the performance, since they complement each other. Widely used ensemble methods are Bagging [38] and Boosting [39], [40]. These type of methods have been also considered in the fuzzy community, where different approaches have been proposed using fuzzy systems as base classifiers [41]–[44]. In [41], the authors combined the FRBCSs obtained in the pareto front of a multi-objective optimization genetic algorithm. An extension of the classical Random Forests (a variant of bagging) using fuzzy decision trees was presented in [42], [45]. Boosting in the fuzzy context was applied in [46]. The authors of [43], [44], [47] developed a methodology to construct FURIA-based multi-classifiers in a series of works, including all the different phases of an ensemble, from its construction (also bagging-based) to the combination procedure presented in the latter work. All these type of models are usually no longer interpretable, and hence the fuzzy classifiers are used instead of other weak classifiers such as the commonly used decision trees to take advantage of the fuzzy decision boundaries to reach highly accurate models, which may need the usage of thousands of rules [44], even though some authors focused on reducing this number [48]. Moreover, FURIA [24] has been one of the most extended base classifier in this framework, which by itself is not as interpretable as classical FRBCSs [1], since it make use of hyper-rectangles adjusted for each rule instead of using the same linguistic labels in each rule. For this reason, in this paper we only deal with decomposition-based ensembles, which may maintain part of the interpretability of the original models.

Otherwise, decomposition strategies have also attracted attention as a way of improving classification in multi-class problems with FRBCSs [26], [32], [34], [49]. In this framework, different base classifiers have been used (Fuzzy Ripper [34], FH-GBML [50] or SLAVE [51]), as well as different combination methods have been proposed such as the Non-Dominance criterion (ND) in [32] or the Learning Valued Preference for Classification (LVPC) [34], [49] already described in Section II-B. In these papers, the authors considered the score-matrix as a preference relation from which the best alternative should be predicted. In order to do so, the conflict and ignorance were modeled in [49] and thereafter applied in the Fuzzy Ripper algorithm presented in [34]. Similarly but with a different approach to output the class from the score-matrix, the authors in [32] proposed the usage of the ND criterion, showing good results with FH-GBML and SLAVE classifiers. In addition, in [26] the authors presented the Top-Down induction of Fuzzy Pattern Trees (PTTD), which made use of OVA approach.

However, recent developments with fuzzy classifiers are not only related to ensemble strategies. Taking into account the good properties of FRBCSs, several approaches [24], [25], [52]–[54] have been proposed aimed at improving the trade-off between accuracy and interpretability [55]. FURIA [24] extended the RIPPER algorithm using fuzzy rules and it provides accurate results. In [52], authors combined a feature selection process using the so-called *modulator functions* and a fuzzy rule extraction mechanism based on fuzzy clustering. Castro et. al [53] defined a fuzzy classifier using general fuzzy rules and a new mechanism aimed at solving the conflicts among them. In [54], authors defined a new approach to design fuzzy classifiers using *k*-means clustering and a memetic algorithm to find the optimal values of fuzzy rules and membership functions. In order to improve the interpretability of TSK fuzzy classifiers, the usage of a minimax probability was proposed in [56]. Finally, Sanz et. al [25] provided a framework to improve the performance of FRBCSs using interval-valued fuzzy sets.

On account of the different fuzzy methodologies described above, we have considered to include those following a similar philosophy to our proposal in the experimental study. More specifically, we have considered both aggregations defined in the fuzzy context (ND and LVPC) [32], [34], [49], the PTTD [26] method as an OVA based fuzzy system as well as FURIA [24] and IVTURS [25] as state-of-the-art fuzzy classifiers. Finally, we should mention that none of the ensemble/multi-classifier approaches previously enumerated has addressed the problem affecting the inference that we aim to overcome in this paper.

III. Introducing n-dimensional overlap functions to adapt FARC-HD behaviour in OVA and OVO

In this paper, we propose to combine FARC-HD with OVA and OVO decomposition strategies in order to improve the performance of FARC-HD in multi-class classification problems. However, the confidences provided by FARC-HD are not adequate for them due to the use of the product to compute the association degree, as we will show in the experimental analysis. Thus, the inference process needs to be adapted for the sake of a better synergy between FARC-HD and decomposition schemes.

In the remainder of this section, we first describe the way in which we introduce FARC-HD in OVO and OVA models, as well as the problems that we have to address when carrying out this combination (Section III-A). Next, we recall the concept of two-dimensional overlap functions and we present the new definition of n-dimensional overlap functions, as well as their construction method (Section III-B). Then, we describe the modification of the inference of FARC-HD using n-dimensional overlap functions aiming at improving the synergy between FARC-HD and decomposition strategies (Section III-C). Finally, we present a new aggregation strategy for the OVO model named WinWV that solves the problems of the WV with the confidences of FARC-HD (Section III-D).

A. Using FARC-HD as base classifier in the OVA and OVO strategies

In order to use OVO and OVA strategies with FARC-HD, we need to fill the score-matrix of OVO (Eq. (7)) and the score-vector of OVA (Eq. (13)) with the confidences provided by FARC-HD for each class. More specifically, we consider as confidences the confidence degree for each class computed using Eq. (5). Both the matching and the association degrees

of the example with the fuzzy rules are computed by Eq. (3) and (4), respectively, using the product t-norm.

When low values are aggregated using the product t-norm, the range in which the result can vary is small (the lower the input values are, the smaller the range becomes), which may happen when computing the matching degree of several fuzzy rules. This effect is further accentuated as the number of antecedents of the rules increases, which implies that the association degrees of those rules with more antecedents will be smaller and will have a lower variation. This behaviour implies that the confidences stored in the score-matrix and score-vector will have low variations as well, which is not desirable for the subsequent aggregation performed in OVO and OVA schemes. Consequently, it seems suitable to modify the inference process in such a way that the aggregation of the values involved in the computation of the association degrees is made using functions whose results are in a wider range, maintaining more information for the aggregation process (Section III-C).

The previous problem does not affect the behaviour of the original FARC-HD, since the confidences obtained after the inference process are not used beyond classification and hence, this variation does not affect the final result given by the algorithm. However, in decomposition strategies, the confidences provided by FARC-HD are used in the aggregation phase, and hence the predictions are used beyond the decision of the class prediction of the base classifiers. Thus, a low variation in the confidences might have a negative effect in OVO and OVA models, which is especially reflected in the unexpected behaviour of robust aggregation strategies used in OVO, such as the WV, as we will show in the experimental study. For this reason, a new aggregation strategy that solves the problems of the WV is needed (see Section III-D).

B. n-dimensional overlap functions

The concept of overlap function [20] was introduced in image processing in order to classify those pixels that it was not clear whether they belonged to the object or to the background. This concept has been applied in many image processing problems [57]–[59] and used in [60] to model the indifference in preference relations. However, the application range of these functions has turned out to be much wider, since they allow one to recover many of the characteristics of the t-norms without imposing the associativity. Precisely because the associativity is not demanded, the extension of the concept of overlap function to dimensions higher than two is not direct. Moreover, this extension is necessary in order to use overlap functions in problems in which the associativity is not necessary or even natural and in which t-norms have been used. In this paper, we propose a definition of overlap function in any finite dimension, which particularly allows one to recover the two-dimensional case. Additionally, we present the construction method of overlap functions using rational expressions.

We first recall the following definition of two-dimensional overlap functions:

Definition 1: [20] A function $O : [0,1] \times [0,1] \to [0,1]$ is an overlap function if satisfies the following conditions :

1) $O(x,y) = O(y,x)$ *for all* $x, y \in [0,1]$.
2) $O(x,y) = 0$ *if and only if* $x \cdot y = 0$.
3) $O(x,y) = 1$ *if and only if* $x \cdot y = 1$.
4) is increasing.
5) is continuous.

Following this concept we define the extension of the previous two-dimensional overlap functions to n dimensions:

Definition 2: An n-dimensional function $O^n : [0,1]^n \to [0,1]$ with $n \geq 2$ is a n-dimensional overlap function if the following properties hold:

1) O^n is symmetric.
2) $O^n(x_1, \ldots, x_n) = 0$ if and only if $\prod_{i=1}^{n} x_i = 0$.
3) $O^n(x_1, \ldots, x_n) = 1$ if and only if $\prod_{i=1}^{n} x_i = 1$.
4) O^n is increasing.
5) O^n is continuous in each of the variables.

Example 1: The following functions are examples of n-dimensional overlap functions:

1) The minimum is a n-dimensional overlap function which is also a t-norm.

$$O^n(x_1 \ldots, x_n) = \min(x_1, \ldots, x_n) \qquad (15)$$

2) Take $p > 0$. Then the function

$$O^n(x_1, \ldots, x_n) = \left(\prod_{i=1}^{n} x_i \right)^p \qquad (16)$$

is a n-dimensional overlap function. Furthermore, O^n is associative if and only if $p = 1$.

a) If $p = 1$ we recover the product, which is a t-norm as well.

$$O^n(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i \qquad (17)$$

b) If $p = \frac{1}{n}$ we have the geometric mean:

$$O^n(x_1, x_2, \ldots, x_n) = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad (18)$$

3) The harmonic mean is a n-dimensional function:

$$O^n(x_1, x_2, \ldots, x_n)$$
$$= \begin{cases} \dfrac{n}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}} & \text{if } x_i \neq 0, \quad \text{for all } i = 1, \ldots, n \\ \\ 0 & \text{otherwise.} \end{cases}$$
$$(19)$$

4) The function

$$O^n(x_1, \ldots, x_n) = \sin\left(\frac{\pi}{2} \left(\prod_{i=1}^{n} x_i \right)^\alpha \right) \qquad (20)$$

where $\alpha \leq \dfrac{1}{2n}$, is another example of n-dimensional overlap function.

As we have shown, both the product, which is used in the original FARC-HD, and the minimum, which is another t-

norm that is commonly used in FRBCs, are examples of n-dimensional overlap functions. Finally, we present a construction method for n-dimensional overlap functions using rational expressions.

Theorem 1: The mapping $O^n : [0,1]^n \rightarrow [0,1]$ is a n-dimensional overlap function if and only if there exist $f, g : [0,1]^n \rightarrow [0,1]$ with

$$O^n(x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n)}$$

where

1) f and g are symmetric.
2) f is non-decreasing and g is non-increasing.
3) $f(x_1, \ldots, x_n) = 0$ if and only if $\prod_{i=1}^{n} x_i = 0$.
4) $g(x_1, \ldots, x_n) = 0$ if and only if $\prod_{i=1}^{n} x_i = 1$.
5) f and g are continuous.

Proof. To see the necessity, assume that O^n is a n-dimensional overlap function. We can define $f(x_1, \ldots, x_n) = O^n(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n) = 1 - f(x_1, \ldots, x_n)$. Hence, the properties (1)-(5) of the Theorem are direct and also:

$$\frac{f(x_1, \ldots, x_n)}{f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n)} = \frac{O^n(x_1, \ldots, x_n)}{1}$$

Let us take a look at what happens with the sufficiency. We must see that the function defined in Eq. (1) is in fact a n-dimensional overlap function. The continuity, the symmetry and the monotony are evident. Also:

$$O^n(x_1, \ldots, x_n) = 0 \text{ iff } f(x_1, \ldots, x_n) = 0 \text{ iff } \prod_{i=1}^{n} x_i = 0$$

and $O^n(x_1, \ldots, x_n) = 1$ iff $f(x_1, \ldots, x_n)$

$$= f(x_1 \ldots, x_n) + g(x_1, \ldots, x_n)$$

that is, if and only if $g(x_1, \ldots, x_n) = 0$ and if and only if $\prod_{i=1}^{n} x_i = 1$.

Example 2: The function

$$O^n(x_1, \ldots, x_n) = \frac{\left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}}{\left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} + \max_{1 \leq i \leq n} (1 - x_i)}$$

is an example of n-dimensional overlap function.

C. Modification of the inference process using n-dimesional overlap functions

Once the n-dimensional overlap functions and the construction methods have been presented, we show the proposed modification of the inference process of FARC-HD in order to improve the aggregation in OVO and OVA strategies. More specifically, we propose to compute the matching degree and the association degree of the example with the fuzzy rules using the previously defined n-dimensional overlap functions. To do so, we replace the t-norm in the matching degree computation (Eq. (3)) by an overlap function (Eq. (21)):

$$\mu_{A_j}(x_p) = O^n(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn_j}}(x_{pn_j})) \quad (21)$$

We must stress that the matching degree is employed to compute both the support and the confidence used in the first step of the learning algorithm shown in Section II-A. Thus, this modification also affects the learning process of the algorithm as well.

Similarly, we substitute the product in the association degree computation (Eq. (4)) by an overlap function, according to Eq. (22).

$$\begin{aligned} b_j(x_p) &= O(\mu_{A_j}(x_p), RW_j) \\ &= O(O^n(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn_j}}(x_{pn_j})), RW_j) \end{aligned} \quad (22)$$

The reason for computing the association degree using an overlap function instead of the product is the same as in the case of the matching degree, that is, the low variation of the association degrees obtained when multiplying the matching degree and the rule weight.

In this paper, we have considered five different overlap functions to observe their effect in the rule base and their influence on the accuracy of the model (for the sake of brevity, in the experiments we take the overlap O in Eq. (22) the same as O^n). Each overlap function returns lower or higher values than the rest for the same input tuple. According to the values returned by the overlap functions, we can establish an order among them. Thus, we will consider that an overlap function is greater than other one if the values returned by the first function are higher than those returned by the second one for the same arguments. A short description of each function is shown below, sorted in ascending order by the returned value:

- *Product (PROD):* The returned value is the product of input values (Eq. (17)). Indeed, this is the case of the original FARC-HD, and hence we are able to recover the original method using the proposed extension of overlap functions.
- *Minimum (MIN):* Returns the minimum of input values (Eq. (15)). This is a t-norm as well, but unlike the product, the returned value does not decrease when the number of arguments increases. The minimum is commonly used in FRBCs.
- *Harmonic Mean (HM):* The returned value is the harmonic mean of input values if all of them are different than zero and 0 otherwise (Eq. (19)).
- *Geometric Mean (GM):* Returns the geometric mean of input values (Eq. (18)).
- *Sine (SIN):* This an example of an overlap function that returns higher values than means (Eq. (20)). The use of this type of functions is interesting in order to check what happens in these cases. In the experiments carried out in Section V, we take $\alpha = \frac{1}{2n}$.

Among the considered overlap functions, the first one is the product (used in the original FARC-HD). The product is a t-norm that returns values with a lower variation than the

other functions and whose returned value decreases as the number of arguments increases. Next, we have the minimum, which is a t-norm as well, but whose returned value does not depend on the number of arguments. Then, we consider the harmonic and the geometric means as representatives of means that return higher values than t-norms [61]. Finally, we have considered a function (SIN) that returns higher values than means. This variety of overlap functions allows us to have a general overview in the experiments (Section V).

According to Eq. (19) and (18), both the harmonic and the geometric means return 0 when one of the arguments is 0. This property is satisfied by t-norms as well, and it is very important to conserve the necessary discrimination capability of FARC-HD. In the experimental study, we will show that another desirable property that the overlaps should satisfy in order to work well in our framework is the idempotence, which is satisfied if:

$$O^n(x, \dots, x) = x \qquad (23)$$

Among the previous overlaps, the minimum, the harmonic mean and the geometric mean satisfy this property.

In Fig. (2a) and (2b) we can graphically observe the previously mentioned differences in the values returned by each overlap function (we depict overlaps with $n = 2$ to ease the visualization of their behavior). Fig. (2a) depicts the behavior of the overlap functions when aggregating a value with the value 1, whereas Fig. (2b) shows the returned values when aggregating a value with itself. Looking at Fig. (2a) and (2b), we can see that the proposed n-dimensional overlap functions provide a higher variation than the product when the input arguments are small. However, both figures show that there is a huge difference between the SIN and the rest of the overlap functions. In fact, in Fig. (2b) we can observe that the returned value is greater than the input arguments when aggregating a value with itself, which might not be a desirable behavior in this framework, as it may produce a loss of the discrimination capabilities of FARC-HD.

D. Adapting the Weighted Voting to FARC-HD confidence estimation: WinWV

In addition to the adaptation of the inference process of the base classifiers to OVO and OVA using overlap functions, we propose a new aggregation method for the OVO strategy named WinWV, which is a modification of the WV that does not achieve the expected results.

As we described in Section II-B, each base classifier provides a pair of confidence degrees $r_{ij}, r_{ji} \in [0, 1]$ in favor of classes C_i, C_j, so that r_{ij} is the confidence predicting the class C_i and r_{ji} is the confidence predicting the class C_j. Although we have improved the confidences provided by FARC-HD making them more suitable for the OVO strategy, we will show that the WV is still not working as expected. From our point of view, the reason is that the confidence estimation of the non-predicted class distorts the aggregation phase in OVO, as we will show in Section V.

Likewise, the LVPC strategy does not work properly with this type of confidences, as it can be observed in [32]. In this case, the confidence for the non-predicted class does not allow one to model the conflict and ignorance degrees properly. Notice that if these terms were not considered the original WV would be recovered. For this reason, we focus on solving the problems of the WV with the confidence estimation of the non-predicted class.

To do so, we propose to consider only the confidence of the predicted class, whereas that of the non-predicted class is not taken into account. Therefore, the WinWV aggregation strategy works as follows:

$$Class = arg \max_{i=1,\dots,m} \sum_{1 \leq j \neq i \leq m} s_{ij} \qquad (24)$$

where s_{ij} is r_{ij} if $r_{ij} > r_{ji}$ and 0 otherwise. Notice that OVA and other combination strategies in OVO such as VOTE and ND (Section II-B) managing the confidence for the non-predicted class differently need not be modified in order to achieve competitive results.

IV. Experimental Framework

In this section, we present the set-up of the experimental framework used to develop the experiments carried out in Section V. First, we describe the datasets selected for the experimental study (Section IV-A). Next, we show the parameter setup considered for each method (Section IV-B). Finally, we introduce the statistical tests that are necessary to assess whether significant differences exist among the results obtained (Section IV-C).

A. Datasets

In order to analyze the performance of our proposal, we have considered twenty datasets selected from the KEEL dataset repository [21]. Table I summarizes the features of the selected datasets, showing for each dataset the number of examples (#Ex.), number of attributes (#Atts.), number of numerical (#Num.) and nominal (#Nom.) attributes, and the number of classes (#Class.).

To carry out the different experiments we consider a *5-fold stratified cross-validation model*, i.e., we randomly split the dataset into five partitions of data, each one containing 20% of the patterns, and we employed a combination of four of them (80%) to train the system and the remaining one to test it. We use three different seeds for the execution of the methods in each partition. In this manner, the result for each dataset is obtained by computing the average of the five partitions using the three seeds in each one. Instead of the commonly used cross-validation and in order to correct the dataset shift, that is, when the training data and the test data do not follow the same distribution [62], [63], we will use a recently published partitioning procedure called *Distribution Optimally Balanced Cross Validation* [64].

B. State-of-the-art fuzzy classification methods used for comparison

In this section we briefly describe the different methods used throughout the experiments and the configuration that we have considered for each one. We have selected three
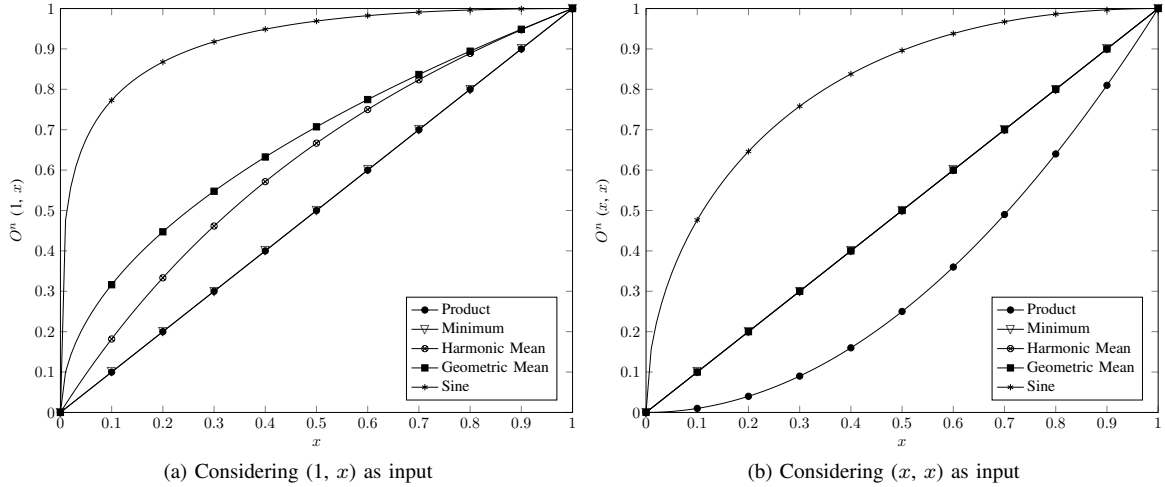
(a) Considering $(1, x)$ as input

(b) Considering (x, x) as input

Fig. 2.   Values returned by the different overlap functions.

TABLE I
SUMMARY OF THE FEATURES OF THE DATASETS USED IN THE
EXPERIMENTAL STUDY.

Id.	Dataset	#Ex.	#Atts.	#Num.	#Nom.	#Class.
aut	autos	159	25	15	10	6
bal	balance	625	4	4	0	3
cle	cleveland	297	13	13	0	5
con	contraceptive	1473	9	6	3	3
eco	ecoli	336	7	7	0	8
gla	glass	214	9	9	0	7
hay	hayes-roth	132	4	4	0	3
iri	iris	150	4	4	0	3
new	newthyroid	215	5	5	0	3
pag	pageblocks	548	10	10	0	5
pen	penbased	1100	16	16	0	10
sat	satimage	643	36	36	0	7
seg	segment	2310	19	19	0	7
shu	shuttle	2175	9	9	0	5
tae	tae	151	5	3	2	3
thy	thyroid	720	21	21	0	3
veh	vehicle	846	18	18	0	4
vow	vowel	990	13	13	0	11
win	wine	178	13	13	0	3
yea	yeast	1484	8	8	0	10

recognized state-of-the-art fuzzy classifiers to compare against our proposal, in addition to the comparison against the original FARC-HD algorithm.

- *FURIA* [24]: This algorithm modifies and extends the RIPPER rule induction algorithm [65]. In particular, FU-RIA learns fuzzy rules of the form given in Eq. (1) instead of conventional rules, using fuzzy sets with trapezoidal membership functions. Additionally, the model built by FURIA learns unordered rule sets instead of rule lists. The learning process is divided in two stages:
  1) Learn a rule set for each class using OVA decomposition. To do so, a modified and extended version of RIPPER is applied, which can be divided into the building and the optimization phase.
  2) Extract the fuzzy rules by fuzzifying RIPPER's rules using a greedy algorithm.

When classifying a new example, the class predicted by FURIA is the one with maximal support. If the example is not covered by any rule, a rule generalization (*stretching*) is carried out and all rules are replaced by their minimal generalizations, which is obtained by deleting all antecedents that are not satisfied by the query. In the case of a tie, a decision in favor of the class with highest frequency is made.

- *IVTURS$_{FARC-HD}$* [25]: This method uses FARC-HD to accomplish the fuzzy rule learning process and then, it substitutes the original fuzzy sets by interval-valued (IV) fuzzy sets and it modifies the inference process using an IV fuzzy reasoning method. This inference process uses IV restricted equivalence functions to increase the relevance of the rules in which the equivalence of the interval membership degrees of the patterns and the ideal membership degrees is greater. In addition, it combines a tuning of the parameters used in the IV fuzzy reasoning method and rule selection, in order to both decrease the complexity and increase the performance of the system.
- *PTTD* [26]: This method constructs a fuzzy pattern tree for each class (OVA decomposition) whose inner nodes are marked with generalized fuzzy logical operators and whose leaf nodes are associated with linguistic terms on input attributes. The learning algorithm used by PTTD builds the pattern tree in a top-down manner.

The configuration of the previous methods and that of our proposal is shown in Table II.

C. Performance measure and statistical tests

In order to test the performance of the different methods, we have used the most common metric, that is, the *accuracy rate*. This metric is defined as percentage of correctly classified examples related to the total number of examples. However, accuracy rate may not properly reflect the behaviour of different algorithms in multi-class problems as they do not take into account the classes of the examples in its computation. Therefore, the usage of additional metrics (not opposite, yet

TABLE II
SETUP OF THE METHODS PARAMETERS.

Algorithm	Parameters
FURIA	Num. of optimizations: 2
	Num. of folds: 3
FARC-HD and IVTURS$_{\text{FARC-HD}}$	Num. of linguistic labels per variable: 5
	Minimum Support: 0.05
	Minimum Confidence: 0.8
	Maximum depth: 3
	Parameter k: 2
	Evaluations: 20000
	Number of individuals: 50
	α parameter: 0.02
	Bits per gen: 30
	Rule weight: certainty factor
	Inference: Additive Combination
PTTD	e parameter: 0.25%
	Beam size: 5

complementary) increases the strength of the experimental study, yielding more complete conclusions. For this reason we have also considered Cohen's kappa [66] measure as an evaluation criterion, which evaluates the portion of hits that can be attributed to the classifier itself (i.e., not to mere chance), relative to all the classifications that cannot be attributed to chance alone.

For multi-class problems, kappa is a very useful, yet simple, meter for measuring a classifier's classification rate while compensating for random successes. The major difference between the classification rate and Cohen's kappa is the scoring of the correct classifications, since Cohen's kappa scores the successes independently for each class and aggregates them. This way of scoring is less sensitive to randomness caused by a different number of examples in each class. Nevertheless, for the sake of space we cannot include the experimental study carried out with this metric, but we provide it as a supplementary material of the paper. Anyway, we should stress that the conclusions drawn are equivalent to those obtained with accuracy along the whole experimental study.

Besides the performance measures used to evaluate the quality of the models, we want to study how the different overlap functions affect the rule base size. To do so, we consider the average number of rules and antecedents by rule for each overlap function in both OVO and OVA models (considering all base classifiers) and the FARC-HD algorithm (directly executed without decomposition strategies).

In order to give a statistical support to the analysis of the results, we carry out some non-parametric tests [22]. More specifically, we use the Wilcoxon signed-ranks test [67] to perform pairwise comparisons, the Aligned Friedman test [68] to check whether there are statistical differences among a group of methods and the Holm *post-hoc* test [69] to find the algorithms that reject the null hypothesis of equivalence against the selected control method. A complete description of these tests and software for their use can be found on the website available at: http://sci2s.ugr.es/sicidm/.

V. EXPERIMENTAL STUDY

In this section, we analyze the results obtained by our proposals developing an experimental study composed of three

steps (the same is done with kappa in the supplementary material):

1) We analyze the effect of overlap functions in the final performance of the model and we also study whether our new aggregation strategy for OVO (WinWV) allows to improve the results of the WV (Section V-A1). Additionally, we show how the usage of overlap functions affects the size of the rule base and the training times (Section V-A2).
2) We show whether our new model is a suitable solution for multi-class problems compared with the original FARC-HD [12] and we analyze which decomposition strategy obtains better results (Section V-B).
3) We study whether our proposal improves the results obtained by some of the state-of-the-art fuzzy classifiers, such as FURIA [24], IVTURS$_{\text{FARC-HD}}$ [25], and PTTD [26] (Section V-C).

A. Study of the behavior of n-dimensional overlap functions

In this section we first study the effect of the different overlap functions in the final performance of the system (Section V-A1) and then we show the impact of these functions in the rule base and in the training times (Section V-A2). Additionally, in order to check whether the proposed new aggregation strategy for OVO (WinWV) solves the problems of the WV with the confidences provided by FARC-HD, a comparison between the WV and the WinWV is performed (also in Section V-A1).

Tables III and IV show the accuracy rate obtained in testing by each method in all datasets, together with the standard deviation (shown with $\pm$). As we can observe in Table III, on the one hand, we execute the FARC-HD algorithm directly (with no decomposition strategies) using the five overlap functions considered in this paper (PROD, MIN, HM, GM, SIN). On the other hand, we present the results of OVA and OVO models considering the previously mentioned overlap functions for those aggregation strategies that are not affected by the confidences of FARC-HD in the case of OVO (ND and VOTE), whereas those being affected (LVPC, WV) are shown in Table IV, together with our proposed solution (WinWV).

1) Analysis of the system's performance:

As we can observe in Tables III and IV, in the case of the original FARC-HD, the replacement of the product by an overlap function does not seem to produce an improvement in the results. However, in the case of OVA and OVO models, we find that the greater overlap function we use, the better results we obtain in general (although the GM does not improve the results of the HM, due to the fact that they exhibit a similar behaviour). The exception to this situation is the usage of the greatest overlap function considered in this paper (SIN). This could be due to the fact that this function returns aggregated values that can be greater than the input ones,

TABLE III
ACCURACY RATE OBTAINED IN TEST BY EACH METHOD.

Dataset	FARC-HD					OVA					OVOND					OVO^{VOTE}				
	PROD	MIN	GM	HM	SIN	PROD	MIN	GM	HM	SIN	PROD	MIN	GM	HM	SIN	PROD	MIN	GM	HM	SIN
aut	80.30±6.79	80.94±3.39	78.85±7.62	80.24±7.33	**81.67±7.50**	77.66±7.03	79.24±6.79	79.29±4.69	79.29±5.85	78.09±7.93	79.09±7.95	80.18±5.38	**81.09±9.12**	80.50±8.10	77.96±10.26	79.27±6.79	80.83±3.39	80.23±7.62	**80.99±7.33**	79.35±7.50
bal	87.51±1.80	86.13±2.84	88.05±2.26	87.73±3.07	**88.53±2.38**	89.22±1.70	88.52±1.92	89.59±1.78	89.76±1.21	**90.08±1.42**	85.27±2.00	85.00±2.61	**86.60±1.81**	86.60±2.41	86.55±2.18	85.43±1.80	84.52±2.84	**86.71±2.26**	85.69±3.07	86.56±2.38
cle	57.80±3.48	56.97±4.92	55.90±4.18	56.80±5.31	57.36±5.50	57.27±4.12	57.32±2.61	56.48±3.61	56.35±4.54	57.95±4.49	54.58±3.91	55.15±3.96	57.06±4.51	**58.07±4.95**	56.39±4.18	57.49±3.48	56.95±4.92	57.62±4.18	**57.74±5.31**	56.95±5.50
con	53.63±2.06	53.07±1.82	53.29±2.31	54.08±1.78	53.63±1.88	55.13±1.62	**55.48±1.76**	55.45±1.47	54.72±1.76	54.60±1.43	54.45±1.98	54.28±3.61	53.77±1.68	**54.99±2.53**	54.65±1.67	54.45±2.06	54.20±1.82	53.77±1.78	**55.15±2.31**	54.27±1.88
eco	82.93±5.00	82.00±4.88	81.57±5.58	81.95±5.60	78.93±6.64	83.57±5.06	83.30±5.64	83.44±4.35	**84.28±5.57**	82.10±5.92	**84.63±4.23**	84.28±3.61	82.53±6.38	83.31±6.06	82.98±5.78	**84.07±5.00**	83.83±4.88	82.90±5.58	83.20±5.60	82.10±6.64
gla	66.28±5.99	65.58±4.19	71.06±5.03	70.06±4.11	**72.54±4.79**	67.85±6.61	68.68±3.55	70.61±6.85	71.31±5.39	**72.15±6.93**	68.40±4.39	71.08±4.16	70.86±4.80	69.90±4.61	**71.41±5.85**	69.90±5.99	**72.63±4.19**	71.23±5.03	70.18±4.11	71.83±6.79
hay	79.21±5.00	79.21±5.00	77.91±4.93	78.42±5.98	72.54±5.07	81.87±4.33	81.87±4.33	81.63±4.81	81.31±5.39	**82.89±5.08**	80.88±4.69	**80.90±4.79**	80.86±4.80	80.90±4.61	80.88±4.69	81.67±5.00	**81.67±5.00**	81.65±4.93	81.19±5.98	81.15±5.07
iri	95.56±4.08	96.22±3.29	95.56±3.29	95.11±3.52	95.78±4.15	95.78±2.95	94.89±3.53	95.33±3.29	95.56±3.25	94.89±3.75	95.33±4.14	95.11±3.53	94.89±3.75	95.11±3.75	93.56±4.27	95.78±4.08	95.33±3.29	95.33±3.29	95.33±3.52	93.78±4.15
new	95.04±2.70	96.28±2.90	95.19±2.13	96.12±2.61	94.26±5.23	95.66±2.62	95.81±2.67	95.66±3.27	95.81±2.19	93.49±2.67	95.35±2.78	96.43±2.90	95.97±2.06	96.43±2.62	92.40±5.52	95.19±2.70	**96.59±2.90**	95.66±2.13	96.43±1.49	92.40±5.23
pag	94.18±1.60	94.06±2.06	94.55±1.34	94.67±1.49	**94.72±1.43**	95.21±1.64	94.36±2.05	94.73±1.63	95.21±1.23	94.78±1.32	94.48±1.44	94.65±1.94	94.90±1.26	96.30±1.52	95.03±1.17	94.12±1.60	94.78±2.06	94.85±1.34	**96.43±1.49**	94.85±1.43
pen	93.05±3.23	91.92±2.47	92.58±2.49	92.58±2.49	93.63±3.48	93.59±3.27	93.59±2.27	93.83±2.52	**94.38±2.20**	92.83±3.10	**94.92±2.15**	94.43±2.26	93.92±2.24	94.92±2.23	94.25±2.22	94.59±2.23	94.34±2.47	94.13±3.23	94.22±2.49	94.25±2.22
sat	80.96±3.68	80.59±3.64	80.68±2.89	80.13±3.26	80.64±1.69	79.49±2.63	79.24±4.84	80.49±3.87	**80.78±1.97**	79.59±2.98	82.15±4.04	82.05±4.19	81.25±3.13	**82.60±2.92**	79.38±2.61	83.49±3.68	83.18±3.64	83.01±2.89	**84.05±3.26**	82.23±1.69
seg	93.62±0.62	92.76±1.14	92.94±1.21	93.02±1.30	**93.64±1.03**	91.85±1.00	**93.80±1.02**	92.83±1.57	92.73±1.34	93.31±1.65	93.46±0.44	94.32±1.06	95.25±1.33	95.22±1.29	**95.95±1.14**	92.99±0.62	94.23±1.14	95.09±1.21	94.99±1.30	**96.12±1.03**
shu	95.50±1.02	99.20±1.81	99.23±0.10	**99.45±0.22**	99.45±0.14	96.46±2.56	96.20±2.11	95.75±1.79	96.66±1.02	97.09±1.94	93.87±1.02	98.65±1.93	**99.74±0.08**	99.60±0.20	99.66±0.17	93.87±1.02	98.69±1.81	**99.71±0.10**	99.59±0.22	99.68±0.14
tae	57.89±11.35	59.03±9.16	57.75±8.94	57.44±8.42	55.97±9.29	56.44±9.40	57.05±9.31	56.40±7.74	56.42±7.63	52.83±9.07	55.65±11.46	57.44±9.40	56.43±8.42	55.65±11.46	60.58±8.70	55.64±11.35	60.55±8.94	**61.63±8.42**	60.55±8.94	59.88±9.29
tyr	93.52±0.38	92.51±0.80	92.46±0.64	92.46±0.64	92.37±0.64	93.20±1.46	92.37±0.88	92.51±0.64	92.51±0.64	92.46±0.62	**93.61±0.34**	92.65±0.80	92.46±0.62	92.51±0.64	92.51±0.64	**93.57±0.38**	92.65±0.80	92.51±0.64	92.51±0.64	92.51±0.64
veh	69.90±2.07	**71.47±3.00**	70.48±3.58	71.15±3.05	70.05±3.61	67.88±1.99	69.04±2.48	69.53±2.49	70.05±2.20	**70.83±2.99**	**73.85±2.60**	73.55±2.26	73.07±3.58	72.25±3.19	72.41±3.54	73.25±2.07	72.72±3.00	72.76±3.58	71.90±3.05	71.85±3.61
vow	75.32±2.38	72.59±3.89	66.26±3.72	67.47±1.14	67.17±2.19	65.99±3.27	68.25±3.41	61.58±1.84	64.34±3.02	66.50±3.93	90.74±2.82	90.24±3.41	90.54±3.63	**91.35±1.49**	89.39±1.86	90.51±2.38	89.26±3.89	90.57±3.72	**90.71±1.14**	90.20±2.19
win	95.56±3.99	93.82±3.03	94.39±3.55	94.39±3.55	92.79±3.59	95.12±2.90	95.55±1.90	96.24±2.57	**97.03±2.21**	94.37±2.37	95.75±3.52	**96.66±2.97**	95.71±3.30	94.94±4.27	94.76±3.55	95.20±3.99	**96.47±3.03**	96.11±3.07	94.55±3.55	94.78±3.59
yea	59.60±2.56	59.01±3.56	59.64±2.88	59.35±3.33	**59.84±3.22**	59.05±2.48	59.55±1.90	60.04±2.90	59.48±2.96	59.80±3.10	60.61±2.67	**60.66±3.34**	60.00±2.91	60.20±3.39	59.67±3.37	59.97±2.56	**60.05±3.56**	59.59±2.88	59.99±3.33	59.37±3.22
AVG	**80.37±3.44**	80.17±3.39	79.89±3.39	80.11±3.46	79.98±3.71	79.92±3.37	80.27±3.38	80.13±3.24	**80.48±3.08**	79.97±3.64	81.45±3.43	81.88±3.32	82.13±3.47	82.18±3.54	81.46±3.69	81.52±3.44	82.03±3.39	82.25±3.39	**82.26±3.46**	81.71±3.71

TABLE IV
ACCURACY RATE OBTAINED IN TEST BY EACH METHOD (II).

Dataset	OVO^{VPC}					OVO^{WV}					OVO^{WuWV}				
	PROD	MIN	GM	HM	SIN	PROD	MIN	GM	HM	SIN	PROD	MIN	GM	HM	SIN
aut	**76.28±7.55**	76.22±2.73	74.74±8.45	74.81±6.06	71.60±8.45	77.48±7.83	**79.55±3.94**	78.93±6.89	76.91±6.50	75.86±7.75	78.65±7.55	**81.40±2.73**	79.43±8.45	80.04±6.06	78.77±8.62
bal	83.35±2.48	83.19±3.03	84.90±2.62	84.81±6.06	84.31±2.02	81.85±2.40	80.46±3.54	81.54±1.72	**82.17±3.03**	80.09±2.62	84.30±2.48	84.08±3.03	85.80±2.09	85.48±2.62	**86.45±2.02**
cle	57.39±2.98	57.27±4.36	**59.43±5.27**	58.98±4.98	56.85±6.43	55.70±3.19	56.93±4.02	58.09±5.47	**58.20±6.26**	55.72±5.33	56.37±2.98	57.16±4.36	**57.86±5.27**	57.62±4.98	56.40±6.43
con	53.86±1.95	54.85±1.67	53.98±1.70	**55.56±2.17**	54.74±1.73	54.04±1.87	54.94±1.73	54.65±1.72	**55.53±2.20**	54.65±1.57	54.47±1.95	53.86±1.67	53.84±1.70	**55.24±2.17**	54.83±1.73
eco	**85.34±5.04**	84.02±4.88	53.98±1.70	82.94±5.60	82.22±5.54	**85.06±4.61**	84.78±5.63	81.56±4.66	82.78±4.69	81.78±5.72	82.38±4.99	84.73±4.88	84.73±4.99	**85.26±5.04**	82.68±5.54
gla	66.72±5.00	68.85±4.60	68.50±4.47	67.00±5.67	**69.26±6.34**	67.75±6.57	**69.26±6.86**	68.85±6.43	68.20±8.06	69.26±6.34	70.77±4.47	71.35±4.60	70.77±4.47	70.01±5.67	**72.38±6.34**
hay	**76.65±4.34**	76.65±4.34	76.17±4.34	76.15±5.14	74.65±4.56	81.14±3.80	81.14±3.80	81.39±3.91	**81.43±4.82**	81.67±4.56	82.16±4.34	82.16±4.34	**82.16±4.34**	81.94±5.14	81.67±4.56
iri	**96.22±4.02**	95.56±3.29	95.56±3.29	95.33±3.52	93.78±4.15	96.22±3.53	91.33±4.58	95.56±3.25	**96.22±3.53**	**97.15±0.48**	96.00±4.34	97.59±3.35	95.56±3.25	95.33±3.52	93.78±4.15
new	**92.09±2.09**	92.85±2.61	92.44±4.14	91.32±2.61	93.33±1.62	95.22±3.53	95.04±3.15	93.95±3.38	95.50±2.98	91.47±5.32	95.35±5.78	58.80±8.56	**96.59±2.90**	96.59±2.61	92.25±5.32
pag	92.59±1.64	92.85±2.21	92.44±1.14	93.15±1.46	91.77±2.17	93.93±1.73	93.74±3.06	94.30±1.60	**95.21±1.52**	91.01±5.19	95.35±2.78	94.12±1.64	94.90±2.21	**96.06±1.46**	93.49±2.47
pen	**94.32±2.01**	93.68±2.79	92.08±2.47	92.49±2.47	91.77±2.17	**94.20±2.42**	93.74±2.60	92.14±2.72	92.49±2.29	91.77±2.17	**94.77±2.01**	94.35±2.79	92.14±1.64	94.12±1.64	93.98±2.17
sat	79.55±3.23	79.55±3.23	77.76±3.02	**81.26±2.53**	70.53±4.79	78.62±3.87	78.62±4.75	79.10±4.32	**80.88±2.95**	70.84±4.75	83.59±3.23	82.97±3.77	82.97±3.23	**83.94±2.53**	79.62±4.79
seg	92.21±0.67	**92.54±1.35**	91.34±1.37	90.85±1.54	92.15±0.78	92.21±0.79	92.54±2.08	91.34±2.11	90.85±1.22	92.06±1.53	93.54±0.67	94.16±1.35	94.42±1.37	93.54±1.54	**95.09±0.78**
shu	91.82±1.00	88.72±3.35	87.93±1.44	89.78±2.28	97.15±0.48	99.23±1.75	91.33±4.58	92.00±4.52	94.15±4.58	**97.29±2.99**	93.86±1.00	97.59±3.35	99.09±1.44	98.74±2.28	**99.54±0.48**
tae	53.25±9.96	54.98±8.91	55.46±9.10	55.33±6.29	**58.58±8.89**	53.03±10.16	54.76±10.90	55.43±8.93	**55.68±11.15**	55.43±8.93	57.68±9.96	**58.80±8.56**	54.13±9.56	57.68±8.91	60.12±9.10
tyr	**93.15±0.40**	92.69±0.75	92.51±0.64	92.51±0.64	92.51±0.64	**93.11±0.68**	92.65±0.80	93.93±1.73	92.51±0.64	92.51±0.64	**93.47±0.40**	92.60±0.75	92.51±0.64	93.47±0.40	92.51±0.64
veh	**73.13±2.69**	72.96±2.60	71.89±3.89	69.81±3.21	70.95±3.45	72.97±2.70	72.92±2.40	72.01±4.35	70.05±3.04	71.19±3.28	**73.84±2.69**	73.08±2.60	72.95±3.89	71.94±3.21	72.13±3.45
vow	**83.27±2.25**	81.82±3.80	82.36±1.96	82.36±1.96	81.04±3.87	**83.27±3.01**	81.82±4.06	81.35±3.25	81.04±3.87	81.04±3.87	86.87±2.25	86.20±3.80	86.03±3.15	**86.90±1.96**	86.26±3.87
win	95.01±3.00	**96.83±3.12**	93.00±4.07	93.00±4.07	95.11±3.65	95.37±3.62	**96.08±3.10**	95.50±3.95	94.92±4.17	95.29±3.06	95.15±3.00	96.08±3.12	95.50±3.95	94.53±4.07	94.92±3.65
yea	59.10±2.29	**59.14±3.40**	56.29±3.29	58.08±3.11	57.08±3.70	59.50±2.88	**60.25±2.61**	57.33±3.18	59.57±3.26	57.08±2.61	59.46±3.29	**60.77±3.40**	57.33±3.11	60.72±3.11	59.38±3.70
AVG	**79.77±3.26**	79.61±3.39	79.06±3.44	79.29±3.54	78.80±3.94	80.19±3.54	80.16±3.77	80.05±3.83	**80.24±4.01**	78.96±3.95	81.50±3.26	81.86±3.39	81.89±3.44	**81.93±3.54**	81.39±3.94

which is not a desirable behaviour in an inference system because part of the discrimination power is lost. Therefore, we can observe that there is a limit beyond which an overlap function might be too great to obtain good results. Anyway, no meaningful conclusions can be extracted without carrying out the proper statistical analysis.

In order to detect significant differences among the results of each overlap function used throughout the experiments, we carry out the Aligned Friedman test and the Holm post-hoc test, whose results are shown in Table V. These results are grouped in columns according to the method used to perform the comparison and in rows according to the overlap function used to compute the association degree (which is the subject of the study). The first column corresponds to the different overlap functions over the original FARC-HD, while in the second one OVA model is considered. The rest of the columns correspond to the different OVO aggregation strategies considered in this work (LVPC, ND, VOTE, WV and WinWV). The value of each cell corresponds to the rank obtained with the Aligned Friedman test that compares the different overlap functions for each method (that is, a Aligned Friedman test is carried out for each group of methods in a column). The value shown in brackets corresponds to the adjusted p-value obtained by the Holm post-hoc test using as control method the one achieving the smallest rank in the same column, which is shown in **bold-face**. The adjusted p-value is underlined when there are statistical differences ($\alpha = 0.1$ considering the ratio between datasets and algorithms).

TABLE V
ALIGNED FRIEDMAN AND HOLM TESTS TO COMPARE THE DIFFERENT OVERLAPS.

	FARC-HD	OVA	OVO$^{\text{ND}}$	OVO$^{\text{VOTE}}$	OVO$^{\text{WinWV}}$	OVO$^{\text{LVPC}}$	OVO$^{\text{WV}}$
PROD	**43.80**	57.90 (0.128)	55.23 (0.327)	56.53 (0.269)	54.40 (0.747)	**37.90**	**42.38**
MIN	48.63 (0.967)	51.72 (0.282)	49.03 (0.708)	49.77 (0.672)	46.42 (1.000)	41.22 (0.717)	42.95 (1.000)
HM	50.22 (0.967)	**38.23**	40.52	40.95	**43.83**	54.05 (0.157)	43.90 (1.000)
GM	56.25 (0.699)	48.95 (0.282)	45.65 (0.708)	43.65 (0.768)	47.95 (1.000)	56.67 (0.122)	49.13 (1.000)
SIN	53.60 (0.856)	55.70 (0.170)	62.08 (0.075)	61.60 (0.097)	59.90 (0.319)	62.65 (0.028)	74.15 (0.002)

As it can be observed in the first column of Table V, in the original FARC-HD the best aggregation method is the product, although there are no statistical differences between the five aggregations. However, in OVA and OVO models, we can observe that the greater the overlap function we use is, the better the results we obtain are (although the GM is greater than the HM, both of them have a similar behaviour), except in the case of the SIN, due to the fact that it can give as output a value which is greater than all the input values, which seems to distort the aggregation process in the OVA and OVO strategies.

Hence, the best method to compute the association degree for OVA and OVO in almost all cases are those obtaining the highest aggregated values preserving the idempotence (although the geometric and harmonic means return similar values, the latter one tends to obtain better results but without statistical differences). The exception to this situation is when we use LVPC and WV strategies, since they are severely affected by the poor quality of the confidences of the non-predicted classes, which is accentuated in LVPC due to the difficulty in modeling the conflict and ignorance terms (as we

have mentioned in Section III-D). However, we must recall that removing these terms the original WV is recovered. On this account, we only focus on WV strategy.

For the sake of solving the problem of the WV with the confidences, we propose a new aggregation strategy for OVO (WinWV) which considers only the confidences of the predicted classes. This way, we want to study whether our proposal allows to improve the results of the WV when using FARC-HD and OVO. In the results presented in Table III, we can observe that the results obtained by WV are different from those obtained by the remainder OVO aggregations (except for LVPC which suffers the same problem), in the sense that using overlap function has no effect on the results. Focusing on the differences between WV and WinWV, we can observe that the usage of WinWV allows to enhance the results of WV.

In order to support this finding, we have carried out a number of pair-wise comparisons using the Wilcoxon signed-ranks test, where we confront the original WV method against the proposed modification for each overlap function considered in this study. Table VI shows the results of these comparisons, where R+ and R- indicate the ranks obtained by WV and WinWV, respectively. As we can observe, the new aggregation strategy statistically outperforms the original WV method with all overlap functions.

TABLE VI
WILCOXON TEST TO COMPARE THE WV AND THE WINWV.

Comparative	R+	R-	p-value	Hypothesis
OVO$^{\text{WV}}_{\text{PROD}}$ vs. OVO$^{\text{WinWV}}_{\text{PROD}}$	7.00	203.00	0.000	Rej. OVO$^{\text{WinWV}}_{\text{PROD}}$ 95%
OVO$^{\text{WV}}_{\text{MIN}}$ vs. OVO$^{\text{WinWV}}_{\text{MIN}}$	21.00	189.00	0.002	Rej. OVO$^{\text{WinWV}}_{\text{MIN}}$ 95%
OVO$^{\text{WV}}_{\text{HM}}$ vs. OVO$^{\text{WinWV}}_{\text{HM}}$	14.50	195.50	0.001	Rej. OVO$^{\text{WinWV}}_{\text{HM}}$ 95%
OVO$^{\text{WV}}_{\text{GM}}$ vs. OVO$^{\text{WinWV}}_{\text{GM}}$	13.50	196.50	0.001	Rej. OVO$^{\text{WinWV}}_{\text{GM}}$ 95%
OVO$^{\text{WV}}_{\text{SIN}}$ vs. OVO$^{\text{WinWV}}_{\text{SIN}}$	5.50	204.50	0.000	Rej. OVO$^{\text{WinWV}}_{\text{SIN}}$ 95%

2) Analyzing the effect of the usage of n-dimensional overlap functions in the rule base size and the training time:

In addition to the performance of the different overlap functions, we want to study the impact of these functions in the size of the rule base and in the time needed for its construction. Table VII shows the average number of rules and antecedents by rule obtained when the considered overlap functions are used in FARC-HD (without decomposition strategies) and in OVA and OVO models, as well as the number of base classifiers (#BC) employed in OVA and OVO for each dataset. As it can be appreciated in Table VII, the usage of a greater overlap function implies a growing trend of the number of rules and a higher complexity of those rules (more antecedents). Thus, there is a relationship between the value returned by the overlap function and the size of the rule base (the greater the overlap function the larger the rule base). It is also interesting to note that the execution times of the methods with the different overlap functions (shown in Table VIII) follows the same trend (the construction of the rule base is only shown since the computational time required by the

TABLE VII

AVERAGE NUMBER OF ANTECEDENTES AND RULES IN FARC-HD, OVA AND OVO.

	FARC-HD												OVA												OVO										
	avg. rules					avg. antecedents					#BC	avg. rules					avg. antecedents					#BC	avg. rules					avg. antecedents				
Dataset	PROD	MIN	HM	GM	SIN	PROD	MIN	HM	GM	SIN		PROD	MIN	HM	GM	SIN	PROD	MIN	HM	GM	SIN		PROD	MIN	HM	GM	SIN	PROD	MIN	HM	GM	SIN
aut	36.67	33.47	33.33	33.53	31.60	2.67	2.66	2.67	2.70	2.80	6	11.40	11.33	11.10	11.59	11.82	2.03	2.02	2.02	2.06	1.99	15	4.68	4.64	4.58	4.56	4.52	1.61	1.59	1.54	1.57	1.59
bal	61.47	60.93	54.93	55.53	55.93	1.83	1.84	1.81	1.81	1.82	3	36.62	37.84	35.64	36.00	34.53	1.50	1.50	1.49	1.48	1.49	3	19.98	21.20	21.53	22.22	22.29	1.51	1.51	1.51	1.51	1.51
cle	62.27	60.40	65.20	63.13	69.00	2.90	2.90	2.88	2.89	2.92	4	17.40	18.41	19.91	19.83	20.56	2.36	2.38	2.43	2.42	2.42	10	13.59	13.43	13.94	14.84	14.92	2.55	2.49	2.54	2.54	2.53
con	64.60	92.40	106.67	107.67	122.33	2.68	2.79	2.86	2.90	2.94	5	48.87	61.96	73.49	77.07	85.80	2.51	2.59	2.70	2.71	2.74	3	49.80	64.67	76.13	79.02	88.69	2.65	2.73	2.81	2.83	2.87
eco	30.73	36.07	38.53	41.87	41.40	2.45	2.56	2.65	2.70	2.76	3	7.99	9.32	9.69	10.01	10.64	1.57	1.60	1.72	1.76	1.78	28	4.00	4.06	4.65	4.75	4.55	1.34	1.39	1.42	1.43	1.44
gla	22.53	25.27	29.40	29.93	37.07	2.56	2.70	2.79	2.77	2.75	6	6.86	7.16	8.14	8.62	10.02	1.60	1.63	1.69	1.71	1.75	21	2.85	2.85	3.46	3.50	3.76	1.14	1.15	1.14	1.13	1.10
hay	18.33	18.33	18.87	17.20	17.33	2.08	2.08	2.12	2.12	2.11	6	14.87	14.87	15.33	15.24	14.78	1.69	1.69	1.71	1.70	1.69	3	9.33	9.33	9.18	9.09	9.00	1.66	1.66	1.64	1.64	1.61
iri	4.20	4.13	5.80	5.80	7.40	1.10	1.10	1.46	1.36	1.70	6	4.29	4.84	5.42	5.60	6.13	1.14	1.11	1.25	1.24	1.44	3	2.60	2.60	2.89	2.89	3.51	1.07	1.10	1.23	1.20	1.28
new	10.00	9.00	10.00	9.87	10.20	1.63	1.54	1.73	1.54	1.68	7	5.78	6.13	6.73	6.09	7.60	1.42	1.30	1.36	1.31	1.46	3	3.42	3.64	3.96	3.82	4.60	1.20	1.19	1.14	1.10	1.23
pag	13.27	13.80	13.20	13.00	14.27	2.43	2.43	2.17	2.28	2.32	3	5.21	5.05	5.45	5.56	5.59	1.80	1.75	1.70	1.78	1.74	10	3.09	2.94	2.97	2.89	3.16	1.34	1.23	1.29	1.31	1.41
pen	78.40	82.47	95.53	101.20	103.00	2.80	2.85	2.90	2.92	2.92	3	15.58	15.71	16.85	17.45	17.38	1.86	1.87	1.99	2.00	1.99	45	4.66	4.71	4.92	4.91	4.90	1.27	1.27	1.34	1.35	1.39
sat	17.27	17.87	23.80	25.40	31.07	2.66	2.83	2.83	2.83	2.87	10	3.94	3.97	7.31	7.42	9.03	1.43	1.63	1.54	1.51	1.52	21	1.98	2.37	3.48	3.74	4.09	1.34	1.33	1.14	1.13	1.27
seg	41.60	44.13	61.07	61.13	69.73	2.66	2.71	2.77	2.78	2.77	4	10.93	11.10	14.06	14.76	15.07	1.72	1.74	1.84	1.85	1.85	10	3.99	4.61	6.11	6.24	6.98	1.33	1.31	1.40	1.40	1.44
shu	7.07	5.13	6.87	6.27	6.73	1.87	1.82	2.00	2.03	1.69	3	3.29	2.37	2.84	2.89	3.36	1.38	1.40	1.38	1.23	1.23	3	2.27	2.16	2.28	2.29	2.19	1.42	1.34	1.29	1.24	1.21
tae	20.40	27.47	28.80	28.33	28.93	2.54	2.63	2.78	2.77	2.76	5	16.40	17.80	19.56	20.60	21.53	2.26	2.39	2.56	2.57	2.59	3	14.18	16.27	17.58	18.71	19.16	2.34	2.45	2.56	2.58	2.58
thy	4.53	3.67	4.93	3.60	5.87	2.13	1.91	1.91	1.80	1.85	5	3.78	3.38	3.71	3.33	3.38	1.78	1.72	1.54	1.22	1.35	6	2.87	3.11	5.76	4.80	5.20	1.79	1.75	1.83	1.75	1.62
veh	44.87	52.33	66.33	75.27	92.33	2.66	2.88	2.94	2.93	2.92	10	20.82	25.90	35.85	39.83	50.98	2.34	2.50	2.60	2.60	2.68	55	13.19	17.01	24.47	27.18	29.88	2.14	2.33	2.52	2.57	2.58
vow	71.53	77.20	82.33	84.07	113.13	2.87	2.91	2.95	2.96	2.97	7	12.20	13.00	13.72	13.25	17.85	1.93	1.91	2.03	2.03	2.06	3	5.50	5.51	6.91	7.43	7.39	1.79	1.86	1.99	2.02	2.04
win	8.13	8.40	9.87	10.13	11.00	1.57	1.81	1.87	1.92	2.02	7	5.91	5.76	6.98	7.09	7.02	1.31	1.46	1.49	1.54	1.54	45	3.53	3.40	3.96	4.07	4.67	1.10	1.08	1.07	1.08	1.23
yea	35.53	41.60	47.47	49.33	57.73	2.64	2.74	2.81	2.82	2.89	5	8.55	9.35	10.07	10.21	13.36	1.66	1.68	1.81	1.84	1.89	45	5.51	5.89	6.75	7.09	8.18	1.64	1.73	1.77	1.76	1.84
AVG.	32.67	35.70	40.15	41.11	46.30	2.34	2.38	2.44	2.44	2.47	5.50	13.03	14.26	16.09	16.62	18.32	1.76	1.79	1.84	1.83	1.86	15.55	8.55	9.72	11.27	11.72	12.58	1.61	1.63	1.66	1.66	1.69

TABLE VIII

AVERAGE EXECUTION TIMES OF FARC-HD, OVA AND OVO. (MM:SS.MS).

	FARC-HD					#BC	OVA					#BC	OVO				
Dataset	PROD	MIN	HM	GM	SIN		PROD	MIN	HM	GM	SIN		PROD	MIN	HM	GM	SIN
aut	01:42.9	03:48.3	03:28.7	04:26.8	06:21.8	6	04:11.6	05:06.8	04:04.5	05:33.4	07:52.3	15	10:39.1	06:51.6	08:43.8	07:19.5	07:01.5
bal	00:22.9	00:32.0	00:23.3	03:13.5	04:15.2	3	01:23.4	01:16.3	00:54.0	02:59.3	04:29.9	3	00:39.6	01:03.2	01:14.2	02:21.1	02:33.7
cle	00:52.3	01:22.8	01:03.9	01:41.9	02:44.4	4	01:54.7	01:48.9	01:37.5	02:37.6	03:59.5	10	02:50.9	02:51.3	03:21.4	03:53.4	03:27.8
con	01:29.5	03:53.9	02:32.4	17:26.7	25:50.9	5	05:43.5	08:52.7	06:17.4	18:20.5	27:14.7	3	04:25.4	06:38.5	05:33.8	14:18.8	18:03.3
eco	00:17.9	00:38.0	00:27.5	00:58.9	01:29.3	3	00:47.0	00:52.3	00:48.9	02:02.3	02:48.9	28	02:00.4	02:37.3	03:21.4	04:03.9	02:17.3
gla	00:27.0	00:55.2	00:43.5	01:02.5	01:27.7	6	00:39.6	00:41.2	00:56.0	01:10.1	01:48.3	21	01:30.4	01:37.4	02:38.9	03:20.3	02:47.9
hay	00:04.3	00:06.3	00:04.3	00:10.3	00:15.0	6	00:13.7	00:12.1	00:09.7	00:25.2	00:34.7	3	00:12.0	00:19.4	00:44.0	01:12.4	00:22.8
iri	00:01.5	00:02.5	00:02.8	00:07.1	00:10.7	6	00:04.5	00:06.0	00:04.7	00:11.7	00:18.3	3	00:07.1	00:17.5	00:35.3	00:46.1	00:14.9
new	00:02.6	00:04.0	00:04.3	00:14.2	00:18.4	7	00:07.4	00:08.4	00:09.4	00:18.8	00:25.6	3	00:14.6	00:17.6	00:47.7	01:02.3	00:25.3
pag	00:13.6	00:22.9	00:20.2	00:30.6	01:01.2	3	00:42.0	00:42.0	00:41.4	01:15.4	02:06.2	10	01:01.6	01:58.0	01:58.0	02:26.8	01:48.8
pen	03:34.5	08:00.9	05:22.8	12:19.9	14:14.2	3	10:35.9	15:02.9	10:15.3	17:08.7	22:39.9	45	13:14.3	18:07.2	23:52.6	31:26.8	44:06.7
sat	08:53.4	16:34.4	15:38.5	20:16.2	22:58.6	10	07:25.6	09:28.7	08:57.3	14:09.8	17:56.2	21	17:25.3	23:37.5	20:10.7	35:55.2	28:34.4
seg	02:31.4	05:44.2	06:04.3	10:49.3	19:15.4	4	09:33.9	11:04.6	10:54.5	19:11.4	24:57.5	10	11:05.8	14:59.9	14:43.6	21:20.6	26:09.2
shu	00:17.4	00:36.2	00:34.7	01:05.6	01:33.9	3	01:28.5	01:26.6	01:15.8	02:35.2	04:15.8	3	01:19.7	02:30.3	02:52.6	03:39.9	03:56.3
tae	00:05.3	00:10.8	00:10.9	00:31.8	00:39.3	5	00:18.5	00:12.4	00:22.5	00:35.8	00:52.9	3	00:16.9	00:26.3	00:42.2	01:20.6	00:42.1
thy	00:11.7	00:28.9	00:23.8	00:50.2	00:55.8	5	00:50.2	00:55.6	00:50.2	01:04.5	01:39.8	6	01:08.7	01:03.5	01:24.7	01:51.0	01:10.3
veh	01:30.5	04:18.9	04:22.5	06:53.2	12:07.2	10	05:24.6	09:54.4	07:40.6	13:41.0	20:09.3	55	04:03.8	27:25.8	08:43.2	11:04.1	13:36.5
vow	02:50.0	06:07.2	05:17.4	07:10.5	11:08.3	7	06:41.7	07:33.2	07:24.9	11:46.4	19:50.3	3	40:17.3	01:02.6	38:41.0	35:54.1	00:36.8
win	00:09.6	00:33.7	00:34.1	00:49.3	01:05.0	7	00:32.0	00:39.7	01:04.2	01:09.5	01:57.8	45	00:23.5	10:13.3	00:41.0	01:30.5	00:52.7
yea	00:46.4	01:57.3	01:54.8	05:00.2	06:58.4	5	04:33.3	05:42.8	05:03.4	10:00.4	15:58.7	45	07:00.9	08:31.3	08:31.3	14:28.6	14:51.3
AVG.	01:19.2	02:48.9	02:28.6	04:46.9	06:44.5	5.5	03:09.5	04:06.7	03:28.5	06:18.8	09:05.8	15.55	05:59.8	06:27.1	07:28.8	09:57.8	11:40.9

different aggregations in OVO is negligible). The reason is that the use of greater overlap functions implies that the aggregation of the matching degrees returns higher values and thus, a higher number of rules and antecedents is needed in order to maintain or improve the discrimination capability (therefore affecting the computational time needed). Moreover, due to the fact that we focus on overlap functions that are independent of the number of antecedents of the rule, the penalization of those rules with more antecedents is minimized. As a consequence, more rules with larger number of antecedents can be learned, which better describe the examples. Table VII confirms this situation, where we can find that the average number of rules of the product is the lowest one among the five overlap functions and increases when considering a greater overlap function.

In Table VII we observe that the average number of rules obtained in OVA and OVO is lower than that obtained in FARC-HD. The reason is that we consider the average of all base classifiers and since they solve binary problems, the definition of the decision boundary in each binary problem is simpler than in the original multi-class problem, which implies that fewer rules are needed in each base classifier. We can also observe that in the case of OVO the number of rules generated by the base classifiers is lower than in OVA, since the binaries problems solved in OVO are simpler. In the same way, the rules generated in OVA and OVO are simpler than those generated in the original FARC-HD algorithm (at the same time the rules generated in OVO are simpler than in OVA, for the reason explained before), but obviously, we have more classifiers in the case of OVA and OVO.

B. Studying the usefulness of decomposition strategies for FARC-HD

In this section we want to check whether our new model improves the performance of the original FARC-HD algorithm when addressing multi-class classification problems. We have shown that the harmonic mean is the overlap function that obtains the best results when using OVO and OVA. Thus, before performing a comparison against FARC-HD, we analyze which aggregation strategy for OVO obtains better performance using the harmonic mean. As we can observe in the results provided by the Aligned Friedman test in Table IX, there are no statistical differences among ND, VOTE, and WinWV (as usually occurs among OVO aggregations [9]). For this reason, we will consider the aggregation strategy that obtains the highest accuracy according to Tables III and IV and the lowest ranks according to Table IX (VOTE).

TABLE IX
ALIGNED FRIEDMAN AND HOLM TESTS TO COMPARE THE DIFFERENT AGGREGATIONS IN OVO.

algorithm	rank	APV
OVO_{HM}^{VOTE}	**32.30**	
OVO_{HM}^{ND}	35.10	1.000
OVO_{HM}^{WinWV}	36.55	1.000
OVO_{HM}^{WV}	69.47	0.002
OVO_{HM}^{LVPC}	79.08	0.000

In order to check whether there are statistical differences among OVA, OVO, and the original FARC-HD, we show the

results of the Aligned Friedman test in Table X. It can be observed that OVO model statistically outperforms the original FARC-HD algorithm and obtains better results than the OVA model, which is in accordance with the findings using other classifiers [9].

TABLE X
ALIGNED FRIEDMAN AND HOLM TESTS TO COMPARE OVA, OVO, AND FARC-HD.

algorithm	rank	APV
OVO_{HM}^{VOTE}	**20.23**	
OVA_{HM}	31.55	0.040
FARC-HD	39.73	0.001

C. Analyzing the quality of FARC-HD_OVO versus state-of-the-art fuzzy classifiers

This section analyzes the performance of our model against three recognized state-of-the-art fuzzy classifiers, i.e., the IVTURS$_{FARC-HD}$ algorithm [25] by Sanz et al., the FURIA algorithm [24] by Hühn and Hüllermeier, and the PTTD method [26] by Senge and Hüllermeier. The results in testing of these three algorithms along with those obtained by OVO_{HM}^{VOTE} (denoted as FARC-HD_OVO$_{HM}^{VOTE}$) are shown in Table XI, where the best of the results obtained in each dataset is highlighted in **bold-face**.

From the results presented in Table XI, we must highlight the notable performance improvement of our proposal respect to IVTURS$_{FARC-HD}$, FURIA, and PTTD, improving their average performances by 3.79%, 1.71%, and 1.22%, respectively. However, we must contrast these results with the proper statistical analysis.

TABLE XI
ACCURACY RATE OBTAINED IN TEST BY THE STATE-OF-THE-ART FUZZY CLASSIFIERS AND THE MOST ACCURATE METHOD OF OUR MODEL BASED ON FARC-HD FUZZY CLASSIFIER.

Dataset	PTTD	FURIA	IVTURS$_{FARC-HD}$	FARC-HD_OVO$_{HM}^{VOTE}$
aut	75.07±6.06	75.66±4.79	77.07±9.11	**80.99**±7.33
bal	**89.28**±1.36	83.14±2.17	85.75±1.92	85.69±3.07
cle	**59.96**±4.36	55.11±1.35	57.47±3.23	57.74±5.31
con	54.17±1.47	**55.37**±2.06	54.47±1.44	55.15±2.31
eco	82.31±5.93	82.96±5.01	81.34±7.33	**83.20**±5.60
gla	63.69±6.61	**72.05**±5.72	69.12±6.17	70.18±4.11
hay	**84.12**±4.00	79.65±5.70	75.46±8.63	81.19±5.98
iri	**96.67**±2.36	94.22±3.44	95.78±2.95	95.33±3.52
new	**96.74**±2.08	94.88±3.31	94.26±1.73	96.28±2.61
pag	95.26±0.75	**96.50**±1.77	94.96±2.10	96.43±1.49
pen	92.47±2.36	91.07±1.67	92.22±2.48	**94.22**±2.49
sat	**86.48**±2.80	83.09±4.47	75.40±3.03	84.05±3.26
seg	93.12±0.89	**97.27**±0.81	90.56±0.91	94.99±1.30
shu	98.48±0.48	**99.68**±0.24	91.88±1.48	99.59±0.22
tae	53.55±8.63	44.51±5.72	54.83±7.80	**60.55**±8.94
tyr	96.53±0.98	**98.37**±1.73	93.85±0.64	92.51±0.64
veh	71.27±3.05	**71.91**±1.73	67.34±2.24	71.90±3.05
vow	75.96±2.54	82.36±2.97	65.99±2.14	**90.71**±1.14
win	**97.20**±1.91	94.79±2.01	95.18±3.03	94.55±3.55
yea	58.43±3.43	58.36±2.51	56.43±2.41	**59.99**±3.33
AVG	81.04±3.10	80.55±2.96	78.47±3.54	**82.26**±3.46

In order to compare these methods, we have applied the Aligned Friedman test. The rankings of the different methods computed using this test are shown in Fig. (3). The p-value

obtained is 0.001, which implies the existence of significant differences among the four methods.
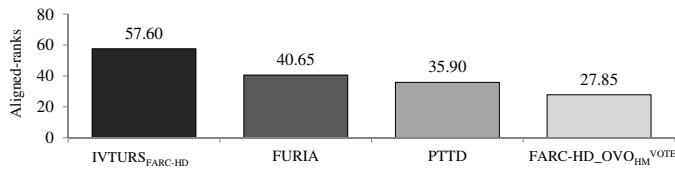

Fig. 3. Rankings of the state-of-the-art fuzzy classifiers along with our proposal.

We now apply the Holm post-hoc test to compare the best ranking method (FARC-HD_OVO$_{HM}^{VOTE}$) with the remaining methods. Table XII shows the results obtained by this test, indicating whether the hypothesis of equivalence is rejected by our proposal and the computed p-value (APV). According to Table XII, the hypothesis of equivalence is rejected in the case of IVTURS$_{FARC-HD}$ with a high level of confidence. Regarding FURIA and PTTD, although the hypothesis is not rejected, the APV values are low, which denotes that the behavior of our proposal is very competitive against these state-of-the-art fuzzy classifiers.

TABLE XII
ALIGNED FRIEDMAN AND HOLM TESTS TO COMPARE OUR PROPOSAL (FARC-HD_OVO$_{HM}^{VOTE}$) WITH RESPECT TO IVTURS$_{FARC-HD}$, FURIA, AND PTTD.

algorithm	rank	APV	Hypothesis
FARC-HD_OVO$_{HM}^{VOTE}$	**27.85**		
PTTD	35.90	0.273	Not Rejected
FURIA	40.65	0.163	Not Rejected
IVTURS$_{FARC-HD}$	57.60	0.001	Rej. FARC-HD_OVO$_{HM}^{VOTE}$ 95%

VI. CONCLUDING REMARKS

In this paper, we have combined the FARC-HD algorithm and OVO and OVA decomposition strategies to improve its performance in multi-class classification problems. We have shown that the confidences returned by FARC-HD may adversely affect the aggregation phase in these decomposition strategies and thus, the final prediction.

In order to minimize this negative effect, we have defined the concept of n-dimensional overlap functions and we have replaced the product t-norm by these functions in the inference system. Additionally, we have proposed a new aggregation strategy for OVO called WinWV, which solves the problems of the WV with the confidences of FARC-HD.

These adaptations have allowed us to show the importance of the inference process when OVO and OVA models are considered, since the confidence values are used beyond the FARC-HD classification. We have shown that the overlap functions that obtain the best results are those which return values with a higher variation and preserve the idempotence. Furthermore, we have observed that there is a relationship between the used overlap functions and the rule base size as well as the computational time spent in its learning. In addition, we have found that the usage of decomposition strategies is suitable for the FARC-HD classifier, but this synergy is better when the inference process is adapted appropriately

and the best results are obtained with OVO scheme which is in accordance with previous works. Moreover, the experimental study shows that our model obtains competitive results in comparison with three state-of-the-art fuzzy classifiers.

In the future, several works remain to be addressed. Among them, the problem of non-competent classifiers [19] must be taken into account when using the OVO model. On the other hand, a more in depth study of how this type of synergies affect the interpretability of the model should be carried out. Furthermore, our proposal might be adapted to different fuzzy classifiers in order to generalize the effect of the usage of overlap functions in the inference process when combining fuzzy classifiers and decomposition strategies. Finally, the comparison and combination between decomposition-based techniques and preprocessing-based fuzzy ensembles such as bagging [44] could be studied, but in this case, only focusing on improving the classification performance using fuzzy techniques.

REFERENCES

[1] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining.* Springer-Verlag, 2004.
[2] F. Chávez, F. Fernández, R. Alcalá, J. Alcalá-Fdez, G. Olague, and F. Herrera, "Hybrid laser pointer detection algorithm based on template matching and Fuzzy Rule-Based Systems for domotic control in real home enviroments," *Applied Intelligence*, vol. 36, no. 2, pp. 407-423, 2012.
[3] C. Tsang, S. Kwong, and H. Wang, "Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection," *Pattern Recognition*, vol. 40, no. 9, p. 2373-2391, 2007.
[4] T. Nakashima, G. Schaefer, and Y. Yokota, "A weighted fuzzy classifier and its application to image processing tasks," *Fuzzy Sets and Systems*, vol. 158, no. 3, p. 284-294, 2007.
[5] J. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, "Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system," *Applied Soft Computing Journal*, 2013, Article in Press.
[6] M. Galar, J. Sanz, M. Pagola, H. Bustince, and F. Herrera, "A preliminary study on fingerprint classification using fuzzy rule-based classification systems," in *2014 IEEE World Congress on Computational Intelligence (IEEE WCCI 2014) - 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2014)*, 2014.
[7] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318-1325, 2012.
[8] V. Bolón-Canedo, N. S.-M. no, and A. Alonso-Betanzos, "An ensemble of filters and classifiers for microarray data classification," *Pattern Recognition*, vol. 45, no. 1, pp. 531-539, 2012.
[9] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761 - 1776, 2011.
[10] A. Lorena, A. Carvalho, and J. Gama, "A review on the combination of binary classifiers in multiclass problems," *Artificial Intelligence Review*, vol. 30, no. 1-4, pp. 19-37, 2008.
[11] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal of Machine Learning Research*, vol. 1, pp. 113-141, 2000.
[12] J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 857-872, 2011.
[13] J. Fürnkranz, "Round robin classification," *Journal of Machine Learning Research*, vol. 2, pp. 721-747, 2002.
[14] J. Fürnkranz, "Round robin ensembles," *Intelligent Data Analysis*, vol. 7, no. 5, pp. 385-403, 2003.

[15] J. A. Sáez, M. Galar, J. Luengo, and F. Herrera, "Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition," *Knowledge and Information Systems*, vol. 38, no. 1, pp. 179-206, 2014.

[16] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera, "Empowering difficult classes with a similarity-based aggregation in multi-class classification problems," *Information Sciences*, vol. 264, pp. 135-157, 2014.

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

[18] E. Hüllermeier and S. Vanderlooy, "Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting," *Pattern Recognition*, vol. 43, no. 1, pp. 128-142, 2010.

[19] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "Dynamic classifier selection for one-vs-one strategy: Avoiding noncompetent classifiers," *Pattern Recognition*, vol. 46, no. 12, pp. 3412-3424, 2013.

[20] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, and R. Orduna, "Overlap functions," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 72, no. 3-4, pp. 1488-1499, 2010.

[21] J. Alcalá-Fdez, A. Fernandez, 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*, vol. 17:2-3, pp. 255-287, 2011.

[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 Comput.*, vol. 13, no. 10, p. 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*, vol. 180, no. 10, pp. 2044-2064, 2010.

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

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

[26] R. Senge and E. Hüllermeier, "Top-down induction of fuzzy pattern trees," *IEEE Transactions on Fuzzy Systems*, vol. 19, pp. 241 - 252, 2011.

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

[28] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *International Conference on Very Large Data Bases*, Santiago de Chile, Chile, September 1994, p. 487-499.

[29] B. Kavsek and N. Lavrac, "Apriori-sd: Adapting association rule learning to subgroup discovery," *Applied Artificial Intelligence*, vol. 20, no. 7, p. 543-583, 2006.

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

[31] J. Friedman, "Another approach to polychotomous classification," Department of Statistics, Stanford University, Tech. Rep., 1996. [Online]. Available: http://www-stat.stanford.edu/~jhf/ftp/poly.ps.Z

[32] A. Fernández, M. Calderón, E. Barrenechea, H. Bustince, and F. Herrera, "Solving mult-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations," *Fuzzy Sets and Systems*, vol. 161, no. 23, pp. 3064-3080, 2010.

[33] E. Hüllermeier and K. Brinker, "Learning valued preference structures for solving classification problems," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2337-2352, 2008.

[34] J. C. Huhn and E. Hullermeier, "FR3: A fuzzy rule learner for inducing reliable classifiers," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 1, pp. 138-149, 2009.

[35] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

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

[37] L. I. Kuncheva, "Diversity in multiple classifier systems," *Information Fusion*, vol. 6, no. 1, pp. 3-4, 2005.

[38] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.

[39] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, pp. 197-227, 1990.

[40] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[41] H. Ishibuchi and Y. Nojima, "Fuzzy ensemble design through multi-objective fuzzy rule selection," in *Multi-Objective Machine Learning*, ser. Studies in Computational Intelligence, Y. Jin, Ed. Springer Berlin Heidelberg, 2006, vol. 16, pp. 507-530.

[42] P. Bonissone, J. M. Cadenas, M. C. Garrido, and R. A. Díaz-Valladares, "A fuzzy random forest," *International Journal of Approximate Reasoning*, vol. 51, no. 7, pp. 729-747, 2010.

[43] K. Trawiński, O. Cordón, and A. Quirin, "On designing fuzzy rule-based multiclassification systems by combining FURIA with bagging and feature selection," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 4, pp. 589-633, 2011.

[44] K. Trawinski, O. Cordon, L. Sanchez, and A. Quirin, "A genetic fuzzy linguistic combination method for fuzzy rule-based multiclassifiers," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 5, pp. 950-965, 2013.

[45] J. M. Cadenas, M. C. Garrido, R. Martínez, and P. P. Bonissone, "Extending information processing in a fuzzy random forest ensemble," *Soft Computing*, vol. 16, no. 5, pp. 845-861, 2012.

[46] A. M. Palacios, L. Sánchez, and I. Couso, "Boosting of fuzzy rules with low quality data," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 19, no. 5-6, pp. 591-619, 2012.

[47] K. Trawiński, O. Cordón, and A. Quirin, "A study on the use of multiobjective genetic algorithms for classifier selection in FURIA-based fuzzy multiclassifiers," *International Journal of Computational Intelligence Systems*, vol. 5, no. 2, pp. 231-253, 2012.

[48] B. Soua, A. Borgi, and M. Tagina, "An ensemble method for fuzzy rule-based classification systems," *Knowledge and Information Systems*, vol. 36, no. 2, pp. 385-410, 2013.

[49] E. Hüllermeier and K. Brinker, "Learning valued preference structures for solving classification problems," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2337-2352, 2008.

[50] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Hybridization of fuzzy GBML approaches for pattern classification problems," *IEEE Transactions on System, Man and Cybernetics B*, vol. 35, no. 2, pp. 359-365, 2005.

[51] A. González and R. Perez, "SLAVE: a genetic learning system based on an iterative approach," *Fuzzy Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 176-191, 1999.

[52] Y.-C. Chen, N. R. Pal, and I.-F. Chung, "An integrated mechanism for feature selection and fuzzy rule extraction for classification," *IEEE Transaction on Fuzzy Systems*, vol. 20, no. 4, pp. 683-698, 2012.

[53] J. Castro-Schez, M. J.M., R. Miguel, and X. Luo, "Knowledge acquisition based on learning of maximal structure fuzzy rules," *Knowledge-Based Systems*, vol. 44, pp. 112-120, 2013.

[54] A. Acilar and A. Arslan, "A novel approach for designing adaptive fuzzy classifiers based on the combination of an artificial immune network and a memetic algorithm," *Information Sciences*, vol. 264, pp. 158-181, 2014.

[55] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," *Information Sciences*, vol. 181, no. 20, pp. 4340-4360, 2011.

[56] Z. H. Deng, L. B. Cao, Y. Z. Jiang, and S. T. Wang, "Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model," *IEEE Transactions on Fuzzy Systems*, 2014.

[57] D. Paternain, M. Pagola, J. Fernandez, R. Mesiar, G. Beliakov, and H. Bustince, "Brain MRI thresholding using incomparability and overlap functions," in *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, 2011, pp. 808-812.

[58] A. Jurio, H. Bustince, M. Pagola, A. Pradera, and R. R. Yager, "Some properties of overlap and grouping functions and their application to image thresholding," *Fuzzy Sets and Systems*, vol. 229, pp. 69-90, 2013.

[59] D. Paternain, J. Fernandez, H. Bustince, R. Mesiar, and G. Beliakov, "Construction of image reduction operators using averaging aggregation functions," *Fuzzy Sets and Systems*, vol. in press, 2014.

[60] H. Bustince, M. Pagola, R. Mesiar, E. Hullermeier, and F. Herrera, "Grouping, overlap, and generalized bientropic functions for fuzzy modeling of pairwise comparisons," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 3, pp. 405-415, 2012.

[61] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*, ser. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, 2007, vol. 221.

[62] J. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," *Pattern Recognition*, vol. 45, no. 1, pp. 521-530, 2012.

[63] V. López, A. Fernández, and F. Herrera, "On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed," *Information Sciences*, vol. 257, pp. 1-13, 2014.

[64] J. Moreno-Torres, J. Saez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1304-1312, 2012.

[65] W. W. Cohen, "Fast effective rule induction," presented at the 12th Int. Conf. Mach. Learn., Lake Tahoe, CA, USA, 1995.

[66] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37-46, 1960.

[67] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, no. 6, pp. 80-83, 1945.

[68] J. L. Hodges and E. L. Lehmann, "Ranks methods for combination of independent experiments in analysis of variance," *Ann. Math. Statist.*, vol. 33, pp. 482-497, 1962.

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

Alberto Fernández received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 2005 and 2010, respectively. He is currently an Assistant Professor with the Department of Computer Science, University of Jaén, Spain. His research interests include data mining, classification in imbalanced domains, fuzzy rule learning, evolutionary algorithms, multiclassification problems with ensembles and decomposition techniques, and Data Science in Big Data applications. Dr. Fernández received the Lofti A. Zadeh Prize of the International Fuzzy Systems Association for the Best paper in 2009-2010 for his work of hierarchical fuzzy rule based classification system with genetic rule selection for imbalanced data-sets, and the 2012 University of Granada Prize for Scientific Excellence Works in the field of Engineering for the paper "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining".

Mikel Elkano received the M.Sc. degree in Computer Science from the Public University of Navarre, Spain, in 2014. He was awarded with the best Master's degree final project in the Artificial Intelligence Summer School 2014, organized by the Spanish Association for Artificial Intelligence. He is currently a Ph.D. student at the Public University of Navarre. His research interests include machine learning, fuzzy systems, ensemble learning, bioinformatics and Big Data.

Edurne Barrenechea is an Assistant Lecturer at the Department of Automatics and Computation, Public University of Navarre. She received an M.Sc. in Computer Science at the Pais Vasco University in 1990. She worked in a private company (Bombas Itur) as analyst programmer from 1990 to 2001, and then she joined the Public University of Navarre as Associate Lecturer. She obtained the Ph.D. in Computer Science in 2005 on the topic interval-valued fuzzy sets applied to image processing. Her publications comprise more than 30 papers in international journals and about 15 book chapters. Her research interests are fuzzy techniques for image processing, fuzzy sets theory, interval type-2 fuzzy sets theory and applications, decision making, and medical and industrial applications of soft computing techniques. She is member of the board of the European Society for Fuzzy Logic and Technology (EUSFLAT).

Mikel Galar received the M.Sc. and Ph.D. degrees in Computer Science in 2009 and 2012, both from the Public University of Navarre, Pamplona, Spain. He is currently an assistant professor in the Department of Automatics and Computation at the Public University of Navarre. His research interests are data-mining, classification, multi-classification, ensemble learning, evolutionary algorithms, fuzzy systems and fingerprint recognition.

Francisco Herrera received his M.Sc. in Mathematics in 1988 and Ph.D. in Mathematics in 1991, both from the University of Granada, Spain. He is currently a Professor in the Department of Computer Science and Artificial Intelligence at the University of Granada. He has been the supervisor of 35 Ph.D. students. He has published more than 290 papers in international journals. He currently acts as Editor in Chief of the international journals "Information Fusion" (Elsevier) and "Progress in Artificial Intelligence" (Springer). He acts as editorial board member of a dozen of journals, among others: International Journal of Computational Intelligence Systems, IEEE Transactions on Fuzzy Systems, Information Sciences, Knowledge and Information Systems, Fuzzy Sets and Systems, Applied Intelligence, Knowledge-Based Systems, 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 System 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.

José Antonio Sanz received the M.Sc. degree in computer sciences and the Ph.D. degree, both form the Public University of Navarre in 2008 and 2011 respectively. He is currently an Associate Lecturer with the Department of Automatics and Computation, Public University of Navarre. He is the author of 15 published original articles in international journals and is involved in teaching artificial intelligence for students of computer sciences. He received the best paper award in the FLINS 2012 international conference and the Pepe Millá award in 2014. His research interests include fuzzy techniques for classification problems, interval-valued fuzzy sets, genetic fuzzy systems, data mining and medical applications of soft computing techniques. Dr. Sanz is member of the European Society for Fuzzy Logic and Technology.

Humberto Bustince received his Bs. C. degree on Physics from the Salamanca University, Spain, in 1983 and his Ph.D. degree in Mathematics from the Public University of Navarre, Pamplona, Spain, in 1994. He has been a teacher at the Public University of Navarre since 1991, and he is currently a Full Professor with the Department of Automatics and Computation. He served as subdirector of the Technical School for Industrial Engineering and Telecommunications from 01/01/2003 to 30/10/2008 and he was involved in the implantation of Computer Science courses at the Public University of Navarre. He is currently involved in teaching artificial intelligence for students of computer sciences. Dr. Bustince has authored more than 120 journal papers (Web of Knowledge), and more than 100 contributions to international conferences. He has also been co-author of four books on fuzzy theory and extensions of fuzzy sets. He is fellow of the IEEE Computational Intelligence Systems society and Member of the board of the European Society for Fuzzy Logic and Applications (EUSFLAT). He currently acts as Editor in chief of the Mathware&Soft Computing Magazine and of Notes on Intuitionistic Fuzzy Sets. He is also guest editor of the Fuzzy Sets and Systems journal and member of the editorial board of the Journal of Intelligence&Fuzzy Systems, the International Journal of Computational Intelligence Systems and the Axioms Journal. His current research interests include interval-valued fuzzy sets, Atanassov's intuitionistic fuzzy sets, aggregation functions, implication operators, inclusion measures, image processing, decision making and approximate reasoning.