Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 465 (2023) 108511

FUZZY sets and systems

www.elsevier.com/locate/fss

Learning positive-negative rule-based fuzzy associative classifiers with a good trade-off between complexity and accuracy

Carmen Biedma-Rdguez [a], María José Gacto [b], Augusto Anguita-Ruiz [c], Rafael Alcalá [a], Concepción María Aguilera [d], Jesús Alcalá-Fdez [a,*]

[a] *Department of Computer Science and Artificial Intelligence, DaSCI, University of Granada, 18071 Granada, Spain*
[b] *Department of Software Engineering, DaSCI, University of Granada, 18071 Granada, Spain*
[c] *Barcelona Institute for Global Health, ISGlobal, 08003 Barcelona, Spain*
[d] *Department of Biochemistry and Molecular Biology II, ibs.GRANADA, University of Granada, 18071 Granada, Spain*

Received 28 September 2022; received in revised form 16 March 2023; accepted 24 March 2023
Available online 29 March 2023

Abstract

Nowadays, the call for transparency in Artificial Intelligence models is growing due to the need to understand how decisions derived from the methods are made when they ultimately affect human life and health. Fuzzy Rule-Based Classification Systems have been used successfully as they are models that are easily understood by models themselves. However, complex search spaces hinder the learning process, and in most cases, lead to problems of complexity (coverage and specificity). This problem directly affects the intention to use them to enable the user to analyze and understand the model. Because of this, we propose a fuzzy associative classification method to learn classifiers with an improved trade-off between accuracy and complexity. This method learns the most appropriate granularity of each variable to generate a set of simple fuzzy association rules with a reduced number of associations that consider positive and negative dependencies to be able to classify an instance depending on the presence or absence of certain items. The proposal also chooses the most interesting rules based on several interesting measures and finally performs a genetic rule selection and adjustment to reach the most suitable context of the selected rule set. The quality of our proposal has been analyzed using 23 real-world datasets, comparing them with other proposals by applying statistical analysis. Moreover, the study carried out on a real biomedical research problem of childhood obesity shows the improved trade-off between the accuracy and complexity of the models generated by our proposal.

Keywords: Fuzzy associative classification; Evolutionary fuzzy systems; eXplainable artificial intelligence; Transparency; Complexity

* Corresponding author.
E-mail addresses: biedmardguez@gmail.com (C. Biedma-Rdguez), mjgacto@ugr.es (M.J. Gacto), augusto.anguita@isglobal.org (A. Anguita-Ruiz), alcala@decsai.ugr.es (R. Alcalá), caguiler@ugr.es (C.M. Aguilera), jalcala@decsai.ugr.es (J. Alcalá-Fdez).

https://doi.org/10.1016/j.fss.2023.03.014

1. Introduction

In recent years, the application of Artificial Intelligence (AI) has experienced a growing interest in a large number of industry and research applications. Much of this interest is due to the high predictive capacity of models derived from different Machine Learning (ML) techniques, such as Deep Learning, in critical contexts [1]. Otherwise, most of these accurate models are very non-transparent, i.e., they still suffer from unreliable interpretability and explainability as it is not clear what information from the input data leads them to their decisions. This is known as the black box problem [2]. Nowadays, the call for transparency is growing among various AI stakeholders due to the need to understand how the decisions derived from methods are taken when they ultimately affect human life and health, as in for example medicine, law or defense, security and finance, among others. These models should be considered decision support systems and, as such, should be endowed with elements that allow their fairness, accountability, transparency and ethics to be taken into account [3,4]. This new trend gives rise to what it is known as eXplainable Artificial Intelligence (XAI) [5,6], which recommends the use of transparent models that are understandable in their own right [7], and/or post-hoc explainability techniques that aim to provide explanations of how a complex model produces its predictions for any given input [8].

Fuzzy Rule Based Classification Systems (FRBCSs) have been successfully used in many real-world applications, such as medical diagnosis [9] and environmental investment [10], among others. They are transparent models that can be understood in their own right, without the need to resort to explanatory techniques to understand them. Their natural relationship to human behavior makes them ideal for understanding and explaining complex systems. In addition, the use of fuzzy logic as a data handling tool provides some important advantages. In terms of semantics, the use of linguistic labels in the fuzzy model structure is a natural representation of knowledge that enables direct human interaction. From a learning perspective, translating the input features into fuzzy variables with fuzzy membership functions allow avoiding the use of unnatural limits in the partitions of the variable domain and yields smoothed descriptive models that adapt well to data with a certain degree of uncertainty [11].

One particular case of FRBCSs are fuzzy associative classifiers (FACs), in which use is made of association discovery techniques to obtain interesting and simple rules from large datasets that allow the end-user to easily understand and interpret the results [12]. These classifiers are built in two stages: first, discover the fuzzy association rules inherent in a database; second, select a small set of relevant rules to build a classifier. This approach yields efficient and accurate classifiers due to the ability of association rule mining methods to find all possible relationships that exist between input data during the rule generation process, generating all the hidden rules that may not have been detected by other classification algorithms [13].

However, these methods often generate a large number of rules (coverage) together with a large number of antecedents/consequents in the rules (specificity). An increase in the coverage and specificity of the model often means an improvement in its accuracy but compromises the interpretability of the model [14]. This problem directly affects the intention of using these models to enable the user to analyze and understand the model, since the transparency of the model decreases as its complexity increases, making it essential to find a good trade-off between accuracy and transparency [5]. In addition, most methods generate rules that only represent positive relationships between the antecedent and the consequent, without taking into account the negative relationships that may be present in them. Rules such as $\neg A \rightarrow C$, can also be interesting to build classifiers by relating the membership of class C to the absence of A. The use of fuzzy negated itemsets in the rule extraction process expands the type of relationships that can be represented, allowing interesting relationships to be represented with fewer rules than with the classical representation. [15,16].

One straightforward way to combine accuracy and transparency in a natural way to obtain XAI models is by means of the positive synergy between FRBCSs and Evolutionary Algorithms, known as Evolutionary Fuzzy Systems (EFSs). These search and optimization methods have been broadly and successfully used to learn and tune fuzzy systems [17]. In this sense, EFSs can preserve the original essence of the comprehensibility of fuzzy systems and enhance their predictive capability by being able to consider the transparency, comprehensibility and understandability of the model in the learning process [18,14]. For instance, in [19,20], the authors proposed an evolutionary learning of accurate and interpretable FACs for traditional and monotonic classification, in which the granularity is provided a priori and the rules follow a classic structure. However, the given granularity may have a critical influence on the final mining results and may result in a high number of rules. On the other hand, in [21], the authors designed a genetic programming algorithm to be as parallel as possible to efficiently buil associative classifiers in Big Data without

affecting their accuracy and interpretability. However, the models obtained have similar accuracy and complexity to those obtained in [19,20].

In this paper, we propose a new FAC method based on positive and negative rules, called PN-FAC, to carry out an evolutionary learning of classifiers with an improved trade-off between complexity and accuracy. To do so, our method consists of 3 stages: 1) genetic learning of the most suitable granularity for each fuzzy partition; 2) mining process of Positive and Negative Fuzzy Rules (PNFRs) by constraining the number of items included in them and rule pre-screening to obtain simple and interesting rules; 3) genetic rule selection and adjustment of the labels of the rules, considering the known positive synergy of both approaches to learn the most suitable context of the selected rule set. These stages focus the search process on potentially interesting and useful information for the model, obtaining a reduced set of simple rules with good precision.

The suitability of our proposal has been evaluated by means of two experimental studies. On one hand, we experimentally test our proposal on 23 real-world datasets with a number of variables ranging from 2 to 90, a number of examples ranging from 106 to 19020, and a number of classes from 2 to 15. We have compared the predictive capacity and complexity of the models obtained by our proposal with 3 associative classification methods as well as with 4 traditional ones, making use of nonparametric statistical tests to analyze the results obtained [22]. On the other hand, we analyze the transparency of our proposal in a real problem related to childhood obesity. Overweight and obesity in children are important risk factors for a number of chronic cardiometabolic alterations during adulthood such as hypertension, insulin resistance (IR), among others, which finally lead to the development of Metabolic Syndrome (MS), considerably increasing population morbimortality. To address this problem effectively, there is an urgent need to implement early-life predictive and understandable models that are capable of anticipate the appearance of MS, when there is still time for clinical intervention. We have used our proposal for the construction of a predictive model for pubertal MS in children with obesity making use of information from the prepubertal stage of children (ages 4-12), which consists of: anthropometry, biochemistry cardiometabolic and inflammatory biomarkers. The results and the model obtained have been analyzed by a real expert.

This contribution is organized as follows. Section 2 provides some relevant concepts of FRBCSs and positive and negative FACs. The different stages of our proposal to obtain accurate and transparent classifiers are detailed in Section 3. Section 4 evaluates the results obtained from the 23 real data sets and analyzes the model obtained from the available information to predict MS in a cohort of children with obesity. Finally, Section 5 summarizes some concluding remarks.

2. Preliminaries

Firstly, some basic concepts about FRBCS are introduced in this section. Then, some definitions of positive and negative association rules and several quality measures for classification problems are briefly introduced.

2.1. Fuzzy rule based classification systems

In a classification problem with D examples $e_d = (e_{d1}, ..., e_{dm}, y_d)$, where e_{di} is the i-th input variable value ($i = 1, 2, ..., m$) and y_d is the class label ($y_d \in \{C_1, \ldots, C_S\}$) of the d-th example, the objective is to learn a model that can accurately classify previously unknown examples. To this end, FRBCSs make use of a fuzzy rule set, called Rule Base (RB), with the following structure:

$$Rule\ R_j:\ If\ e_1\ is\ A_{j1}\ \text{and}\ \ldots\ \text{and}\ e_m\ is\ A_{jm}\ \text{then}\ Class\ =\ y_j\ \text{with}\ RW_j$$

where R_j is the j-th rule, $e = (e_1, ..., e_m)$ is an m-dimensional vector which represents the example values, A_{ji} is a fuzzy set defined over the domain of the i-th input variable, y_j is the class label of the j-th rule, and RW_j is the rule weight. Due to their significant influence on FRBCS performance, several rule weight specifications are available in the literature [23,24]. We have used the fuzzy confidence or Certainty Factor (CF), which is the most common specification. This specification is defined as [25]:

$$RW_j = CF_j = \frac{\sum_{e_d \in Class\ y_j} \mu_{A_j}(e_d)}{\sum_{d=1}^{D} \mu_{A_j}(e_d)}$$

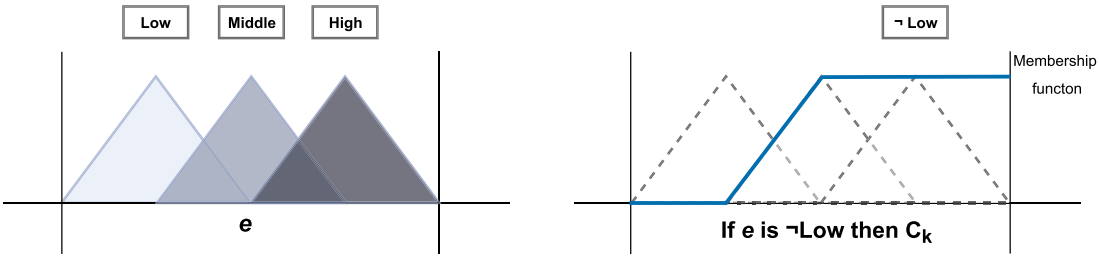

Fig. 1. A simple example of a fuzzy associative rule involving negated items.

where $\mu_{A_j}(e_d) = T(\mu_{A_{j1}}(e_{d1}), \ldots, \mu_{A_{jm}}(e_{dm})))$ is the matching degree of the example e_d with the antecedent of R_j. In order to classify a new example $e_d = (e_{d1}, ..., e_{dm})$, we have used the fuzzy reasoning method called additive combination [25]. This method calculates a confidence degree (CD) for each class label and classifies the example e_d with the label with the maximum degree. The CD of a class label C_k $(k = 1, 2, ..., S)$ is calculated by adding the product of the matching degree and the weight of each rule in the RB with class label C_k:

$$CD_{C_k}(e_d) = \sum_{R_{jy_j=C_k} \in RB} \mu_{A_j}(e_d) \cdot RW_j$$

The performance of classifiers is usually evaluated using the accuracy ratio. However, accuracy is not the most appropriate measure when there are significant differences in the proportion of examples belonging to each class of the problem, known as imbalanced datasets, because it does not distinguish between the number of correctly classified examples from the different classes. Thus, the predominance of the most frequent class (known as the negative class) may mask poor classification performance in the infrequent class (known as the positive class). To alleviate this problem, several researchers have proposed other measures that can be used to evaluate classification performance in the presence of class imbalance [26]. In this work, we have also used the Kappa [27] and G-mean [28] measures account for class imbalance of the classes in the analysis of the experimentation carried out. These measures are defined as follows in a classification problem with two classes (positive and negative):

$$Kappa = \frac{Acc_o - Acc_e}{1 - Acc_e} \tag{1}$$

$$G - mean = \sqrt{\frac{TP}{TP + FN} \frac{TN}{TN + FP}} \tag{2}$$

where Acc_e is the expected accuracy ratio, Acc_o is the obtained accuracy ratio, TP and TN indicate the correctly classified examples of the positive and negative classes, and FN and FP indicate the misclassified examples of the positive and negative classes. These measures have been calculated following a macro approach on multiclass datasets.

2.2. Positive and negative fuzzy association rules for classification

Fuzzy association rules are expressions of type $A \rightarrow B$ which express frequent dependencies that exist between fuzzy items (pairs $\{variable, label\}$) in a high-confidence dataset [29,30]. These rules can be used as classification rules if in the consequent part they contain only one of the class labels C_k $(k = 1, 2, ..., S)$ of the problem, so many approaches have made use of association discovery techniques to utilize their abilities for classification [12].

Most of the mining methods focus on discovering positive dependencies between fuzzy items in the dataset, however, negative dependencies can also be of interest as they relate the absence of fuzzy items to the class label they belong to. Negative rules, such as $\neg A \rightarrow C_k$, not only consider the fuzzy items included in the positive rules but may also include negated fuzzy items (couples $\{variable, \neg label\}$) in the antecedent of the rules. Fig. 1 shows a simple example of a negated rule and the membership function of the negated item involved in the rule.

As with fuzzy association rules, the measures commonly used to assess fuzzy associative classification rules are support and confidence. They are based on the support of a fuzzy itemset FI, which is calculated as follows:

$$Support(FI) = \frac{\sum_{e_d \in D} \mu_{FI}(e_d)}{\mid D \mid} \tag{3}$$

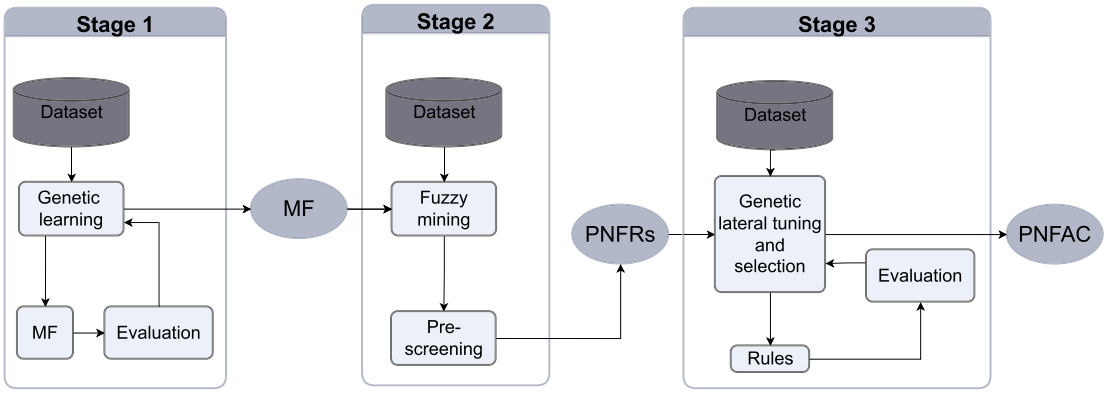

Fig. 2. Workflow of PN-FAC.

where $\mu_I(e_d)$ is the matching degree of the example e_d with the fuzzy itemset FI, and $\mid D \mid$ is the quantity of training examples. Thus, given a rule $A \rightarrow C_k$, the support and confidence are calculated as follows:

$$Support(A \rightarrow C_k) = Support(AC_k) \tag{4}$$

$$Confidence(A \rightarrow C_k) = \frac{Support(AC_k)}{Support(A)} \tag{5}$$

Classical methods try to discover fuzzy itemsets whose support is greater than a minimum support threshold ($minsup$), known as frequent fuzzy itemsets, and then generate rules whose confidence is greater than a minimum confidence threshold ($minconf$). Nevertheless, a number of studies have highlighted various problems associated with these measures of rule assessment [31]. As result, several researchers have proposed other measures to assess their potential interest to the user [32]. In this work, the wWRAcc measure [33] has been extended to choose the most interesting rule subset from the extracted rule set.

3. Positive and negative fuzzy associative classifier: PN-FAC

We present our proposal for performing an evolutionary learning of FAC with a good trade-off between accuracy and complexity. In the following, we will explain each of the stages of which it is composed in detail. Fig. 2 shows a workflow for the proposed method.

3.1. Genetic learning of the granularity

Since the given granularity significantly influences the FRBCS performance, we perform granularity learning by considering strongly equidistributed fuzzy partitioning with triangular membership functions (MFs). This learning is carried out using the CHC genetic algorithm [34], which is a good choice for problems with complex search spaces as it offers a good balance between exploitation and exploration in the search process. To encourage convergence in the search process, CHC uses a population-based selection mechanism that picks out the P best individuals from the current population and their offspring at each iteration to take part in the next population. To introduce diversity in the search, rather than a mutation operator this algorithm considers an incest prevention mechanism and a restart process. Two individuals are crossed only if they are different to a degree above a predefined threshold value L, which is decremented each time a new individual is not included in the next population. On the other hand, the population is restarted when L becomes zero. Fig. 3 shows a workflow of the evolutionary model of the CHC algorithm.

Population chromosomes are vectors of integer values that encode the number of labels for each of the m input variables of the problem $(c_1, \ldots, c_m)$. Each gene takes values in the set $\{2, \ldots, 7\}$ and, additionally, it may take the value 0 to indicate that the corresponding variable is not used. In order to provide diversity, the initial population consists of two different subsets of individuals. In the first subset, each chromosome has the same number of labels for all the variables and contains one chromosome for each of the values in the set $\{2, \ldots, 7\}$. In the second one, the rest of the population is initialized randomly. In order to encourage the selection of interesting variables and to avoid

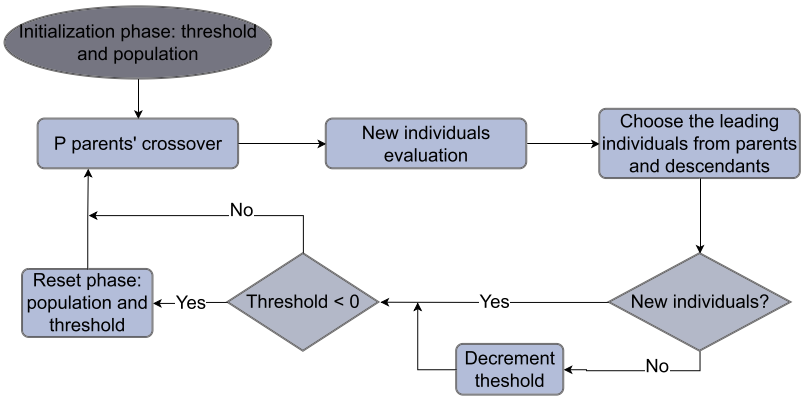

Fig. 3. Workflow of the evolutionary model of the CHC algorithm.

the generation of a large number of rules in the extraction process, in problems with more than 5 input variables the number of used variables is randomly reduced to 5 for half of the individuals of the second subset ($c_i = 0$ with i selected at random).

To evaluate each chromosome, the encoded granularity for each variable is used to generate the uniform fuzzy partitions from which a RB is generated by using the same rule mining process as in the next stage. The classification accuracy of the generated RB will be considered the fitness value of the chromosome.

The offspring population is generated by applying the HUX crossover operator [35], which randomly exchanges half of the genes that are different in the parents. This operator maximizes the distance to the parents, encouraging exploration in the search process. Note that this operator will only be applied if half of the hamming distance between the parents is greater than threshold L.

To move away from local optima, the restarting process reinitializes the population when the threshold L becomes zero. In this case, the best chromosome is maintained in the population and the rest of the individuals are initialized as in the second subset of the initial population. The value of L will also be reset.

3.2. *Fuzzy rule mining process and prescreening*

Once we have the equidistributed strong fuzzy partitions derived from the granularity learned for each variable in the previous phase, positive and negative association rules are extracted by extending the frequent fuzzy itemset extraction method proposed in [19] to also consider negated fuzzy items in the search process. This method uses a search tree to explore all fuzzy itemsets that are frequent for each class. All variables assume the order of appearance in the dataset and all fuzzy 1-itemsets corresponding to each variable along with their respective negated fuzzy 1-itemsets are placed in the first level of the tree according to their order. The children of a fuzzy 1-itemset node of a variable are the 2-itemsets that are obtained by combining it with the fuzzy 1-itemsets of the variables that are larger in the predefined order, and so on.

If the n-itemset of a node is not frequent, it will no longer be combined because any itemset generated from it will also be infrequent. Note that the *minsup* is adjusted to the frequency of examples that the class has in the training set. In the same way, if an itemset produces a classification rule with a confidence higher than the maximum confidence required by the user, it is not necessary to continue combining it because it has reached the required quality. The tree depth is also limited to three in order to avoid generating very specific itemsets that provide complex rules. These properties, together with the variables removed in the previous phase ($c_i = 0$), make it possible to substantially reduce the number of nodes generated during the search. Once the frequent fuzzy itemsets have been extracted, positive and negative classification rules can be generated with higher confidence than $minconf$ by setting the itemset in the antecedent part of the rule and the corresponding class in the consequent. This process is repeated with each of the problem's classes.

However, this method may provide a large number of rules and may find many misleading rules due to the drawbacks of support and confidence measures. Therefore, we have selected the most interesting subset of rules following an example weighting scheme. In this scheme, each example has an associated weight which is used to calculate the

evaluation criteria of the rules. This weight is reduced each time we select a rule that covers it ($w(e_d, nr) = \frac{1}{nr+1}$, where nr represents the number of selected rules covering it), encouraging the selection of rules that cover examples with higher weights. The weight becomes zero when more than T selected rules cover the example. At each iteration, we sort the unselected rules according to their value for the evaluation criterion, select the best rule, and adjust the weights of the covered examples. This process is repeated until all the examples have a weight of zero.

We have used the wWRAcc measure as an evaluation criterion to select the fuzzy rules [33,19]. This measure is defined as follows:

$$wWRAcc(A \to C_k) = \frac{n''(A \cdot C_k)}{n'(C_k)} \cdot (\frac{n''(A \cdot C_k)}{n''(A)} - \frac{n(C_k)}{\mid D \mid}) \tag{6}$$

where $\mid D \mid$ is the number of training examples, $n''(A)$ is the sum of the products of the matching degree of the examples with the antecedent of the rule multiplied by their weights, $n''(A \cdot C_k)$ is the sum of the products of the matching degree of the examples with the rule multiplied by their weights, $n'(C_k)$ is the sum of the example weights of the class C_k and $n'(C_k)$ is the number of examples of the class C_k. However, this measure could potentially value rules with low and high confidence equally. In this measure, the first term refers to the proportion of uncovered examples of class C_k covered by the rule, and the second term refers to the difference between the confidence of the rule and the confidence of the default rule of class C_k. Let us consider two rules, R_1 and $R2$, which cover a proportion of uncovered examples of class C_k of 0.7 and 0.4 (both have a good coverage score) and have a confidence score of 0.65 and 0.9 respectively, and the default rule has a confidence score of 0.2. In this case, the R_1 rule has a higher value for the wWRACC measure (0.31) than the R_2 rule (0.28), so R_1 (with a confidence score of 0.65) would be selected before R_2 (with a confidence score of 0.9).

In this work, we have modified this measure to obtain a weighted average of both terms in order to encourage the confidence of the rule against its support, avoiding the selection of low quality rules. The new measure is defined as follows:

$$wWRAcc'(A \to C_k) = (1-\lambda)\frac{n''(A \cdot C_k)}{n'(C_k)} + \lambda(\frac{n''(A \cdot C_k)}{n''(A)} - \frac{n(C_k)}{\mid D \mid}) \tag{7}$$

where $lambda$ allows the expert to determine the trade-off between support and confidence in the rule prescreening process (where $lambda = 0.7$ represents a good performance).

3.3. Genetic rule selection and tuning of the labels of the rules

From the interesting rules selected in the previous phase, we selected a reduced subset of rules and tune the context of their MFs in order to improve classification accuracy. To this end, we have made use of the evolutionary method proposed in [36], which also employs the CHC genetic algorithm (more information on the CHC evolutionary model can be found in Subsection 3.1) to perform the selection and lateral tuning of the rule labels based on the representation of the linguistic 2-tuples. This representation reduces the search space by using a single parameter (α) to represent a small offset of the labels between their 2 lateral labels in the initial fuzzy partition, maintaining their original shapes and making the rule selection within the same process easier to handle. This parameter takes values in the interval [-0.5, 0.5), where -0.5 represents the lateral offset of the MF associated to the label up to half the distance between the MF core of the label and the MF core of the label to its left (offsets are to the right with positive values). Fig. 4 shows an example of a rule based on the representation of linguistic 2-tuples.

In our proposal, the domain of the α parameter has been reduced to the interval $[-0.25, 0.25)$ to obtain small label offsets, avoiding a significant loss of label interpretability. In addition, from these offsets we can calculate the equivalent MFs and generate a FRBCS based on classical Mamdani rules.

To learn these offsets and select the rules, each chromosome in the population is composed of two vectors. On the one hand, there is a binary vector to encode the selected rules, where a value of 1 indicates that the rule is selected and 0 that it is not. On the other hand, there is a real vector to encode the offset of each of the labels involved in each rule, taking values in the interval $[-0.25, 0.25)$. To provide diversity to the search, half of the chromosomes are initialized with all the selected rules and the randomly generated offsets, and in the rest of the population both offsets and rules are randomly selected. A chromosome with all selected rules and all offsets at 0 will also be included in the initial population to consider the initial set of rules in the search process.

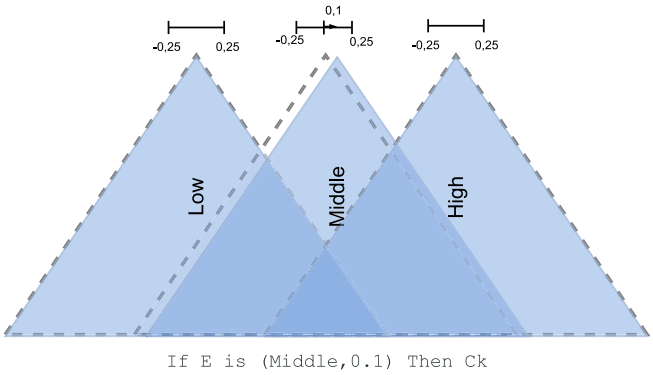

Fig. 4. Example of fuzzy rule based on the representation of lingüistic 2-tuples.

Each chromosome is evaluated considering its classification ratio penalized by the number of selected rules. Therefore, the following function is maximized:

$$Fitness = \frac{Hits}{D} - \delta \cdot \frac{PNFARs_{initial}}{PNFARs_{initial} - PNFARs + 1.0}, \tag{8}$$

where $Hits$ is the number of correctly classified examples, D is the number of examples, $PNFARs_{initial}$ is the number of rules in the initial set of rules, $PNFARs$ is the number of rules selected, and $delta$ is a weight predefined by the expert that allows them to determine the trade-off between the accuracy and complexity of the PNFAC.

We have used a different crossover operator for each vector of the chromosome. In the binary vector, we have used the HUX operator to generate two offspring by exchanging half of the genes that are different in the parents. In the real vector, the Parent Centric BLX (PCBLX) operator [37] was used to generate two offspring in which each gene is randomly selected within a small interval centered on each gene of each parent. Combining the results obtained from both operators we can generate four offspring, but only the best two will be considered to be part of the next population.

To avoid premature convergence, the incest prevention mechanism will only allow parents to cross over if their hamming distance divided by 2 is greater than the threshold L (initially predefined by the expert), which will be decremented each time the next generation does not include a new offspring. When L becomes zero, the population will be reinitialized, making sure that the best solution in the population is still the most accurate. Half of the population will be reset based on the best solution, selecting the same rules and generating offsets that are similar to those of the best solution, and in the rest of the population both rules and offsets are randomly generated.

4. Experimentation

The performance of our proposal has been analyzed by carrying out a double experimental study:

- We have analyzed the predictive capacity and complexity of the models obtained by our proposal with respect to other methods in the literature by comparing the results obtained for several measures on 23 real problems (subsection 4.1).
- We have studied the transparency of the model obtained by PN-FAC for a real biomedical research problem related to childhood obesity, studying the relationships identified from the prepubertal stage information to predict pubertal MS in children with obesity in a longitudinal cohort (Subsection 4.2).

4.1. Accuracy and complexity analysis with respect to other methods.

In this study, 23 problems were used with a number of variables ranging from 2 to 90, a number of examples ranging from 106 to 19,020, and a number of classes ranging from 2 to 15. Table 1 summarizes their main properties, indicating the number of input variables ($\#Var$), the number of examples ($\#Exams$), and the number of classes ($Classes$). Note that the Cleveland, Crx, and Wisconsin datasets include examples with missing values, which have

Table 1
Number of variables, examples and classes of datasets used in the experimental study.

Name	#Var	#Exams	Classes	Name	#Var	#Exams	Classes
Appendicitis	7	106	2	Pen-based	16	10,992	10
Banana	2	5,300	2	Phoneme	5	5,404	2
Bupa	6	345	2	Pima	8	768	2
Cleveland	13	297	5	Sonar	60	208	2
Crx	15	653	2	Spectfheart	44	267	2
Ecoli	7	336	8	Texture	40	5,500	11
German	20	1,000	2	Twonorm	20	7,400	2
Heart	13	270	2	Vehicle	18	846	4
Iris	4	150	3	Vowel	13	990	11
Newthyroid	5	215	3	Wine	13	178	3
Magic	10	19,020	2	Wisconsin	9	683	2
MovementLibras	90	360	15				

Table 2
Parameters considered for the comparison.

Algorithms	*Parameters*
Method	Parameters
CMAR	$Minconf = 0.5$, $Minsup = 0.01$, $DifferenceThreshold = 20\%$, $\delta = 4$
CPAR	$min\_gain = 0.7$, $\delta = 0.05$, $k = 5$, $\alpha = 0.66$
FARC-HD	$Minsup = 0.05$, $Maxconf = 0.8$, $Depth_{max} = 3$, $k_t = 2$, $Pop = 50$, $Evaluations = 15,000$, $BITSGENE = 30$, $\delta = 0.15$
C4.5	$Pruned = yes$, $Confidence = 0.25$, $InstancesPerLeaf = 2$
SLAVE2	$Pop = 100$, $Iter_{change} = 500$, $P_c = 0.6$, $P_m = 0.01$, $\kappa_1 = 0$, $\kappa_2 = 1$, $Op_{or} = 0.1$, $Op_{and} = 0.1$, $P_{gen} = 0.1$, $P_{rot} = 0.1$, $\alpha = 0.15$
SGERD	$Q = heuristics$
FURIA	$n\_folds = 3$, $optimizations = 2$
PN-FAC	$Minsup = 0.05$, $Maxconf = 0.8$, $T = 2$, $Pop = 50$, $Eval.G = 750$, $Eval.ST. = 10,000$, $BITSGENE = 30$, $\delta = 0.15$

been removed for this study. All these datasets can be downloaded from the KEEL-Dataset public repository[1] [38]. In this study we have used a cross-validation mechanism with 5 stratified folds. Each method has been run 3 times on each fold, and therefore has been run 15 times on each dataset.

Our proposal has been compared with 7 different methods from the literature that are available in the KEEL software tool [39]: 3 associative classification methods (CMAR [40], CPAR [41], and FARC-HD [19]) and 4 traditional rule-based classification methods (C4.5Rules [42], SLAVE2 [43], SGERD [44], and FURIA [45]). The parameters used for each method are shown in Table 2. The parameters of our proposal have been experimentally adjusted so that they can be recommended for use in future work. The rest of the method parameters have been selected following the recommendations indicated by their authors when they were published, which are the default parameters in KEEL. The CMAR and CPAR methods cannot be executed on datasets with continuous variables, so the Entropy method [46] has been previously applied to discretize them. For the fuzzy methods, the initial linguistic partition is composed of 5 fuzzy labels with uniformly distributed triangular MFs, with the exception of the method SGERD that uses uniform fuzzy partitions with different granularities (with 2, 3, 4 and 5 labels respectively).

To analyze the accuracy of our proposal, the tables of results show the mean results (the standard deviation is shown in parentheses) of the means obtained from the 15 runs performed by each method on each dataset, including the values for the measures of accuracy, G-mean and Kappa in order to consider possible imbalances between the classes in the datasets (see Subsection 2.1). The mean results obtained on each dataset can be found on the web page associated with the paper at https://sci2s.ugr.es/PN-FAC. To perform the statistical analysis of the results, the Friedman test was used to refute the equality hypothesis among the results obtained, thus obtaining a ranking of the methods for each

[1] https://sci2s.ugr.es/keel/datasets.php.

Table 3
Average results obtained by traditional classification methods on the measures Acc, G-mean, and Kappa.

	Methods				
Measures	C4.5Rules	SLAVE2	SGERD	FURIA	PN-FAC
Acc (σ)	0.793 (0.12)	0.798 (0.12)	0.708 (0.15)	0.829 (0.11)	**0.833** (0.11)
G-mean (σ)	0.769 (0.17)	0.747 (0.17)	0.577 (0.21)	0.780 (0.19)	**0.822** (0.14)
Kappa (σ)	0.611 (0.23)	0.594 (0.26)	0.459 (0.24)	0.654 (0.24)	**0.690** (0.23)

Table 4
Statistical test results for traditional classification methods on the measures Acc, G-mean, and Kappa. Friedman rank and adjusted p-values for Holm's test.

Method	Ranking Acc	Ranking G-mean	Ranking Kappa	Apv Acc	Apv G-mean	Apv Kappa
C4.5Rules	3.4	3.1	3.2	0.001	0.003	0.002
SLAVE2	3.1	3.1	3.3	0.007	0.003	<0.001
SGERD	4.6	4.8	4.7	<0.001	<0.001	<0.001
FURIA	2.1	2.4	2.2	0.608	0.093	0.208
PN-FAC	**1.8**	**1.6**	**1.6**	-	-	-

Table 5
Average results obtained by classical rule-based associative classification methods on the measures Acc, G-mean, and Kappa.

	Methods			
Measures	CMAR	CPAR	FARC-HD	PN-FAC
Acc (σ)	0.754 (0.18)	0.78 (0.16)	0.832 (0.12)	**0.833** (0.11)
G-mean (σ)	0.664 (0.23)	0.661 (0.20)	0.801 (0.17)	**0.822** (0.14)
Kappa (σ)	0.570 (0.25)	0.595 (0.25)	0.679 (0.24)	**0.691** (0.22)

measure. Once the equality hypothesis was rejected, the Holm test was used to perform the multiple comparisons [47]. For each comparison, we have calculated the adjusted p-value (Apv) for which the equality hypothesis is rejected.

Table 3 shows the results obtained by the traditional classification methods, where σ is the standard deviation, Acc is the mean value obtained for the accuracy measure, $Gmean$ is the mean value obtained for the Gmean measure and $Kappa$ is the mean value obtained for the Kappa measure. The statistical analysis is shown in Table 4, including the Friedman ranking for each measure and the Apv for comparisons between the method with the best Friedman ranking and the other methods. PN-FAC obtains the best Friedman ranking in all measures and the $Apvs$ are less than 0.1 in all comparisons except for Acc and Kappa in the comparison with FURIA, rejecting the equality hypothesis with greater than 90% confidence. Although statistically there are no significant differences with FURIA for the Acc measure, we can see how the G-Mean and Kappa measures show a better performance of our proposal when taking into account the balancing of the classes in the problems. These results highlight the good performance of our proposal when compared to these methods.

The results of the comparison of our proposal with the associative classification methods are shown in Tables 5 and 6 (these tables follow a similar structure to Tables 3 and 4). As in the previous comparisons, PN-FAC obtains the best Friedman ranking in all measures and the $Apvs$ obtained are less than or equal to 0.05 except for Acc and G-mean in the comparison with FARC-HD, rejecting the equality hypothesis with greater than 95% confidence. Again, our proposal shows no statistical differences with FARC-HD for the Acc and G-mean measures, but for the Kappa measure also show better performance again when class balancing is taken into account.

In order to compare the complexity of the models obtained with our proposal, we have analyzed the number of rules and the number of variables included in the antecedent of the rules generated by our proposal with respect to those obtained by the methods that have shown better performance in the accuracy analysis (FURIA and FARC-HD).

Table 6
Statistical test results for classical rule-based associative classification methods on the measures Acc, G-mean, and Kappa. Friedman rank and adjusted p-values for Holm's test.

Method	Ranking Acc	Ranking G-mean	Ranking Kappa	APV Acc	APV G-mean	APV Kappa
CMAR	3.0	3.2	3.2	0.018	<0.001	<0.001
CPAR	2.8	3.5	3.1	0.033	<0.001	<0.001
FARC-HD	2.3	1.9	2.4	0.253	0.120	0.050
PN-FAC	**1.9**	**1.4**	**1.5**	-	-	-

Table 7
Complexity of the methods with the best accuracy in the study on the 23 datasets: average number of rules, average number of variables involved, product between the mean number of rules and variables included.

	Methods		
Measures	FURIA	FARC-HD	PN-FAC
#Rules (σ)	28.33 (31.28)	39.95 (37.45)	28.91 (33.02)
#VarInc (σ)	5.01 (1.57)	2.36 (0.55)	2.57 (0.47)
Complex (σ)	170.67 (238.31)	105.60 (104.88)	82.78 (100.28)

Table 8
Statistical test results for the Complex measure. Friedman rank and adjusted p-values for the Holm's test.

Method	Ranking Complex	APV Complex
FURIA	2.3	0.010
FARC-HD	2.2	0.012
PN-FAC	**1.4**	-

To evaluate the complexity of the models obtained, we have used the product of the number of rules obtained and the number of variables involved in the rules as a measure of complexity ($Complex$).

Table 7 shows the mean values of the averages obtained from the 15 runs performed by each method on each dataset, where $\#Rules$ is the mean number of rules included in the model, $\#VarInc$ is the mean number of variables included in the rules generated, and $Complex$ is the complexity measure. Table 8 shows the results of the statistical analysis on the complex measure, which has been normalized by dividing it by the value of the measure of the most complex model generated. At first glance, we can observe how PN-FAC and FARC-HD are able to obtain models involving rules with few variables in the antecedent but PN-FAC is able to generate simpler models, including on average 25% fewer rules. On the other hand, PN-FAC and FURIA generate models involving a similar number of rules, but FURIA generates more complex rules that include on average twice as many variables as the rules generated by PN-FAC. Statistical results show that our proposal obtains the best Friedman ranking with Apv lower than 0.05, rejecting the equality hypothesis with a confidence higher than 95%.

In order to determine how much the MFs have changed from the originals by tuning their context and to detect possible inconsistent rules in the model, we have made use of the interpretability measures GM3M and Rule Mean Index (RMI) [48] to analyze the models obtained by PN-FAC and FARC-HD. Note that FURIA has no analyzed due to the fact that it is an approximate method that generates a specific label for each variable of each rule without using an initial uniform partition of reference with which we can determine when the MFs have been modified.

Table 9 shows the mean values obtained by the FARC-HD and PN-FAC methods for the RMI and GM3M interpretability measures. Analyzing the results shown in Table 9, we can see how the models obtained by FARC-HD and PN-FAC reach high average values for the GM3M measure so that only small offsets are performed on the MFs, avoiding a significant loss of interpretability of the original label. On the other hand, the values for RMI show how

Table 9
Results obtained for GM3M and RMI interpretability measures for the FARC-HD and PN-FAC methods.

	Methods	
Measures	FARC-HD	PN-FAC
RMI (σ)	0.59 (0.20)	**0.63 (0.23)**
GM3M (σ)	**0.82 (0.02)**	0.81 (0.01)

these small offsets enable a proper adjustment of the context of the MFs associated with the labels, improving the consistency of the rules generated with respect to FARC-HD.

4.2. Analysis of the biological relevance of the model obtained to predict insuline resistance in children with prepubertal obesity.

To analyze the transparency of the method in the biomedical domain, we present an study showcase in a multi-center cohort of Spanish children with obesity. Particularly, we used data derived from the project "Puberty and metabolic risk in obese children - Epigenetic alterations and pathophysiological and diagnostic implications" (PUBMEP). This project is a longitudinal study based on the follow-up of a cohort of Spanish children who previously participated in the GENOBOX study [49,50]. The main objective of the project was to unveil the molecular mechanisms behind the appearance of obesity and cardiometabolic complications from the early stages of life. In the PUBMEP study, prepubertal boys and girls initially enrolled in the GENOBOX study who had already initiated puberty at the time of the PUBMEP study start were invited to participate. During the PUBMEP study (2012–2018), children remained under regular medical monitoring by the same paediatricians. MS is a dysfunctional situation that frequently appears in obese patients affecting cardiometabolic parameters such as blood pressure, lipid and glucose metabolism, increasing the risk of cardiovascular disease and preceding the development of type 2 Diabetes. The early life identification of children with obesity at risk for this condition, as well as the understanding of the external factors and biological mechanisms behind it, is key to implementing preventive interventions and developing new therapies. By applying our method to this dataset, we aimed to predict the presence or absence of MS in these children.

A total of 374 subjects were contacted in the PUBMEP study, of which 49 were not located, 36 could not participate because they had changed their place of residence or met any of the exclusion criteria, and 97 declined the invitation. Finally, 192 children were accepted to participate and were included in the analyzes. This dataset presents the typical challenges of biological databases, including a longitudinal design, high-dimensionality, high-correlation between features and strong variability between subjects. The dataset was composed of 58 variables including anthropometric and clinical measurements (such as body mass index, waist circumference (WC) measurements, systolic and diastolic blood pressure (SBP and DBP) measurements, or city of origin), biochemical analyzes (mainly lipid and glucose metabolism biomarkers such as glucose, insulin-resistance index (HOMA-IR), triglycerides (TG) and cholesterol fasting blood levels, or sexual hormone concentrations), proteomic data (plasma adipokines, inflammation, and cardiovascular risk biomarkers), and physical activity (measured through accelerometry; ActiGraph GT3X+ accelerometers), all of them, parameters related to the metabolic health status of the children. These data were available both at the prepubertal stage (T0), and at the pubertal stage (T1), although in our showcase only input data from the prepubertal stage (T0) were employed.

The class to predict in our use case was the MS status of the children (healthy or unhealthy), which was defined at the pubertal stage (T1) according to the next criteria; WC $\geq$ 90th percentile Ferrandez growth-chart [51], SBP or DBP $\geq$ 90th percentile [52,53], TG $\geq$ 90th percentile [54], HDL-cholesterol $\leq$ 10th percentile [54], and HOMA-IR above 3.38 in pubertal boys and 3.905 in pubertal girls [55]. A child was classified as unhealthy if presenting altered WC plus two other altered metabolic compounds among the remaining four. Among the 192 subjects, we had 163 healthy and 29 unhealthy children.

Before applying our method, missing data imputation was performed using the missforest approach as described elsewhere [56]. The experimental set-up followed in our use case involved using prepubertal data (T0) for the 58 input variables to predict the MS status of the children at the pubertal stage (T1), which was measured 5 years later. From the application of our proposal, a predictive model was generated with a good ability to identify both healthy and

unhealthy classes (Acc: 0.97, G-mean: 0.92, Kappa: 0.88). With only 15 rules and a mean of 2.8 variables involved in the antecedent of the rules, this model constituted a concise and easily interpretable simplification of the complex scenario of MS and obesity that we were facing.

The method selected important input prepubertal (T0) variables as predictors of the MS status at the pubertal stage (T1), including anthropometric measurements (Hip-Circumference or BMI Z-Score), environmental data (the city of origin) or biochemical biomarkers (lipid profile biomarkers such as TG and glucose metabolism biomarkers as HOMA-IR and glucose plasmatic levels), which participate in the definition of the MS class and it has been demonstrated that one of the highest risk factors for presenting MS at the pubertal stage is to have presented obesity and MS during early childhood [57]. Interestingly, the method also highlighted other important biomarkers of inflammation and cardiovascular disturbances (including pro-inflammatory cytokines (TNF-a, adiponectin) or coagulation and adhesion molecules (aPAI or sICAM)). This finding is interesting since inflammation is one of the main mediators between obesity and metabolic disease; the inability of adipose tissue to expand properly provokes the infiltration of inflammatory cells into the tissue and implies the release of pro-inflammatory biomarkers into the bloodstream, favoring a systemic inflammatory status in the body and the development of insulin-resistance [58]. Furthermore, the model pointed to the dysregulation of hepatic enzymes like AST, preceding the development of MS. This makes sense given the known association between NAFLD (non-alcoholic fatty liver disease) and metabolic disturbances.

Although the model is composed of 15 rules, a clear separation is made with the first rule, which affects 65% of the population (support: 0.65 and CF: 0.91). This rule establishes that children with low-to-medium hip-circumference values (54 to 93.4 cms) at the prepubertal stage probably will be healthy at the pubertal stage. This is in line with the well-known role of abdominal (visceral) obesity as a predominant risk factor for MS [59]. The rest of the rules refer to specific situations (with low support) typical of MS, referred to inflammatory and cardiovascular biomarkers (high lipid profiles, insulin resistance, or inflammation). This reflects the high heterogeneity that is gathered under the umbrella of MS definition, and how the course of events leading to its appearance can be different in every patient. The fuzzy partitions defined for each variable and the rules generated can be found on the web page associated with the paper at https://sci2s.ugr.es/PN-FAC.

5. Conclusions

In this paper, we propose a new method based on three stages to obtain associative classifiers consisting of a few simple positive and negative fuzzy rules with good classification accuracy. For this purpose, we carried out three stages: first, we performed an evolutionary learning of the most appropriate granularity for each variable to define an uniform fuzzy partition with triangular MFs over its domain; second, we applied an extraction process to obtain interesting positive and negative rules involving a low number of variables for each class; and third, we made use of an evolutionary method to select a reduced subset of rules and tune the context of their MFs.

The results obtained from the experimental analysis on 23 datasets show how our proposal allows us to obtain classifiers with a good performance for different quality measures, obtaining statistically similar or better results than the rest of the analyzed methods. The obtained classifiers consist on average of a reduced sets of rules that involve few variables. In addition, their MFs only present small lateral offsets that allow an easy understanding of the rules from an expert's point of view, providing classifiers with a good trade-off between complexity and accuracy.

The suitability of our proposal has been evaluated by means of two experimental studies. On one hand, we experimentally test our proposal on 23 real-world datasets with a number of variables ranging from 2 to 90, a number of examples ranging from 106 to 19020, and a number of classes from 2 to 15.

The classifier obtained on a real-life longitudinal biomedical dataset presents a good predictive ability considering the high difficulty behind giving 5-year forecasts in humans. At the time, it favors a proper interpretation of how decisions are taken, opening a horizon for its application in the biomedical field not only with predictive/classification intentions but also for the discovery-understanding of the molecular alterations leading to disease.

Finally, the learning of the most appropriate granularity for each variable, the generation of simple rules involving few variables using a rule representation that allow to represent certain interesting relationships with fewer rules than with the classical representation, and the selection of interesting rules based on the measure $wWRAcc'$, allow to considerably reduce the search space and to handle datasets with different dimensionality (variables, examples and classes). Moreover, this proposal could be extended to handle large scale datasets (Big Data) following the framework based on the MapReduce paradigm and using Apache Spark proposed in [60], in which different MapReduce phases

are used to obtain the rules from different splits of data applying any algorithm to discover association rules (such as Fuzzy FP-growth [61]) and to evaluate them using all the splits of the entire input dataset.

Declaration of competing interest

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

Data availability

The authors have used benchmark datasets available in public repositories and data from a real biomedical problem for which the authors do not have permission to share the data.

Acknowledgement

This work was supported in part by the ERDF, by the Regional Government of Andalusia/Ministry of Economic Transformation, Industry, Knowledge and Universities (grant numbers P18-RT-2248 and B-CTS-536-UGR20), by the Health Institute Carlos III/Spanish Ministry of Science, Innovation and Universities (grant number PI20/00711), and by the Spanish Ministry of Science and Innovation (grant number PID2019-107793GB-I00). Funding for open access charge: Universidad de Granada / CBUA.

References

[1] M. Attaran, P. Deb, Machine learning: the new 'big thing' for competitive advantage, Int. J. Knowl. Eng. Data Min. 5 (2018) 277–305.
[2] D. Castelvecchi, Can we open the black box of AI?, Nature 538 (2016) 20–23.
[3] A. Shaban-Nejad, M. Michalowski, J.S. Brownstein, D.L. Buckeridge, Guest Editorial Explainable AI: towards fairness, accountability, transparency and trust in healthcare, IEEE J. Biomed. Health Inform. 25 (2021) 2374–2375.
[4] D. Shin, Y.J. Park, Role of fairness, accountability, and transparency in algorithmic affordance, Comput. Hum. Behav. 98 (2019) 277–284.
[5] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible ai, Inf. Fusion 58 (2020) 82–115.
[6] J.M. Alonso Moral, C. Castiello, L. Magdalena, C. Mencar, Toward Explainable Artificial Intelligence Through Fuzzy Systems, Springer International Publishing, 2021, pp. 1–23.
[7] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (2019) 206–215.
[8] I. Stepin, J.M. Alonso, A. Catala, M. Pereira-Fariña, A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence, IEEE Access 9 (2021) 11974–12001.
[9] S.M. Mousavi, S. Abdullah, S.T.A. Niaki, S. Banihashemi, An intelligent hybrid classification algorithm integrating fuzzy rule-based extraction and harmony search optimization: medical diagnosis applications, Knowl.-Based Syst. 220 (2021) 106943.
[10] L.-H. Yang, F.-F. Ye, J. Liu, Y.-M. Wang, H. Hu, An improved fuzzy rule-based system using evidential reasoning and subtractive clustering for environmental investment prediction, Fuzzy Sets Syst. 421 (2021) 44–61.
[11] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules, 1st ed., Springer Berlin, Heidelberg, Berlin, 2005.
[12] N. Abdelhamid, F. Thabtah, Associative classification approaches: review and comparison, J. Inf. Knowl. Manag. 13 (2014) 1450027.
[13] A. Bechini, F. Marcelloni, A. Segatori, A mapreduce solution for associative classification of big data, Inf. Sci. 332 (2016) 33–55.
[14] C. Biedma-Rdguez, M.J. Gacto, A. Anguita-Ruiz, J. Alcalá-Fdez, R. Alcalá, Transparent but accurate evolutionary regression combining new linguistic fuzzy grammar and a novel interpretable linear extension, Int. J. Fuzzy Syst. 24 (2022) 3082–3103.
[15] D. Martín, A. Rosete, J. Alcalá-Fdez, F. Herrera, A new multiobjective evolutionary algorithm for mining a reduced set of interesting positive and negative quantitative association rules, IEEE Trans. Evol. Comput. 18 (2014) 54–69.
[16] X. Dong, F. Hao, L. Zhao, T. Xu, An efficient method for pruning redundant negative and positive association rules, Neurocomputing 393 (2020) 245–258.
[17] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends, Fuzzy Sets Syst. 141 (2004) 5–31.
[18] A. Fernández, F. Herrera, O. Cordón, M.J. del Jesus, F. Marcelloni, Evolutionary fuzzy systems for explainable artificial intelligence: why, when, what for, and where to?, IEEE Comput. Intell. Mag. 14 (2019) 69–81.
[19] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (2011) 857–872.

[20] J. Alcalá-Fdez, R. Alcalá, S. González, Y. Nojima, S. García, Evolutionary fuzzy rule-based methods for monotonic classification, IEEE Trans. Fuzzy Syst. 25 (2017) 1376–1390.
[21] F. Padillo, J.M. Luna, S. Ventura, A grammar-guided genetic programing algorithm for associative classification in big data, Cogn. Comput. 11 (2019) 331–346.
[22] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability, Soft Comput. 13 (2009) 959–977.
[23] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 13 (2005) 428–435.
[24] F. Li, C. Shang, Y. Li, J. Yang, Q. Shen, Approximate reasoning with fuzzy rule interpolation: background and recent advances, Artif. Intell. Rev. 54 (2021) 4543–4590.
[25] O. Cordón, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, Int. J. Approx. Reason. 20 (1999) 21–45.
[26] A. Fernández, S. García, M. Galar, R.C. Prati, B. Krawczyk, F. Herrera, Performance measures, in: Learning from Imbalanced Data Sets, 1 ed., Springer, Cham, 2018, pp. 47–61.
[27] A. Ben-David, A lot of randomness is hiding in accuracy, Eng. Appl. Artif. Intell. 20 (2007) 875–885.
[28] M. Kubat, S. Matwin, Addressing the curse of imbalanced training sets: one-sided selection, in: Fourteenth International Conference on Machine Learning (ICML), Morgan Kaufmann, San Francisco, CA, USA, 1997, pp. 179–186.
[29] E. Hullermeier, Y. Yi, In defense of fuzzy association analysis, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 37 (2007) 1039–1043.
[30] N. Marín, M.D. Ruiz, D. Sánchez, Fuzzy frameworks for mining data associations: fuzzy association rules and beyond, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 6 (2016) 50–69.
[31] F. Berzal, I. Blanco, D. Sanchez, M. Vila, Measuring the accuracy and interest of association rules: a new framework, Intell. Data Anal. 6 (2002) 221–235.
[32] L. Geng, H.J. Hamilton, Interestingness measures for data mining: a survey, ACM Comput. Surv. 38 (2006) 1–32.
[33] B. Kavsek, N. Lavrac, APRIORI-SD: adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (2006) 543–583.
[34] L. Eshelman, The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination, in: G. Rawlin (Ed.), Foundations of Genetic Algorithms, vol. 1, Morgan Kaufmann, 1991, pp. 265–283.
[35] L. Eshelman, J. Schaffer, Real-coded genetic algorithms and interval schemata, in: D. Whitley (Ed.), Foundations of Genetic Algorithms, vol. 2, Morgan Kaufmann, 1993, pp. 187–202.
[36] R. Alcalá, J. Alcalá-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Trans. Fuzzy Syst. 15 (2007) 616–635.
[37] M. Lozano, F. Herrera, N. Krasnogor, D. Molina, Real-coded memetic algorithms with crossover hill-climbing, Evol. Comput. 12 (2004) 273–302.
[38] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Mult.-Valued Log. Soft Comput. 17 (2011) 255–287.
[39] I. Triguero, S. González, J.M. Moyano, S. García López, J. Alcalá Fernández, J. Luengo Martín, A.L. Fernández Hilario, M.J.d. Jesús Díaz, L. Sánchez, F. Herrera Triguero, et al., Keel 3.0: an open source software for multi-stage analysis in data mining, Int. J. Comput. Intell. Syst. 10 (2017) 1238–1249.
[40] W. Li, J. Han, J. Pei, Cmar: accurate and efficient classification based on multiple class-association rules, in: Proceedings IEEE International Conference on Data Mining (ICDM), California, USA, 2001, pp. 369–376.
[41] X. Yin, J. Han, Cpar: classification based on predictive association rules, in: Proceedings of 3rd SIAM International Conference on Data Mining (SDM), San Francisco, CA, USA, 2003, pp. 331–335.
[42] J. Quinlan, Mdl and categorical theories (continued), in: Machine Learning: Proceedings of the Twelfth International Conference, 1995, pp. 464–470.
[43] D. García, A. González, R. Pérez, Overview of the SLAVE learning algorithm: a review of its evolution and prospects, Int. J. Comput. Intell. Syst. 7 (2014) 1194–1221.
[44] E.G. Mansoori, M.J. Zolghadri, S.D. Katebi, SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzzy Syst. 16 (2008) 1061–1071.
[45] J. Hühn, E. Hüllermeier, Furia: an algorithm for unordered fuzzy rule induction, Data Min. Knowl. Discov. 19 (2009) 293–319.
[46] U. Fayyad, K. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, in: 13th International Joint Conference on Uncertainly in Artificial Intelligence, IJCAI93, 1993, pp. 1022–1029.
[47] S. García, F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, J. Mach. Learn. Res. 9 (2008) 2677–2694.
[48] M. Galende, M.J. Gacto, G. Sainz, R. Alcalá, Comparison and design of interpretable linguistic vs. scatter frbss: Gm3m generalization and new rule meaning index for global assessment and local pseudo-linguistic representation, Inf. Sci. 282 (2014) 190–213.
[49] M. Latorre-Millán, A.I. Rupérez, E.M. González-Gil, A. Santaliestra-Pasías, R. Vázquez-Cobela, M. Gil-Campos, C.M. Aguilera, A. Gil, L.A. Moreno, R. Leis, G. Bueno, Dietary patterns and their association with body composition and cardiometabolic markers in children and adolescents: genobox cohort, Nutrients 12 (2020) 3424.
[50] F.J. Llorente-Cantarero, F.J. Aguilar-Gómez, R. Leis, G. Bueno, A.I. Rupérez, A. Anguita-Ruiz, R. Vázquez-Cobela, M.D. Mesa, L.A. Moreno, A. Gil, C.M. Aguilera, M. Gil-Campos, Relationship between physical activity, oxidative stress, and total plasma antioxidant capacity in Spanish children from the genobox study, Antioxidants 10 (2021) 320.
[51] A. Ferrández, Curvas y tablas de crecimiento (estudios longitudinal y transversal), Fundación Andrea Prader, Zaragoza, 2005.
[52] R. Kuczmarski, C. Ogden, L. Grummer-Strawn, K. Flegal, S. Guo, R. Wei, Z. Mei, L. Curtin, A. Roche, C. Johnson, CDC growth charts: United States, Adv. Data 314 (2020) 1–27.

[53] J.T. Flynn, D.C. Kaelber, C.M. Baker-Smith, D. Blowey, A.E. Carroll, S.R. Daniels, S.D. de Ferranti, J.M. Dionne, B. Falkner, S.K. Flinn, S.S. Gidding, C. Goodwin, M.G. Leu, M.E. Powers, C. Rea, J. Samuels, M. Simasek, V.V. Thaker, E.M. Urbina, S. on Screening, M. of High Blood Pressure in Children, Clinical practice guideline for screening and management of high blood pressure in children and adolescents, Pediatrics 140 (2017) 20171904.

[54] NCEP Expert Panel on Blood Cholesterol Levels in Children and Adolescents, National cholesterol education program (NCEP): highlights of the report of the expert panel on blood cholesterol levels in children and adolescents, Pediatrics 89 (1992) 495–501.

[55] A. Anguita-Ruiz, A. Mendez-Gutierrez, A.I. Ruperez, R. Leis, G. Bueno, M. Gil-Campos, I. Tofe, C. Gomez-Llorente, L.A. Moreno, A. Gil, C.M. Aguilera, The protein s100a4 as a novel marker of insulin resistance in prepubertal and pubertal children with obesity, Metabolism 105 (2020) 154187.

[56] Á. Torres-Martos, A. Anguita-Ruiz, M. Bustos-Aibar, S. Cámara-Sánchez, R. Alcalá, C.M. Aguilera, J. Alcalá-Fdez, Human multi-omics data pre-processing for predictive purposes using machine learning: a case study in childhood obesity, in: I. Rojas, O. Valenzuela, F. Rojas, L.J. Herrera, F. Ortuño (Eds.), International Work-Conference on Bioinformatics and Biomedical Engineering, in: Springer International Publishing, vol. 13347, Gran Canaria, Spain, 2022, pp. 359–374.

[57] R.E. Jones, J. Jewell, R. Saksena, X. Ramos Salas, J. Breda, Overweight and obesity in children under 5 years: surveillance opportunities and challenges for the who European region, Front. Public Health 5 (2017).

[58] K. Bornfeldt, I. Tabas, Insulin resistance, hyperglycemia, and atherosclerosis, Cell Metab. 14 (2011) 575–585.

[59] J. Després, I. Lemieux, Abdominal obesity and metabolic syndrome, Nature 444 (2006) 881–887.

[60] D. Martín, M. Martínez-Ballesteros, D. García-Gil, J. Alcalá-Fdez, F. Herrera, J. Riquelme-Santos, MRQAR: a generic mapreduce framework to discover quantitative association rules in big data problems, Knowl.-Based Syst. 153 (2018) 176–192.

[61] C.-H. Wang, W.-H. Lee, C.-T. Pang, Applying fuzzy FP-growth to mine fuzzy association rules, World Acad. Sci., Eng. Technol. 65 (2010) 956–962.