Multi-objective evolutionary algorithms for fuzzy classification in survival prediction

Fernando Jiménez*[1], Gracia Sánchez[1], and José M. Juárez[1]

[1]Department of Information and Communication Engineering. Faculty of Computer Science. University of Murcia, 30100, Murcia, Spain.

fernan@um.es, gracia@um.es, jmjuarez@um.es

December 3, 2013

Abstract

Objective: This paper presents a novel rule-based fuzzy classification methodology for survival/mortality prediction in severe burnt patients. Due to the ethical aspects involved in this medical scenario, physicians tend not to accept a computer-based evaluation unless they understand why and how such a recommendation is given. Therefore, any fuzzy classifier model must be both accurate and interpretable.

Methods and materials: The proposed methodology is a three-step process: (1) multi-objective constrained optimization of a patient's data set, using Pareto-based elitist multi-objective evolutionary algorithms to maximize accuracy and minimize the complexity (number of rules) of classifiers, subject to interpretability constraints; this step produces a set of alternative (Pareto) classifiers; (2) linguistic labeling, which assigns a linguistic label to each fuzzy set of the classifiers; this step is essential to the interpretability of the classifiers; (3) decision making, whereby a classifier is chosen, if it is satisfactory, according to the preferences of the decision maker. If no classifier is satisfactory for the decision maker, the process starts again in step (1) with a different input parameter set.

*Contacting author: fernan@um.es, tel. +34 868 884 630, fax +34 868 884 151.

Results: The performance of three multi-objective evolutionary algorithms, niched pre-selection multi-objective algorithm, elitist Pareto-based multi-objective evolutionary algorithm for diversity reinforcement (ENORA) and the non-dominated sorting genetic algorithm (NSGA-II), was tested using a patient's data set from an intensive care burn unit and a standard machine learning data set from an standard machine learning repository. The results are compared using the hypervolume multi-objective metric. Besides, the results have been compared with other non-evolutionary techniques and validated with a multi-objective cross-validation technique. Our proposal improves the classification rate obtained by other non-evolutionary techniques (decision trees, artificial neural networks, Naive Bayes, and case-based reasoning) obtaining with ENORA a classification rate of 0.9298, specificity of 0.9385, and sensitivity of 0.9364, with 14.2 interpretable fuzzy rules on average.

Conclusions: Our proposal improves the accuracy and interpretability of the classifiers, compared with other non-evolutionary techniques. We also conclude that ENORA outperforms niched pre-selection and NSGA-II algorithms. Moreover, given that our multi-objective evolutionary methodology is non-combinational based on real parameter optimization, the time cost is significantly reduced compared with other evolutionary approaches existing in literature based on combinational optimization.

Keywords:

Fuzzy classification; Multi-objective evolutionary computation; Severity scores; Intensive care burns unit

1 Introduction

Severely burned patients require specialized medical care to minimize mortality [1]. Great efforts have been made to analyze this problem from the clinical, the epidemiological and national health system perspective [2–6]. Although the survival rates for burn patients have improved substantially due to medical care in specialized burn centers, patient mortality is still the primary outcome measure for burn care [7].

Early mortality prediction after admission is essential before an aggressive or conservative therapy can be recommended. Severity scores are simple but useful tools for physicians when evaluating the state of the patient. Scoring systems aim to use the most predictive pre-morbid and injury factors to yield an expected likelihood of death for a given patient [7]. In gen-

eral practice, physicians only use a few scores, usually internationally accepted ones involving very simple calculations, such as the simplified acute physiology score (SAPS I-II), the acute physiology and chronic health evaluation score (APACHE II), or the sequential organ failure assessment (SOFA) [8, 9]. However, their daily use in burn centers gives rise to a potential problem, since the above mentioned scores are not designed for critically ill burn patients.

Open burn wounds increase vulnerability to environmental contamination, the larger the burn size the more vulnerable it is to contamination. Burn-related deaths depend on the total body surface area (TBSA). Moreover, from the epidemiological point of view, sex, weight and age are also significant for mortality from burn wounds [1, 10].

Some effort has been dedicated to predicting mortality using a scoring system after burn trauma based on these evidences. Baux and prognostic burn index (PBI) scores provide a mortality rate by summing age and TBSA [11, 12]. Other authors also consider respiratory problems, such as the model presented in [13] using a respiratory score, the CapeTown score, which extends the Baux score with an inhalation score [14], or the score based on logistic regression presented in [15].

Some authors suggest that, for mortality prediction in burns, TBSA, age and inhalation injury outweigh other factors significantly enough to serve as the only components that need to be used to give a prediction [7], for example, the probability of death estimation described in [16]. On the other hand, the widely used abbreviated burns severity index (ABSI) considers gender, age, inhalation injury, %TBSA and presence of the burns severity score proposed in [17].

However, infections and co-morbidity are also relevant aspects to consider. Burn patients with larger burn injuries (over 30%TBSA) and those colonized by multiple resistant organisms require special precautions [1]. Indeed, 75% of deaths are currently related to infectious complications and inhalation injuries from burn wounds when the TBSA is over 40% [1, 18, 19]. According to [19], about 75% of mortality is related to infections.

If infections and co-morbidity are considered, the number of parameters to analyze increases considerably. The two main problems are the difficulty involved in selecting relevant parameters and the need to provide a survival model that can be easily interpreted by physicians.

Further efforts can be made to combine infection and co-morbidity factors in an attempt to provide a death estimation. For instance, in [4] a mortality prediction model is proposed based on multivariate analysis. In this sense, the combination of evolutionary computation and

fuzzy logic, namely evolutionary fuzzy systems [20, 21], helps to solve this kind of problem. In particular, we focus on rule-based fuzzy classification since rules can be easily interpreted by physicians [22–26]. In our proposal, these rules are obtained from the medical data sets using a multi-objective constrained optimization model which maximizes the classification rate and minimizes the number of rules of the classifier, subject to interpretability constraints. This optimization model is solved by using Pareto-based elitist multi-objective evolutionary algorithms [27–30].

The remainder of this paper takes the following form: Section 2 reviews the main works developed in the fields of fuzzy classification and evolutionary computation, particularly when applied to artificial intelligence in medicine. The main differences and advantages of our proposed approach compared to existing studies are shown. Section 3 describes a fuzzy optimization process for mortality scoring, where a fuzzy classification model is described and a multi-objective constrained optimization model is proposed to learn accurate and comprehensible fuzzy classifiers. In section 4 two Pareto-based elitist multi-objective algorithms (niched pre-selection and ENORA) are proposed to learn fuzzy classifiers according to the proposed multi-objective constrained optimization model. In addition, the well known multi-objective evolutionary algorithm NSGA-II is briefly described. Section 5 includes the experiments carried out and the results obtained for the problem of classifying infection-related mortality in patients suffering from severe burns. In order to compare the results obtained by our algorithm and those obtained by other authors in the scientific community, this work also includes experiments for the well-known problem of the Iris data set classification. An analysis of the obtained results is also included. Section 6 discusses the novelties and benefits of our suggested methodology and the main conclusions of the paper are outlined.

2 Background

In this section we review some of the efforts made to lend weight to prediction of mortality using artificial intelligence techniques. We then analyze fuzzy classification and evolutionary computation techniques and their impact in the medical field.

2.1 Mortality scoring in Artificial Intelligence

There is increasing interest from the medical community to support mortality scoring by artificial intelligence techniques [31–35]. In particular, commonly used scoring systems have been deeply analyzed for critical patient care, such as APACHE or SOFA and severity and organ failure scores for Intensive Care Units (ICU) [8, 9]. In [36, 37] mortality prediction models are presented, combining APACHE score and artificial neuronal networks. The work described in [38] presents a case-based reasoning system using APACHE to support clinical decisions. In [9, 39] SOFA-based models are analyzed to support mortality predictions in ICUs. In [40], the T-CARE system based on temporal case-based reasoning is presented to support severity scores in burns units.

In medical scenarios where mortality prediction depends on a large number of features, some authors propose the use of evolutionary computing. For example, in [41], the authors propose a survival prediction for breast cancer based on genetic programming. A Bayesian model optimized by a genetic algorithm is described in [42] for mortality prediction.

2.2 Fuzzy classification in medicine

One illustrative milestone of these first works is the MYCIN system, a diagnosis support system for infectious diseases in which the medical knowledge is provided from the physician's team in the form of rules [43].

Fuzzy sets [44] have been recognized for their ability to introduce notions of continuity into deductive thinking. Because its continuous nature, the behavior of fuzzy systems is more likely to be closer to medical reality than the behavior of classical systems. Additionally, fuzzy sets allow symbolic models to be used. Fuzzy sets can bridge the gap between the discrete world of reasoning and the continuity of reality, which is the main reason why they are considered useful in [45].

One of the most important areas of application in the fuzzy set theory is *fuzzy rule-based systems (FRBSs)*. These fuzzy logic systems constitute an extension of the classical rule-based systems, because they deal with *if-then* rules, whose antecedents and consequences are composed of fuzzy logic statements, rather than classical logic ones. In a broad sense, an FRBS is a rule-based system in which fuzzy logic is used as a tool for representing different forms of knowledge about a problem, as well as for modeling the interactions and relationships that

exist between its variables. Due to this property, fuzzy logic principles have been successfully applied to a wide range of problems in different domains in which uncertainty and vagueness emerge in varying ways.

Fuzzy classification is one of the most common applications of FRBSs. Some examples of applications in medicine are the classification of medical images [46], interpretation of mammograms [47], classification of the malformation of cortical development [48] and medical diagnosis [49]. Basically, a classifier is an algorithm that assigns a class label to an object, based on the object's description. Therefore, a classifier predicts the class label. The object description comes in the form of a vector which constraints the attribute values relevant for the classification task. Classifiers use a training algorithm and a training data set to learn to predict class labels.

An essential issue for medical decision support systems is the *interpretability* or *comprehensibility* of the classifier. As described in [24], comprehensibility of the fuzzy partitions basically relies on the linguistic interpretability of the fuzzy sets, and the simplicity or compactness of the fuzzy rule base. The former is related to the separation of the neighboring fuzzy sets, also called *transparency* [50], and to the number of different fuzzy sets, while the latter depends on the number of input variables and number of rules (*compactness*). In a medical environment, expert users do not accept a computer based evaluation, unless they understand why and how a certain recommendation is given [26]. Interpretability is considered to be the main advantage of fuzzy systems over other alternatives like neural networks or statistical models [22, 23]. Interpretability means that users are able to understand the fuzzy system's behavior by inspecting the rule base [25], considering: (i) easily and reliably verify the acquired knowledge and to relate it to user's domain knowledge; (ii) facilitate debugging and improve the fuzzy model and the related learning algorithm; (iii) validate the system, for its maintenance; (iv) convince the user that the model's behavior is reliable.

2.3 Evolutionary fuzzy systems

Several approaches have been proposed to generate the if-then fuzzy rules automatically from numerical a data set. Due to the complexity of the problem, *Evolutionary Computation* [51] is one of the most widely used techniques. Evolutionary computation makes use of a metaphor of natural evolution. According to this metaphor, a problem plays the role of an environment where lives a population of individuals, each representing a possible solution to the problem. The

degree of adaptation of each individual to its environment is expressed by an adequacy measure known as *fitness function*. Like evolution in nature, evolutionary algorithms potentially produce progressively better solutions to the problem. The algorithms begin with a initial population of random solutions and, in each iteration, the best individuals are selected and combined using variation operators such as crossing and mutation to build the next generation. This process is repeated until some stop criterion, typically when a number of iterations is reached. An overview of evolutionary computation in medicine is shown in [52].

Evolutionary Computation has successfully been applied to learn fuzzy models [53–55], leading to many complex algorithms, generally called *evolutionary (or genetic) fuzzy systems (EFS)* [20, 21], which have been particularly applied to fuzzy classification [56]. An EFS is a kind of hybrid system that melds the approximate reasoning method of fuzzy systems with the adaptation capabilities of evolutionary algorithms. On the one hand, fuzzy systems have demonstrated the ability to formalize in a computationally efficient manner the approximate reasoning typical of humans. On the other hand, evolutionary algorithms constitute a robust technique in complex optimization, identification, learning, and adaptation problems, including classification. While in the past much attention has been paid to system accuracy, in recent years an increasing number of papers have focused on a balance between interpretability and accuracy [21, 26, 50, 57–60]. One of the current trends in the search of accurate and interpretable fuzzy models is the use of *multi-objective evolutionary algorithms* (MOEAs) [27,29,30,61]. MOEAs have proved to be very effective in searching for optimal solutions to problems that incorporate competing multiple performance criteria, called *Multi-objective Optimization Problem* (MOP).

A MOP is formulated as a set of minimization/maximization problems of a tuple of n objectives functions $f_1(\vec{x}), \ldots, f_n(\vec{x})$ where $\vec{x}$ is a vector of parameters belonging to a given domain.

A set $\mathcal{F}$ of solutions for a MOP is *not dominated* (or *Pareto optimal*) if and only if for each $\vec{x} \in \mathcal{F}$, there exists no $\vec{y} \in \mathcal{F}$ for which: 1) there exists i ($1 \leq i \leq n$) so that $f_i(\vec{y})$ improves $f_i(\vec{x})$, and 2) for every j, ($1 \leq j \leq n$, $j \neq i$), $f_j(\vec{x})$ does not improve $f_j(\vec{y})$.

MOEAs are particularly suitable for multi-objective optimization [28,29], as they search for multiple optimal solutions in parallel. MOEAs are capable of finding a set of optimal solutions in its final population in a single run. Once the set of optimal solutions is available the most satisfactory can be chosen by applying a preference criterion. Thus, the aim of a multi-objective search algorithm is to discover a family of solutions that are a good approximation of the Pareto

front. In the case of multi-objective EFS, each solution in the front represents an EFS with an associated trade-off between accuracy and interpretability.

The first approaches involving multi-objective EFS for learning fuzzy rules appeared in the late nineties of the last century [62–64]. In [63], a non-Pareto MOEA was used to minimize the classification error and the number of rules. A three-objective EA for linguistic rule extraction was proposed in [65] adding a new objective to the previous two-objective EA [63] to minimize the length of the rules. As recognized in [66], the idea of using Pareto-based MOEAs to optimize multiple objectives in fuzzy modeling was suggested in [62]. This theoretical work was soon applied in [67, 68] for function approximation and dynamic modeling in standard test problems studied in the literature. In [64], a single objective genetic algorithm is used to minimize the approximation error, complexity, sensitivity to noise and continuity of rules by means of a weighted approach.

In [69], MOEAs are used to perform feature selection and fuzzy set granularity learning in order to obtain compact and comprehensible FRBSs with high classification capacity. In [70], MOEAs were used to select fuzzy rules after extracting a large number of candidate rules by a heuristic approach. The goal is to find non-dominated subsets of candidate fuzzy rules with respect to both accuracy maximization and complexity minimization. In [71] and [72], sets of fuzzy rule bases are generated with different trade-offs between accuracy and complexity/interpretability. Each rule is represented as a integer vector in which each integer represents a fuzzy set, and a set of rules is represented by concatenating as many vectors as rules. In [73], a two-objective MOEA is used to improve interpretability in Takagi-Sugeno-Kang (TSK) fuzzy models. The algorithm minimizes the mean squared error and the similarity of the fuzzy sets. A real parameter representation is used to encode Gaussian fuzzy sets and a multi-objective extension of the pre-selection schema proposed in [51]. An explicit niche formation technique is incorporated to ensure a minimum number of solutions for each number of rules. A rule based simplification technique is used to simplify the rule set after crossing and mutation. TSK fuzzy models are interpreted as radial basis function neural networks whose consequence parameters are trained with a gradient method after simplification. The obtained FRBS achieves high accuracy with an acceptable degree of interpretability. In [74], a fuzzy modeling scheme using evolutionary algorithms is proposed. In this scheme, a Non-Pareto multi-objective evolutionary algorithm is applied twice in a batch process, in which similarity is first optimized in combination with a rule-based simplification technique, before the accuracy is optimized.

In [75], a MOEA is used to concurrently learn rule and data base of a FRBS. In this case, two objectives are considered: the first measures the complexity as the sum of the input variable labels used in each of the rules and the second corresponds to the mean square error. In [76], fuzzy classifiers for imbalanced and cost-sensitive data sets are generated with a three objective MOEA. The first and second objectives are sensitivity and specificity, which express, respectively, how well the system classifies patterns belonging to the positive class and to the negative class. The third objective is a complexity measure computed as the sum of the conditions that compose the antecedents of the rules, which is minimized. In [77] a two-objective MOEA for fuzzy modeling in high dimensional regression problems is described. The first objective is to minimize the number of rules and the second objective is to minimize the mean square error. A double coding scheme is used to represent both granularity and translation parameters and a Wang and Mendel algorithm is considered for an ad hoc data-driven method to learn a rule base from each data set definition within the evolutionary process by adding a cropping mechanism. In [78], a decision support system is reported for assigning a liver from a donor to a recipient on a waiting-list that maximizes the probability of belonging to the survival graft class after a year of transplant and minimizes the probability of belonging to the non-survival graft class. A multi-objective evolutionary algorithm is proposed to generate a Pareto front of neural network models for survival classification. A rule based decision support system uses these models to determine the best matching donor-recipient. Nevertheless, this proposal does not take into account interpretability aspects of the models.

An overview, case study and suggestions for future research for Pareto-based multi-objective machine learning are shown in [66, 79, 80].

2.4 Discussion of related works

In the reviewed literature, the combination of fuzzy logic, particularly fuzzy classification, with evolutionary computation has been proved to be one of the most effective techniques in the search for prediction models. Besides, rule-base fuzzy classification is very appropriate in medicine since such models can be easily interpreted by physicians. However, the literature offers a wide variety of EFS involving different methodologies. Two main groups of techniques can be identified: single objective and multi-objective EFSs. Single objective EFSs usually perform the modeling process by a batch execution of different evolutionary algorithms, each of them specialized in a given process, e.g., model structure identification, linguistic fuzzy

partitions, tuning of the fuzzy sets, etc. The main disadvantage of these approaches is the loss of information that occurs between one process and another. Multi-objective EFSs are therefore more appropriate, as the optimization process is performed by a single algorithm in a global search process. For this reason, we focus on multi-objective EFSs in this paper.

Within multi-objective EFSs, two main categories can be identified: one based on combinatorial optimization and the other on real parameter optimization. Combinatorial optimization based EFSs identify the rule set by the combinatorial search for fixed input variable partitions, where each partition has an associated linguistic label. The goal of these approaches is to improve the accuracy of highly comprehensible descriptive fuzzy models. The first disadvantage of these systems is the large number of evaluations needed to reach convergence when high dimensional data sets are considered and therefore much time is required to generate the rule base. Another disadvantage of these approaches is the lack of flexibility of the fixed input variable partitions, which leads to obtain models that, despite being highly comprehensible, may not be accurate enough. On the other hand, real parameter optimization based EFSs are more efficient in the search for models with high dimensional data sets and they also perform a more flexible search as they are not constrained by a fixed input variable partitions leading to high accurate models. The goal of these approaches is to improve the interpretability of highly accurate approximative fuzzy models and therefore these approaches usually need ad hoc rule simplification techniques in order to obtain comprehensible models. We focus on this second category of techniques as our goal is to learn highly accurate fuzzy systems with an acceptable comprehensibility. Besides, it is well known that the power of evolutionary algorithms lies in real parameter optimization. For combinatorial optimization, other non-linear optimization techniques such as Branch-And-Bound, Monte Carlo, etc, may be more suitable.

In this paper we propose a Pareto-based elitist multi-objective evolutionary algorithm for real parameter multi-objective constrained optimization to obtain accurate and comprehensible fuzzy classifiers. This is a novel proposal that differs from all existing works.

3 A fuzzy classification methodology for mortality scoring

In this section, we propose a fuzzy classification methodology to classify input patterns with both real and categorized data. The categorized data may have any number of classes. This fuzzy classification methodology can be used specifically for the classification of mortality by

infection in severe burn patients. Section 3.1 shows the general scheme used to obtain the fuzzy classifier. Section 3.2 describes the fuzzy classifier model used in this work, identifying the reasoning method and rule weight assignment. Finally, section 3.3 shows the multi-objective constrained optimization model used in the optimization phase of the fuzzy classifiers. The criteria and constraints taken into account in the optimization model are identified in order to generate accurate and comprehensible fuzzy classifiers.

3.1 General scheme for fuzzy classification

In this section, we describe a general methodology for fuzzy classification that is graphically illustrated in figure 1. The process starts by extracting instances of patients observed by a physician in a hospital. For each patient, input and output variable values are collected, and the *data set* is built with this set of instances. Next, a MOEA, previously validated with a cross-validation process, is run with a set of parameters supplied by both physician and decision maker. These parameters are shown in table 7. The MOEA searches for fuzzy classifiers which maximize the model classification rate and minimize the number of rules, subject to interpretability constraints, such as the fuzzy set similarity being below a given threshold, and the number of rules being between a minimum and a maximum value. Given that the optimization problem is a real parameter constrained MOP, the solution is composed of a set of Pareto-optimal single solutions. A linguistic labeling process is performed over these solutions to assign a linguistic label to each fuzzy set. Next, a decision-making process is performed over this set of solutions in order to choose the most satisfactory solution. The decision-making process simultaneously takes into account the satisfaction degree with respect to the accuracy and the interpretability of the solutions. If a solution of the Pareto set satisfies both accuracy and interpretability, the solution is shown as an output. In other cases, if no solution satisfies both criteria, the process is started again with a new set of parameters. This process is repeated until a satisfactory solution is found. In the following subsections, we show the fuzzy classifier model, the multi-objective constrained optimization model and the characteristics of the two MOEAs proposed in this paper.

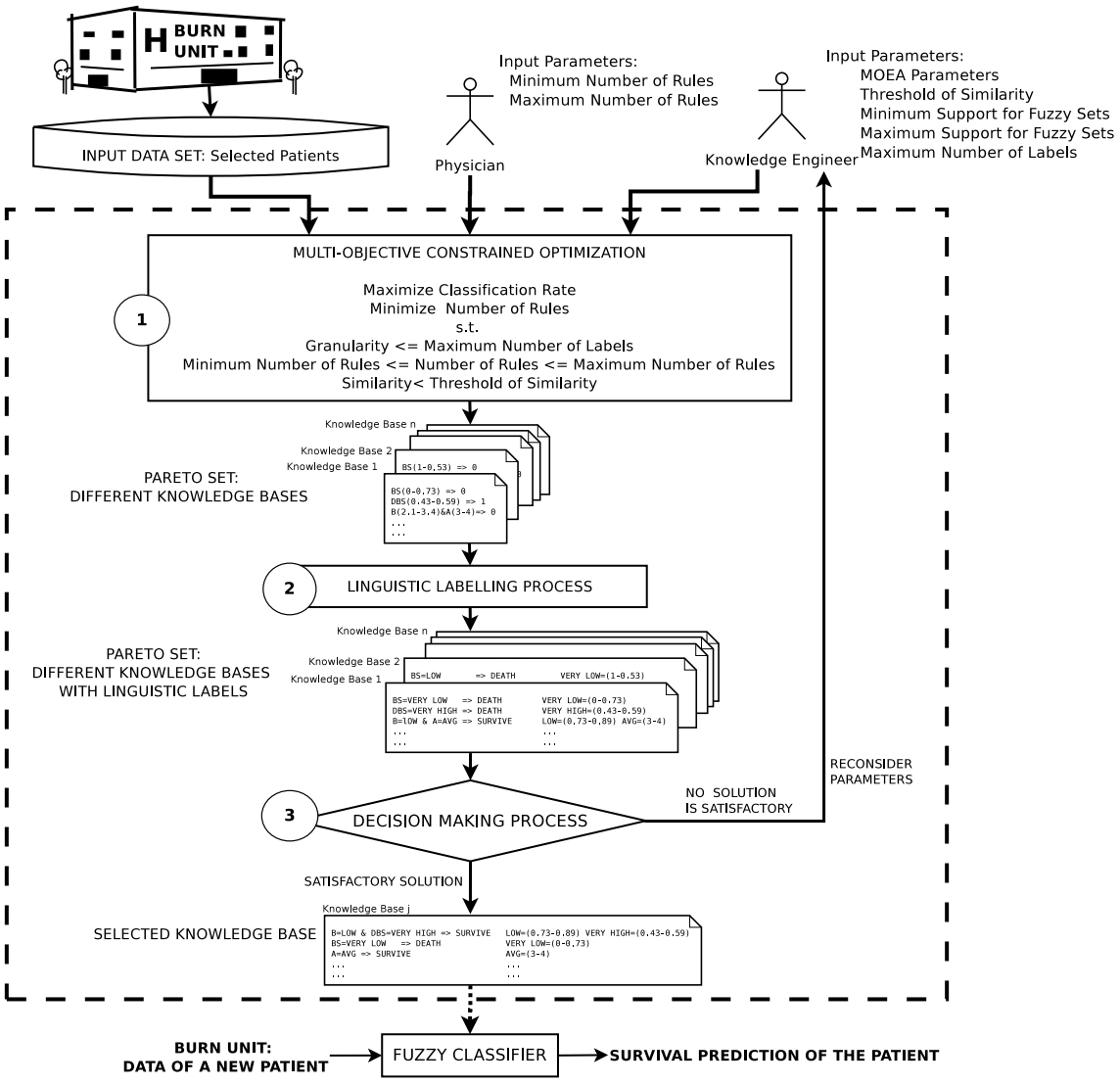


Figure 1: Proposed methodology for fuzzy classification.

3.2 A fuzzy classifier model

Let us consider a data set of N input datum $\Delta = \{d_1, \ldots, d_N\}$. Each input data d_k, $(k = 1, \ldots, N)$ has the following form $d_k = (\vec{x_k}, \vec{w_k}, y_k)$, where

- A vector of real input attributes $\vec{x_k} = \left(x_1^k, \ldots, x_p^k\right)$, $x_i^k \in [l_i, u_i] \subset \Re$, $i = 1, \ldots, p$, $p \geq 0$, where l_i and u_i are the lower and upper values for the real input attribute i, that is, $l_i = \min\limits_{k=1,\ldots,N} \left\{x_i^k\right\}$ and $u_i = \max\limits_{k=1,\ldots,N} \left\{x_i^k\right\}$.
- A vector of integers for the categorized input attributes $\vec{w_k} = \left(w_1^k, \ldots, w_q^k\right)$, $w_i^k \in \{1, \ldots, v_i\}$, $i = 1, \ldots, q$, $q \geq 0$, where $v_i > 1$ is the number of classes for the categorized input attribute i.

- An integer value for the categorized output attribute $y_k \in \{1, \ldots, z\}$, where $z > 1$ is the number of classes for the categorized output attribute.

Note that a boolean input or output attribute can be represented by a categorized attribute w_i or y, so that $v_i = 2$ or $z = 2$. It is also assumed that $p + q > 0$.

Let us consider a fuzzy classifier $\Gamma \in \mathcal{T}$ formed by M_Γ fuzzy rules $R_1^\Gamma, \ldots, R_{M_\Gamma}^\Gamma$ where $\mathcal{T}$ is the space of all possible fuzzy classifiers. We assume that at least one rule should exist for each of the z output classes, so that $M_\Gamma \geq z$. Besides, we assume that the number of rules M_Γ is limited by a lower value, $M_{min} \geq z$, and an upper value, $M_{max} \geq M_{min}$, so that $z \leq min \leq M_\Gamma \leq max$, where the values M_{min} and M_{max} are supplied by a physician. Each fuzzy rule R_j^Γ $(j = 1, \ldots, M_\Gamma)$ contains p fuzzy sets $\widetilde{A}_{ij}^\Gamma$ $(i = 1, \ldots, p)$ associated to p real input attributes, q integer values $B_{ij}^\Gamma \in \{1, \ldots, v_i\}$ $(i = 1, \ldots, q)$ associated to q categorized input attributes, and an integer value $C_j^\Gamma \in \{1, \ldots, z\}$ associated to the categorized output attribute. The set of fuzzy rules $R_1^\Gamma, \ldots, R_{M_\Gamma}^\Gamma$ is usually called [20] *rule base* (RB).

Each fuzzy set $\widetilde{A}_{ij}^\Gamma$ $(i = 1, \ldots, p)$ $(j = 1, \ldots, M_\Gamma)$ can be described by a membership function $\mu_{\widetilde{A}_{ij}^\Gamma} : \mathcal{X}_i \to [0, 1]$, where $\mathcal{X}_i = [l_i, u_i] \subset \Re$ is the domain of the real input attribute x_i.

In our model, we use, for suitability [81], gaussian membership functions:

$$\mu_{\widetilde{A}_{ij}^\Gamma}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - a_{ij}^\Gamma}{\sigma_{ij}^\Gamma}\right)^2\right] \tag{1}$$

where $a_{ij}^\Gamma \in \Re$ is the center, and $\sigma_{ij}^\Gamma \in \Re$ is the variance. We assume $a_{ij}^\Gamma \in [l_i, u_i] \subset \Re$ and $\sigma_{ij}^\Gamma \in \left[\frac{u_i - l_i}{\gamma_1}, \frac{u_i - l_i}{\gamma_2}\right] \subset \Re$, with $\gamma_1 > \gamma_2 > 0$. In our case, we choose the values $\gamma_1 = 30$ which will lead to interpretable models with less than 7 linguistic labels (see section 4.8 for a detailed explanation), and $\gamma_2 = 2$ which ensures that at least 47.72% of any fuzzy set will be within the variable domain. The set of membership functions $\mu_{\widetilde{A}_{ij}^\Gamma} : \mathcal{X}_i \to [0, 1]$ is usually called [20] *data base* (DB).

A fuzzy rule R_j^Γ, therefore, has the following structure:

$$\begin{aligned} R_j^\Gamma: \ & if \quad x_1 \ is \ \widetilde{A}_{1j}^\Gamma \quad and \ \ldots \ and \quad x_p \ is \ \widetilde{A}_{pj}^\Gamma \quad and \\ & \quad\ \ w_1 \ is \ B_{1j}^\Gamma \quad and \ \ldots \ and \quad w_q \ is \ B_{qj}^\Gamma \\ & then \quad y \ is \ C_j^\Gamma \end{aligned} \tag{2}$$

A fuzzy rule based classifier is composed of an RB and a DB, which is usually called [20] *knowledge base*.

Besides, a fuzzy rule based classifier is characterized by its *reasoning method*, which uses the information from the rule to determine a class label for a specific input data. A classifier is also characterized by a *rule weight method*. In our case, the weight of a rule R_j^Γ, i.e., the degree of certainty of the classification in the class C_j^Γ for an input datum belonging to the fuzzy subspace defined by the antecedent of the rule, is equal to 1 and for any other class $C_k \in \{1, \ldots, z\}$ such as $C_k \neq C_j^\Gamma$, is equal to 0.

The *firing degree* (or *matching degree*) of a rule R_j^Γ for an input datum $(\vec{x}, \vec{w})$ is the strength of activation of the if-part of the rule, $\varphi_j^\Gamma(\vec{x}, \vec{w})$, calculated as:

$$\varphi_j^\Gamma(\vec{x}, \vec{w}) = \left(\phi_j^\Gamma(\vec{w}) + 1\right) \prod_{i=1}^{p} \mu_{\widetilde{A}_{ij}^\Gamma}(x_i)$$

where $\phi_j^\Gamma(\vec{w})$ is the number of integer input attributes, so that $w_j = B_{ij}^\Gamma$. The firing degree is obtained by applying a t-norm product to the degree of satisfaction of the clauses $x_i \ \ is \ \widetilde{A}_{ij}^\Gamma$ multiplied by the number of matches of the integer input data $w_i \ \ is \ B_{ij}^\Gamma$.

The *degree of association* of an input datum $(\vec{x}, \vec{w})$ and an output class $C \in \{1, \ldots, z\}$, $\lambda_C^\Gamma(\vec{x}, \vec{w})$ is calculated by summing the firing degree $\varphi_j^\Gamma(\vec{x}, \vec{w})$ of each rule R_j^Γ $(j = 1, \ldots, M_\Gamma)$ whose value for the integer output attribute C_j^Γ is equal to C, that is:

$$\lambda_C^\Gamma(\vec{x}, \vec{w}) = \sum_{\substack{j = 1, \ldots, M_\Gamma \\ C_j^\Gamma = C}} \varphi_j^\Gamma(\vec{x}, \vec{w})$$

This degree of association $\lambda_C^\Gamma(\vec{x}, \vec{w})$ is a *soundness degree* of the classification of the input data $(\vec{x}, \vec{w})$ in class C for the classifier Γ.

The *classification* for input data $(\vec{x}, \vec{w})$ or output of the classifier Γ, $f_\Gamma(\vec{x}, \vec{w})$, corresponds to the class $C \in \{1, \ldots, z\}$ whose degree of association $\lambda_C^\Gamma(\vec{x}, \vec{w})$ is maximum, that is:

$$f_\Gamma(\vec{x}, \vec{w}) = \arg_C \max_{C=1}^{z} \lambda_C^\Gamma(\vec{x}, \vec{w})$$

3.3 A multi-objective constrained optimization model for accurate and comprehensible fuzzy classifiers

In this section, we propose the following optimization model that allows the identification of both accurate and comprehensible fuzzy classifiers:

$$
\begin{aligned}
&Maximize && \mathcal{CR}\left(\Gamma, \Delta\right) \\
&Minimize && \mathcal{NR}\left(\Gamma\right) \\
&Subject\ to: && M_\Gamma \geq M_{min} \\
& && M_\Gamma \leq M_{max} \\
& && \mathcal{S}\left(\Gamma\right) \leq g_s \\
& && \mathcal{NL}\left(\Gamma\right) \leq L_{max}
\end{aligned}
\tag{3}
$$

with the following boundary constraints:

$$
\begin{aligned}
&a_{ij}^{\Gamma} \in [l_i, u_i] \subset \Re,\ i = 1, \ldots, p\ ,\ j = 1, \ldots, M_\Gamma \\
&\sigma_{ij}^{\Gamma} \in \left[\tfrac{u_i - l_i}{\gamma_1}, \tfrac{u_i - l_i}{\gamma_2}\right] \subset \Re,\ i = 1, \ldots, p\ ,\ j = 1, \ldots, M_\Gamma \\
&B_{ij}^{\Gamma} \in \{1, \ldots, v_i\}\ ,\ i = 1, \ldots, q\ ,\ j = 1, \ldots, M_\Gamma \\
&C_{j}^{\Gamma} \in \{1, \ldots, z\}\ ,\ j = 1, \ldots, M_\Gamma
\end{aligned}
\tag{4}
$$

The optimization model (3) is a two objective optimization problem with three constraints. The first objective implements the accuracy criterion by maximizing the *classification rate* of the classifier, $\mathcal{CR}\left(\Gamma\right)$. This value is calculated from the set of N input data as follows:

$$
\mathcal{CR}\left(\Gamma, \Delta\right) = \frac{\Phi\left(\Gamma, \Delta\right)}{N}
$$

where $\Phi\left(\Gamma, \Delta\right)$ is the number of input data from Δ, $d_k = \{\vec{x_k}, \vec{w_k}, y_k\}$ $(k = 1, \ldots, n)$ for which $f_\Gamma\left(\vec{x_k}, \vec{w_k}\right) = y_k$.

The second objective implements the simplicity (or compactness) criterion by minimizing the number of rules of the classifier, i.e.:

$$
\mathcal{NR}\left(\Gamma\right) = M_\Gamma
$$

Note that, by minimizing the number of rules M_Γ, the number L_Γ of different fuzzy sets is also minimized. Besides, the model ensures that there are no repeated rules since a solution with a repeated rule would be dominated by another solution without repeated rules.

The constraints $M_\Gamma \leq M_{min}$ and $M_\Gamma \geq M_{max}$ ensure that the number of rules of the fuzzy classifier Γ is between a minimum M_{min} and a maximum M_{max} number.

The constraint $\mathcal{S}\left(\Gamma\right) \leq g_s$ is imposed to reach transparency by ensuring a minimum similarity g_s $(0 < g_s \leq 1)$ between the fuzzy sets. To measure the degree of separation between the fuzzy sets we use the *similarity* [82]. The similarity of a fuzzy classifier, $\mathcal{S}\left(\Gamma\right)$, is calculated as the maximum similarity between different fuzzy sets in the following way:

$$\mathcal{S}(\Gamma) = \begin{array}{c} \max \\ j,k = 1,\ldots,M_\Gamma \\ i = 1,\ldots,p \\ \widetilde{A}_{ij}^{\Gamma} \neq \widetilde{A}_{ik}^{\Gamma} \end{array} Sim(\widetilde{A}_{ij}^{\Gamma}, \widetilde{A}_{ik}^{\Gamma}) \tag{5}$$

The similarity Sim between two different fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ can be measured by:

$$Sim(\widetilde{A}, \widetilde{B}) = \max\left\{\frac{|\widetilde{A} \cap \widetilde{B}|}{|\widetilde{A}|}, \frac{|\widetilde{A} \cap \widetilde{B}|}{|\widetilde{B}|}\right\} \tag{6}$$

Note that the similarity value of a classifier Γ represents the maximum value of overlapping among their fuzzy sets for any input variable. The constraint $\mathcal{S}(\Gamma) \leq g_s$ ensures, therefore, that there will not be two fuzzy sets for any input variable overlapped more than a g_s value. In our case, we choose a value $g_s = 0.1$ which ensures a maximum overlapping of 10% between any two fuzzy sets of the same variable.

Finally, the constraint $\mathcal{NL}(\Gamma) \leq L_{max}$ ensures that the maximum number of different fuzzy sets for the classifier Γ, calculated as $\mathcal{NL}(\Gamma)$, is smaller than a given value L_{max}. The value $\mathcal{NL}(\Gamma)$ is calculated as:

$$\mathcal{NL}(\Gamma) = \max_{i=1,\ldots,p} \frac{u_i - l_i}{S_i^{\Gamma}}$$

where S_i^{Γ} is the minimum distance between the centers of any two different fuzzy sets of the classifier Γ for the i real input variable:

$$S_i^{\Gamma} = \begin{array}{c} \min \\ j,k = 1,\ldots,M^{\Gamma} \\ \widetilde{A}_{ij}^{\Gamma} \neq \widetilde{A}_{ik}^{\Gamma} \end{array} \left|a_{ij}^{\Gamma} - a_{ik}^{\Gamma}\right|,$$

4 Multi-objective evolutionary algorithm for learning fuzzy classifiers

We propose an evolutionary learning system to search for multiple Pareto-optimal solutions (classifiers) simultaneously, taking into account criteria of accuracy and comprehensibility according to the multi-objective constrained optimization model (3). In this section, the main components of the proposed evolutionary algorithm are described. These components are the

Common characteristics
Pareto-based elitist multi-objective evolutionary algorithms.
Variable-length representation with real and categorized input variables with a Pittsburgh approach.
Handling constraints using a repair algorithm.
Adaptive variation operators.
Specific characteristics
Niched pre-selection: Steady-state substitution ($n = 2$), explicit niche formation technique.
ENORA: Ordering solutions by non-domination level of the individual in its niche and density.
NSGA-II: Ordering solutions by non- domination level of the individual in the population and density.

Table 1: Common and specific characteristics of niched pre-selection, ENORA and NSGA-II.

solution representation, constraints handling, initial population and variation operators. With these common components, we study three elitist Pareto-based multi-objective evolutionary algorithms with different selection, sampling and generational replacement techniques, namely niched pre-selection, ENORA and NSGA-II. The niched pre-selection technique was initially developed by the authors for function approximation and dynamic modeling with TSK fuzzy models [73]. The ENORA technique was also proposed by the authors for multi-objective constrained real parameter optimization in [83]. NSGA-II is the well known technique proposed in [29] which was also initially developed for real parameter optimization in multi-objective constrained optimization problems. Table 1 resumes both common and specific components of these algorithms.

4.1 Representation of solutions and evaluation

We use a length-variable representation involving the codification of real and discrete numbers, using the Pittsburgh approach [63]. Each individual I of a population contains a variable number of rules M_I. Each rule R_j^I, $j = 1, \ldots, M_I$ codifies the following components:

- The fuzzy sets $\widetilde{A}_{ij}^I$ associated to the real input attributes x_i, $i = 1, \ldots, p$, by means of real numbers $a_{ij}^I \in [l_i, u_i]$ and $\sigma_{ij}^I \in \left[\frac{u_i - l_i}{\gamma_1}, \frac{u_i - l_i}{\gamma_2}\right]$, which define the centers and variances, respectively.

- The discrete values associated to the categorized input attributes w_i, $i = 1, \ldots, q$, by means of integer numbers $b_{ij}^I \in \{1, \ldots, v_i\}$.

Parameters for the codification of the fuzzy set rules.													
Rules	Centers of fuzzy sets				Variances of fuzzy sets				Categorized inputs				Outputs
R_1^I =	a_{11}^I	a_{21}^I	$\dots$	a_{p1}^I	σ_{11}^I	σ_{21}^I	$\dots$	σ_{p1}^I	b_{11}^I	b_{21}^I	$\dots$	b_{q1}^I	c_1^I
R_2^I =	a_{12}^I	a_{22}^I	$\dots$	a_{p2}^I	σ_{12}^I	σ_{22}^I	$\dots$	σ_{p2}^I	b_{12}^I	b_{22}^I	$\dots$	b_{q2}^I	c_2^I
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$R_{M_I}^I$ =	$a_{1M_I}^I$	$a_{2M_I}^I$	$\dots$	$a_{pM_I}^I$	$\sigma_{1M_I}^I$	$\sigma_{2M_I}^I$	$\dots$	$\sigma_{pM_I}^I$	$b_{1M_I}^I$	$b_{2M_I}^I$	$\dots$	$b_{qM_I}^I$	$c_{M_I}^I$
Parameters to carry out adaptive crossing and mutation.													
d_I =	Associated crossing.												
e_I =	Associated mutation.												
e_c^I =	Amplitude of the mutation of the centers of the fuzzy sets.												
e_v^I =	Amplitude of the mutation of the variations of the fuzzy sets.												

Table 2: Chromosome coding for an individual I.

- The discrete value associated to the categorized output attribute by means of an integer number $c_j^I \in \{1, \dots, z\}$.

Additionally, to carry out adaptive crossing and mutation, each individual has two discrete parameters $d_I \in \{0, \dots, \delta\}$ and $e_I \in \{0, \dots, \epsilon\}$ associated to crossing and mutation, where $\delta \geq 0$ is the number of crossing operators and $\varepsilon \geq 0$ is the number of mutation operators. In order to perform adapting mutation, there are also two real parameters $e_c^I, e_v^I \in [0, 1]$ that define the amplitude of the mutation of the centers and variances of the fuzzy sets, respectively. Table 2 summarizes the chromosome coding for an individual I.

An individual I is evaluated with two fitness functions, f_1^I and f_2^I, corresponding to the two objectives of the multi-objective constrained optimization model (3):

$$\begin{aligned} f_1^I &= \mathcal{CR}\left(\Gamma_I, \Delta\right) \\ f_2^I &= \mathcal{NR}\left(\Gamma_I\right) \end{aligned} \tag{7}$$

where Γ_I is the fuzzy model represented by the individual I.

4.2 Handling constraints

We use problem-specific knowledge for constraints handling. The first two constraints, $M_\Gamma \leq M_{min}$ and $M_\Gamma \geq M_{max}$, are satisfied by means of a specialized initialization procedure and variation operators which always generate individuals with a number of rules (M_I) between M_{min} and M_{max}. The constraints $\mathcal{S}\left(\Gamma\right) \leq g_s$ and $\mathcal{NL}\left(\Gamma\right) \leq L_{max}$ are more difficult to satisfy.

To fulfil these constraints, we propose a repair algorithm. These algorithms are very popular techniques in evolutionary computation for constraints handling [84], and they use problem-specific knowledge to transform unfeasible individuals into feasible ones.

Algorithm 1 shows the proposed repair method. In this algorithm, fuzzy sets are first joined (steps 7 to 8), merged (steps 11 to 14) or separated (step 17) in order to make any difference between fuzzy sets greater than the minimum allowed; to obtain classifiers with a maximum number of fuzzy sets L_{max}, the minimum separation between fuzzy sets for input real variable i must be $S_i \leftarrow \frac{u_i - l_i}{L_{max}}$. Following the variance of the fuzzy sets is reduced (steps 24 to 25), when needed, in order to get a maximum similarity smaller than the g_s threshold allowed so that the constraint $\mathcal{S}(\Gamma) \leq g_s$ is satisfied. Finally, the algorithm searches for identical rules from the rule set, which are then removed (steps 30 to 32). Notice that, for a given similarity threshold of g_s and a minimum separation S_i established by a maximum number of fuzzy sets L_{max}, a minimum variance σ_i is established and therefore γ_1 parameter must be established to ensure that this value σ_i is greater than the minimum variance $\frac{u_i - l_i}{\gamma_1}$. In our case, for a value $g_s = 0.1$ and a $L_{max} = 7$, it is necessary to set $\gamma_1 = 30$.

Algorithm 1 is an improvement of the rule base simplification algorithm proposed in [82]. In our algorithm the fuzzy sets are merged and separated while in the algorithm proposed in [82] the fuzzy sets are only merged.

For the sake of illustration, we consider a fictitious problem with a single real input variable with domain in $[0, 10]$, and a boolean output variable. The Repair algorithm is executed with a similarity threshold $g_s = 0.1$ and $L_{max} = 7$. Table 3 shows, on the left hand side, a non-feasible classifier with 6 rules, 6 fuzzy sets and similarity $\mathcal{S} = 0.996 > g_s = 0.1$. During the repair process, fuzzy set $\widetilde{A}_{12}$ is separated from fuzzy set $\widetilde{A}_{11}$ the minimum distance 1.43 and fuzzy sets $\widetilde{A}_{15}$ and $\widetilde{A}_{16}$ is merged into a single fuzzy set $\widetilde{A}'_{15}$; following the variance of the fuzzy sets are narrowed, when needed, in order to achieve a similarity smaller than the minimum allowed. With this process, the new rules R'_5 and R'_6 are alike and therefore R'_6 is erased. After the repair algorithm, the new classifier has 5 rules, 5 fuzzy sets and a similarity $\mathcal{S} = 0.094$ and can therefore be considered a feasible classifier ($\mathcal{S} = 0.098 \leq g_s = 0.1$). Table 4 shows a more complex example for the same problem with an initial non-feasible classifier with 10 rules, 10 fuzzy sets and $\mathcal{S} = 0.997 > g_s = 0.1$. In this case, during the repair fuzzy, multiple fuzzy sets are separated and merged and multiple rules are removed. After the repair algorithm, a feasible classifier with 6 rules, 5 fuzzy sets and $\mathcal{S} = 0.074 \leq g_s = 0.1$ is obtained.

Algorithm 1 Repair method

Require: g_s ($0 < g_s < 1$) {Threshold value for similarity}

Require: M_{min} ($M_{min} > 0$) {Minimum number of rules}

Require: L_{max} ($L_{max} > 0$) {Maximum number of fuzzy sets}

Require: I {Individual to repair.}

$\{R_j^I, j = 1, \ldots, M_I$: Rules of the model for individual $I\}$

$\{\widetilde{A}_{ij}^I, i = 1, \ldots, p, j = 1, \ldots, M_I$: Fuzzy sets for the real input variables$\}$

$\{a_{ij}^I, i = 1, \ldots, p, j = 1, \ldots, M_I$: Real values of fuzzy sets centers for the real input variables$\}$

$\{\sigma_{ij}^I, i = 1, \ldots, p, j = 1, \ldots, M_I$: Real values of fuzzy sets variances for the real input variables$\}$

for $i = 1$ to p **do**

  Order $\widetilde{A}_{ij}^I$ so that $a_{ij}^I \leq a_{ij+1}^I, j = 1, \ldots, M_I - 1$

  {Make separation between any fuzzy sets centers greater than minimum allowed}

  $S_i \leftarrow \frac{u_i - l_i}{L_{max}}$ {Minimum separation between fuzzy sets centers for i real input variable}

  for $j = 1$ to $M_I - 1$ **do**

    if $\widetilde{A}_{ij+1}^I \neq \widetilde{A}_{ij}^I$ **then**

      if $\left(a_{ij+1}^I - a_{ij}^I < 0\right)$ {Join $\widetilde{A}_{ij}^I$ and $\widetilde{A}_{ij+1}^I$} **then**

        $\sigma_{ij}^I \leftarrow \max\left(\sigma_{ij}^I, \sigma_{ij+1}^I\right)$

        $\widetilde{A}_{ij+1}^I \leftarrow \widetilde{A}_{ij}^I$

      end if

      if $\left(a_{ij+1}^I - a_{ij}^I \geq 0\right)$ *and* $\left[\left(a_{ij+1}^I - a_{ij}^I < \frac{S_i}{2}\right)\right.$ *or* $\left.\left(a_{ij}^I + S_i > u_i\right)\right]$ {Merge $\widetilde{A}_{ij}^I$ and $\widetilde{A}_{ij+1}^I$} **then**

        $\theta \leftarrow \frac{\sigma_{ij}^I}{\sigma_{ij}^I + \sigma_{ij+1}^I}$

        $a_{ij}^I \leftarrow \theta a_{ij}^I + (1 - \theta)\, a_{ij+1}^I$

        $\sigma_{ij}^I \leftarrow \max\left(\sigma_{ij}^I, \sigma_{ij+1}^I\right)$

        $\widetilde{A}_{ij+1}^I \leftarrow \widetilde{A}_{ij}^I$

      end if

      if $\left(\frac{S_i}{2} \leq a_{ij+1}^I - a_{ij} < S_i\right)$ and $\left(a_{ij}^I + S_i \leq u_i\right)$ {Separate $\widetilde{A}_{ij+1}^I$ from $\widetilde{A}_{ij}^I$} **then**

        $a_{ij+1}^I \leftarrow a_{ij}^I + S_i$

      end if

    end if

  end for

  {Reduce variance of fuzzy sets with similar greater than η}

  for $j = 1$ to $M_I - 1$ **do**

    if $\widetilde{A}_{ij+1}^I \neq \widetilde{A}_{ij}^I$ {Reduce variance of $\widetilde{A}_{ij}^I$ and $\widetilde{A}_{ij+1}^I$} **then**

      while $Sim\left(\widetilde{A}_{ij}^I, \widetilde{A}_{ij+1}^I\right) > g_s$ **do**

        $\sigma_{ij}^I \leftarrow 0.9\sigma_{ij}^I$

        $\sigma_{ij+1}^I \leftarrow 0.9\sigma_{ij+1}^I$

      end while

    end if

  end for

end for

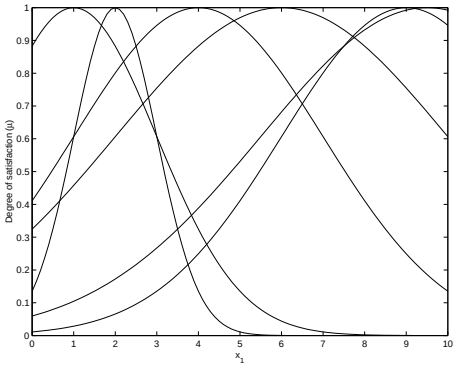
{Remove identical rules}

while $M_I > M_{min}$ and exists $j, k = 1, \ldots, M_I$ so that $R_j^I = R_k^I$ **do**

  Remove R_k^I

end while

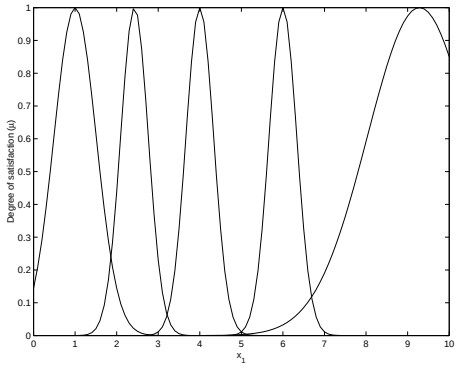
	$\boldsymbol{x_1}$	$\boldsymbol{y}$
$R_1:$	$\widetilde{A}_{11} = (1.0; 2.0)$	1
$R_2:$	$\widetilde{A}_{12} = (2.0; 1.0)$	1
$R_3:$	$\widetilde{A}_{13} = (4.0; 3.0)$	1
$R_4:$	$\widetilde{A}_{14} = (6.0; 4.0)$	1
$R_5:$	$\widetilde{A}_{15} = (9.0; 3.0)$	2
$R_6:$	$\widetilde{A}_{16} = (9.5; 4.0)$	2



Initial Classifier:

6 rules. 6 fuzzy Sets. $\mathcal{S} = 0.996$.

	$\boldsymbol{x_1}$	$\boldsymbol{y}$
$R'_1:$	$\widetilde{A}'_{11} = (1.0; 0.5)$	1
$R'_2:$	$\widetilde{A}'_{12} = (2.4; 0.3)$	1
$R'_3:$	$\widetilde{A}'_{13} = (4.0; 0.3)$	1
$R'_4:$	$\widetilde{A}'_{14} = (6.0; 0.3)$	1
$R'_5:$	$\widetilde{A}'_{15} = (9.3; 1.3)$	2

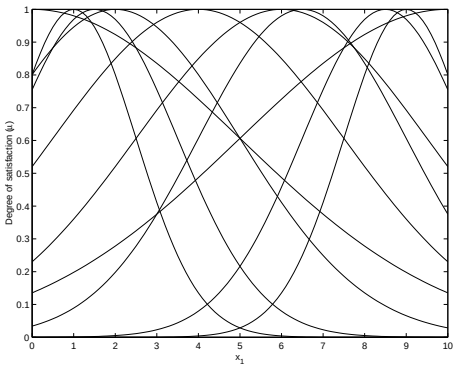


Classifier after repair algorithm:

5 rules. 5 fuzzy sets. $\mathcal{S} = 0.098$.

Table 3: Example of repair algorithm ($\eta = 0.1$, $L_{max} = 7$). Two fuzzy sets have been split; two fuzzy sets have been merged and a rule has been removed.

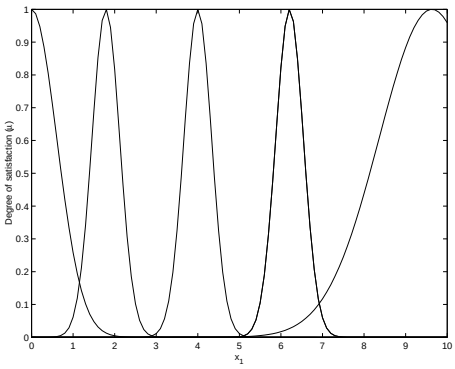
	$\boldsymbol{x_1}$	$\boldsymbol{y}$
$R_1:$	$\widetilde{A}_{11} = (0.0; 5.0)$	1
$R_2:$	$\widetilde{A}_{12} = (10.0; 5.0)$	1
$R_3:$	$\widetilde{A}_{13} = (6.0; 3.5)$	2
$R_4:$	$\widetilde{A}_{14} = (2.0; 3.0)$	1
$R_5:$	$\widetilde{A}_{15} = (1.0; 1.5)$	1
$R_6:$	$\widetilde{A}_{16} = (9.0; 1.5)$	2
$R_7:$	$\widetilde{A}_{17} = (6.5; 2.5)$	2
$R_8:$	$\widetilde{A}_{18} = (4.0; 3.5)$	1
$R_9:$	$\widetilde{A}_{19} = (8.5; 2.0)$	2
$R_{10}:$	$\widetilde{A}_{110} = (1.5; 2.0)$	1



Initial classifier:

10 rules. 10 fuzzy sets. $\mathcal{S} = 0.997$.

	$\boldsymbol{x_1}$	$\boldsymbol{y}$
$R'_1:$	$\widetilde{A}'_{11} = (0.0; 0.6)$	1
$R'_2:$	$\widetilde{A}'_{12} = (9.6; 1.3)$	1
$R'_3:$	$\widetilde{A}'_{13} = (1.8; 0.3)$	1
$R'_4:$	$\widetilde{A}'_{14} = (6.2; 0.3)$	1
$R'_5:$	$\widetilde{A}'_{15} = (4.0; 0.3)$	1
$R'_6:$	$\widetilde{A}'_{16} = (6.2; 0.3)$	2



Classifier after repair algorithm:

6 rules. 5 fuzzy sets. $\mathcal{S} = 0.074$.

Table 4: Example of repair algorithm ($\eta = 0.1$, $L_{max} = 7$). Multiple fuzzy sets have been split and mixed and multiples rules have been erased.

4.3 Initial population

The initial population (algorithm 2) is randomly generated with the following conditions:

- The individuals are uniformly distributed with respect to the number of rules with values between M_{min} and M_{max} (steps 8 to 13). This ensures a proper initial diversity in the second objective of the optimization model.
- All individuals contain at least one rule for any output value between 1 and z (steps 23 to 27).
- The rest of parameters are randomly generated within its domain.
- The repair method (algorithm 1) is applied to all individuals in order to satisfy the constraints (step 33).

4.4 Variation operators

Evolutionary computation uses probabilistic transition rules to transform the populations in order to exploit and explore the search space. They use random choice as a tool to find search space regions of likely improvement. These mechanisms have the form of variation operators that work over two individuals, making a crossover between them, or over one individual, mutating it. In order to obtain a proper exploration of the search space, the variation operators have to work at the different levels of the individuals. In our case, as the individuals represent fuzzy set classifiers, the variation operators can operate at rule level or at fuzzy set level. Furthermore, we use adaptive probabilities of crossover and mutation to realize the twin goals of maintaining diversity in the population and sustaining the convergence capacity of the evolutionary algorithm. In an adaptive evolutionary algorithm [85], the probabilities of crossover and mutation varies depending on the fitness value of the solutions. By using adaptive probabilities of variation operators, it is not necessary to set a priori the probabilities of application of the different operators. We use three crossover operators and five mutation operators which work on different levels of the fuzzy classifier. The selection of the operators is made by means of the adaptive technique that uses parameters d_I and e_I to indicate which crossover and mutation is carried out to the individual I. This technique for crossing and mutation is shown in algorithms 4 and 5 respectively. Additionally, adapting mutation is performed by using the real parameters

Algorithm 2 Initialize population

Require: l_i,u_i ($l_i < u_i$), $i = 1, \ldots, p$ {Lower and upper limits for the i real input variable}

Require: $v_i > 1$, $i = 1, \ldots, q$ {Number of classes for the i categorized input variable}

Require: $z > 1$, {Number of classes for the categorized output variable}

Require: γ_1, γ_2 ($\gamma_1 > \gamma_2 > 0$) {Minimum and maximum variance parameter}

Require: $\delta > 0$ {Number of crossing operators}

Require: $\epsilon > 0$ {Number of mutation operators}

Require: M_{min},M_{max} ($z \leq M_{min} \leq M_{max}$) {Minimum and maximum number of rules}

Require: $N \geq z$ {Number of individuals in the population}

$P \leftarrow$ Empty Population
for $i = 1$ to p **do**
    $\alpha_i \leftarrow \frac{u_i - l_i}{\gamma_1}$ {Lower limit for variance for i real input variable}
    $\beta_i \leftarrow \frac{u_i - l_i}{\gamma_2}$ {Upper limit for variance for i real input variable}
end for
for $index = 1$ to N **do**
    $I \leftarrow$ new Individual
    $nClasses \leftarrow M_{max} - M_{min} + 1$
    if $index \leq \frac{N}{nClasses}$ **then**
        $M_I \leftarrow (index \bmod nClasses) + 1$
    else
        $M_I \leftarrow$ Int *Random*(0,M_{max})
    end if
    {Random rule R_j^I}
    for $j = 1$ to M_I **do**
        {Random fuzzy set $\widetilde{A}_{ij}^I$}
        for $i = 1$ to p **do**
            $a_{ij}^I \leftarrow$ Real *Random*(l_i,u_i)
            $\sigma_{ij}^I \leftarrow$ Real *Random*(α_i,β_i)
        end for
        {Random Discrete Input variables}
        for $i = 1$ to q **do**
            $b_{ij}^I \leftarrow$ Int *Random*(1,v_i)
        end for
        {Random Discrete Output variable}
        if $j < z$ **then**
            $c_j^I = j$
        else
            $c_j^I \leftarrow$ Int *Random*(1,z)
        end if
    end for
    {Random Discrete and Real parameters for adaptive variation}
    $d_I \leftarrow$ Int *Random*(0,δ)
    $e_I \leftarrow$ Int *Random*(0,ϵ)
    $e_c^I \leftarrow$ Real *Random*(0,1)
    $e_v^I \leftarrow$ Real *Random*(0,1)
    Repair I
    Add I to population P
end for
return P

e_c^I and e_v^I that define the amplitude of the mutation of the centers and variances of the gaussian fuzzy sets, respectively, as algorithms 9 and 10 show.

Algorithm 3 is used to generate two children from two parents by adaptive crossing, adaptive mutation and repair.

Algorithm 3 Variation

Require: $Parent1$, $Parent2$ {Individuals to vary}
1: $Child1 \leftarrow Parent1$
2: $Child1 \leftarrow Parent2$
3: Adaptive crossover $Child1$, $Child2$
4: Adaptive mutation $Child1$
5: Adaptive mutation $Child2$
6: Repair $Child1$
7: Repair $Child2$
8: **return** $Child1, Child2$

Algorithm 4 Adaptive crossover

Require: I, J {Individuals to cross}
Require: p_v $(0 < p_v < 1)$ {Probability of variation}
Require: $\delta > 0$ {Number of different crossover operators ($\delta = 3$ in our case)}
1: **if** A random Bernoulli variable of probability p_v takes the value 1 **then**
2: $d_I \leftarrow$ Int *Random*(0,δ)
3: **end if**
4: $d_J \leftarrow d_I$
5: Carry out the type of crossover specified by d_I:
{0: No cross}
{1: Fuzzy set crossover}
{2: Rule crossover}
{3: Rule incremental crossover}

The following sections describe the different each of the crossover and mutation operators.

4.4.1 Fuzzy set crossover

Given two individuals, I and J, this operator exchanges two fuzzy sets randomly selected from the two individuals (algorithm 6).

4.4.2 Rule crossover

Given two individuals, I and J, this operator exchanges two rules randomly selected from the two individuals (algorithm 7).

Algorithm 5 Adaptive mutation

Require: I {Individual to mutate}

Require: p_v ($0 < p_v < 1$) {Probability of variation}

Require: $\epsilon > 0$ {Number of different mutation operators ($\epsilon = 5$ in our case, $\epsilon = 4$ if there were no discrete input variables)}

1: **if** A random Bernoulli variable of probability p_v takes the value 1 **then**

2: $e_I \leftarrow$ Int *Random*(0,ϵ)

3: **end if**

4: Carry out the type of mutation specified by e_I:

{0: No mutation}

{1: Gaussian set center mutation}

{2: Gaussian set variance mutation}

{3: Fuzzy set mutation}

{4: Rule incremental mutation}

{5: Integer mutation}

Algorithm 6 Fuzzy set crossover

Require: I, J {Individuals to cross}

Require: $p > 0$ {Number of real input variables}

1: $i \leftarrow$ Int *Random*(0,p)

2: $j \leftarrow$ Int *Random*(0,M_I)

3: $k \leftarrow$ Int *Random*(0,M_J)

4: Exchange fuzzy sets $\widetilde{A}_{ij}^I$ and $\widetilde{A}_{ik}^J$

Algorithm 7 Rule crossover

Require: I, J {Individuals to cross}

1: $j \leftarrow$ Int *Random*(0,M_I)

2: $k \leftarrow$ Int *Random*(0,M_J)

3: Exchange rules R_j^I and R_k^J

4.4.3 Rule incremental crossover

Given two individuals, I and J, this operator adds to each individual a rule randomly selected from the other individual if its number of rules is less than the maximum number of rules (algorithm 8):

Algorithm 8 Rule incremental crossover

Require: I, J {Individuals to cross}

Require: $M_{max} > 1$ {Maximum number of rules}

if $M_I < M_{max}$ **then**
    $j \leftarrow$ Int *Random*(0,M_J)
    add R_j^J to individual I
end if
if $M_J < M_{max}$ **then**
    $i \leftarrow$ Int *Random*(0,M_I)
    Add R_i^I to individual J
end if

4.4.4 Gaussian set center mutation

Given an individual I, this operator carries out the mutation of the center of a fuzzy set randomly selected from the individual. Additionally, all the fuzzy sets which are equal to the selected fuzzy set, are mutated in the same way (algorithm 9).

Algorithm 9 Gaussian set center mutation

Require: I {Individual to mutate}

Require: l_i,u_i ($l_i < u_i$), $i = 1, \ldots, p$ {Low and Upper limits for i real input variable}

Require: p_v ($0 < p_v < 1$) {Probability of variation}

if A random Bernoulli variable of probability p_v takes the value 1 **then**
    $e_c^I \leftarrow$ Int *Random*(0,1)
end if
$i \leftarrow$ Int *Random*(0,*p*)
$j \leftarrow$ Int *Random*(0,M_I)
$\zeta \leftarrow$ Real *Random*(0,1)
$a \leftarrow a_{ij}^I + e_c^I \left(l_i + \zeta \left(u_i - l_i \right) - a_{ij}^I \right)$
for $k = 1$ to M_I **do**
    if $k \neq j$ and $a_{ik}^I = a_{ij}^I$ **then**
        $a_{iK}^I \leftarrow a$
    end if
end for
$a_{ij}^I \leftarrow a$

4.4.5 Gaussian set variance mutation

Given an individual I, this operator carries out the mutation of the variance of a fuzzy set randomly selected from the individual. Additionally, all the fuzzy sets which are equal to the selected fuzzy set, are mutated in the same way (algorithm 10).

Algorithm 10 Gaussian set variance mutation

Require: I {Individual to mutate}

Require: l_i,u_i ($l_i < u_i$), $i = 1, \ldots, p$ {Low and Upper limits for i real input variable}

Require: γ_1, γ_2 ($\gamma_1 > \gamma_2 > 0$) {Minimum and maximum variance parameter}

Require: p_v ($0 < p_v < 1$) {Probability of variation}

if A random Bernoulli variable of probability p_v takes the value 1 **then**
    $e_v^I \leftarrow$ Int *Random*(0,1)
end if
$i \leftarrow$ Int *Random*(0,*p*)
$j \leftarrow$ Int *Random*(0,M_I)
$\zeta \leftarrow$ Real *Random*(0,1)
$\alpha_i = \frac{u_i - l_i}{\gamma_1}$
$\beta_i = \frac{u_i - l_i}{\gamma_2}$
$\sigma \leftarrow \sigma_{ij} + e_v \left(\alpha_i + \zeta \left(\beta_i - \alpha_i\right) - \sigma_{ij}\right)$
for $k = 1$ to M_I **do**
    if $k \neq j$ and $\sigma_{ik}^I = \sigma_{ij}^I$ **then**
        $\sigma_{ik}^I \leftarrow \sigma$
    end if
end for
$\sigma_{ij}^I \leftarrow \sigma$

4.4.6 Fuzzy set mutation

Given an individual I, this operator exchanges two fuzzy sets randomly selected from the individual (algorithm 11).

Algorithm 11 Fuzzy set mutation

Require: I {Individual to mutate}

Require: $p > 0$ Number of real input variables

$i \leftarrow$ Int *Random*(0,*p*)
$j \leftarrow$ Int *Random*(0,M_I)
$k \leftarrow$ Int *Random*(0,M_I)
$\widetilde{A}_{ij}^I \leftarrow \widetilde{A}_{ik}^I$

4.4.7 Rule incremental mutation

Given an individual I, this operator adds a new rule to the individual if the number of rules of I is less than the maximum number of rules (algorithm 12).

Algorithm 12 Rule incremental mutation

Require: I {Individual to mutate}

Require: M_{max} {Maximum number of rules}

1: **if** $M_I < M_{max}$ **then**

2: Add a new random rule to I

3: **end if**

4.4.8 Integer mutation

Given an individual I, this operator carries out the uniform mutation of an integer random selected value from the individual (algorithm 13).

Algorithm 13 Integer mutation

Require: I {Individual to mutate}

Require: $v_i > 1, i = 1, \ldots, q$ {Number of values for i categorized input variable}

1: $i \leftarrow$ Int *Random*(0,q)

2: $j \leftarrow$ Int *Random*(0,M_I)

3: $b_{ij}^I \leftarrow$ Int *Random*(1,v_i)

4.5 Niched pre-selection algorithm

The *pre-selection* mechanism [86] was one of the first efforts to induce niched-like behavior in evolutionary algorithms. In this scheme, an offspring replaces the parent if the offspring's fitness exceeds that of the inferior parent. In this way, diversity is maintained because individuals replace individuals similar to themselves (one of their parents). This technique is, implicitly, a niche formation technique and also an elitist strategy [51].

We propose an adaptation of the pre-selection scheme for multi-objective optimization problems. An explicit niche formation technique is incorporated to improve the diversity of individuals of the population. The search space is divided into n niches so that an individual I belongs to a niche $\mathcal{N}(I) \in \{1, \ldots, n\}$. The explicit niche formation technique ensures that the number of individuals in each niche is greater than or equal to NS_{min} and smaller than or equal to NS_{max}. That is, given a population P and n niches, the following diversity property is

satisfied:

$$\forall i \in \{1, \ldots, n\},\ NS_{min} \leq \mathcal{NC}(P, i) \leq NS_{max} \tag{8}$$

where $\mathcal{NC}(P, i)$ (*niche count*) is the number of individuals of the population P so that they belong to niche i, i.e., $\mathcal{N}(I) = i$.

To solve the multi-objective optimization problem of the model (3), we use a number of niches $n = M_{max} - M_{min} + 1$ and an individual I belongs to the niche $\mathcal{N}(I)$ determined by its number of rules M_I as follows:

$$\mathcal{N}(I) = M_I - M_{min} + 1 \tag{9}$$

In this way, for each number of rules $i \in \{M_{min}, \ldots, M_{max}\}$, the number of individuals of the population with i rules is between NS_{min} and NS_{max}.

The selection and generational replacement process is the following (algorithm 14): two individuals, $Parent1$ and $Parent2$, are randomly selected from the population P. These individuals are crossed and mutated NC (number of children pairs) times, where $NC > 0$ is an algorithm parameter, producing two groups of NC individuals. A random non-dominated individual of the first group, $Best1$, replaces $Parent1$ if $Best1$ dominates $Parent1$ and the diversity property (8) is satisfied for population $P - \{Parent1\} \bigcup \{Best1\}$. Similarly, a random non-dominated individual of the second group, $Best2$, replaces $Parent2$ if $Best2$ dominates $Parent2$ and the diversity property (8) is satisfied for population $P - \{Parent2\} \bigcup \{Best2\}$.

4.6 ENORA: Elitist Pareto-based multi-objective evolutionary algorithm for diversity reinforcement

ENORA is an elitist Pareto-based multi-objective evolutionary algorithm that uses a $(\mu + \lambda)$ survival. The $(\mu + \lambda)$ survival, where μ corresponds to the population size $popsize$ and λ refers to the number of children created, was originally developed in [87] as an *E*volution Strategy, using selection, adapting mutation and a population of size one, called $(1 + 1) - ES$. Recombination and populations with more than one individual were later introduced in [88]. The $(\mu + \lambda)$ technique allows the μ best children and parents to survive and is, therefore, an elitist method. ENORA uses a $(\mu + \lambda)$ survival with $\mu = \lambda = popsize$, binary tournament selection, recombination and adapting mutation for multi-objective evolutionary optimization (algorithm 15).

Algorithm 14 Niched pre-selection algorithm for multi-objective optimization

Require: $T > 1$ {Number of iterations}

Require: $N > 1$ {Number of individuals in the population}

Require: $NC > 0$ {Number of children}

Require: NS_{min}, NS_{max} $(0 < NS_{min} < NS_{max})$ {Minimum and maximum size of niche}

Initialize P with N individuals
Evaluate all individuals of P
$t \leftarrow 0$
while $t < T$ **do**
    $Parent1 \leftarrow$ Random Selection from P
    $Parent2 \leftarrow$ Random Selection from P
    $Best1 \leftarrow Parent1$
    $Best2 \leftarrow Parent2$
    $i \leftarrow 0$
    while $i < NC$ **do**
        $Child1, Child2 \leftarrow$ Variation $Parent1, Parent2$
        Evaluate $Child1$
        Evaluate $Child2$
        if $Child1$ dominates $Best1$ and $P - \{Best1\} \bigcup \{Child1\}$ satisfies the diversity property (8) **then**
            $P \leftarrow P - \{Best1\} \bigcup \{Child1\}$
            $Best1 \leftarrow Child1$
        end if
        if $Child2$ dominates $Best2$ and $P - \{Best2\} \bigcup \{Child2\}$ satisfies the diversity property (8) **then**
            $P \leftarrow P - \{Best2\} \bigcup \{Child2\}$
            $Best2 \leftarrow Child2$
        end if
        $i \leftarrow i + 1$
    end while
    $t \leftarrow t + 1$
end while
return Non dominated individuals from P

Algorithm 15 implements a $(\mu + \lambda)$ strategy for multi-objective optimization. The algorithm begins with the initialization and evaluation of a population P of N individuals.

For each of the T generations, a pair of parents are selected by a *Binary Tournament Selection* from the population P (algorithm 16). This selection algorithm returns the best from two random individuals according to the Rank-Crowding-Better Function (algorithm 17). With this function, an individual I is better than an individual J if its rank is better (lower) than the rank of individual J in the population P. The rank of an individual I in a population P, $rank\,(P, I)$, is the *non-domination level* of the individual I among the individuals J of the population P so that $slot\,(I)$=$slot\,(J)$, where the $slot$ function is calculated according to equation (10) where $d = \left\lfloor \sqrt[n-1]{N} \right\rfloor$ and h_j^I is the objective function f_j^I normalized in $[0, 1]$.

$$
\begin{aligned}
slot\,(I) &= \sum_{j=1}^{n-1} d^{j-1} \lfloor d \frac{\alpha_j^I}{\pi/2} \rfloor \\
\alpha_j^I &= \begin{cases} \frac{\pi}{2} & if\ h_j^I = 0 \\ \arctan(\frac{h_{j+1}^I}{h_j^I}) & if\ h_j^I \neq 0 \end{cases}
\end{aligned}
\tag{10}
$$

If two individuals have the same rank, the best individual is the individual with the greater crowding distance. The crowding distance of an individual I in a population P is a measure of the search space around individual I which is not occupied by any other individual in the population P. This quantity serves as an estimate of the perimeter of the cuboid formed by using the nearest neighbors as the vertices. The crowding distance is calculated as follows:

$$
crowding\_distance\,(P, I) = \begin{cases} \infty,\ if\ f_j^i = f_j^{max}\ or\ f_j^i = f_j^{min}\ for\ any\ j \\ \sum_{j=1}^{n} \frac{f_j^{sup_j^I} - f_j^{inf_j^I}}{f_j^{max} - f_j^{min}},\ in\ other\ case \end{cases}
\tag{11}
$$

where $f_j^{max} = \max_{I \in P} \left\{ f_j^i \right\}$, $f_j^{min} = \min_{I \in P} \left\{ f_j^i \right\}$, $f_j^{sup_j^I}$ is the value of the jth objective for the individual higher adjacent in the jth objective to the individual I, and $f_j^{inf_j^I}$ is the value of the jth objective for the individual lower adjacent in the jth objective to the individual I.

The selected pair of parents is crossed, mutated, repaired, evaluated and added to an initially empty auxiliary population Q. This process is repeated until Q contains a number N of individuals. An auxiliary population R is obtained with the union of the populations P and Q. Next, the rank of all individuals in the population R is calculated (algorithm 17). Finally, the N

best individuals of R according to the Rank-Crowding Better Function (algorithm 17) survive to the next generation.

Algorithm 15 $(\mu + \lambda)$ strategy for multi-objective optimization

Require: $T > 1$ {Number of iterations}
Require: $N > 1$ {Number of individuals in population}
Initialize P with N individuals
Evaluate all individuals of P
$t \leftarrow 0$
while $t < T$ **do**
    $Q \leftarrow \emptyset$
    $i \leftarrow 0$
    while $i < N$ **do**
        $Parent1 \leftarrow$ Binary tournament selection from P
        $Parent2 \leftarrow$ Binary tournament selection from P
        $Child1, Child2 \leftarrow$ Variation $Parent1, Parent2$
        Evaluate $Child1$
        Evaluate $Child2$
        $Q \leftarrow Q \bigcup \{Child1, Child2\}$
        $i \leftarrow i + 2$
    end while
    $R \leftarrow P \bigcup Q$
    $P \leftarrow N$ Best individuals from R according to the Rank-crowding better function in population R
    $t \leftarrow t + 1$
end while
return Non dominated individuals from P

Algorithm 16 Binary tournament selection

Require: P {Population}
$I \leftarrow$ Random selection from P
$J \leftarrow$ Random selection from P
if I is better than J according to the Rank-crowding better function in population P **then**
    return I
else
    return J
end if

4.7 NSGA-II: elitist non-dominated sorting genetic algorithm

NSGA-II [29] is an elitist Pareto-based multi-objective evolutionary algorithm which improves the previous NSGA algorithm by incorporating an explicit diversity technique. NSGA-II is, perhaps, one of the most used Pareto-based multi-objective evolutionary algorithms described in the literature.

Algorithm 17 Rank-crowding better function

Require: P {Population}
Require: I, J {Individuals to compare}
1: **if** $rank\,(P, I) < rank\,(P, J)$ **then**
2: **return** $True$
3: **end if**
4: **if** $rank\,(P, J) < rank\,(P, I)$ **then**
5: **return** $False$
6: **end if**
7: **return** $crowding\_distance\,(P, I) > crowding\_distance\,(P, J)$

NSGA-II uses, as ENORA, a $(\mu + \lambda)$ strategy (algorithm 15) with a binary tournament selection (algorithm 16) and a rank crowding better function (algorithm 17). The difference between NSGA-II and ENORA is how the calculation of the rank of the individuals in the population is performed. In ENORA, the rank of an individual in a population is the non-domination level of the individual in its slot, whereas in NSGA-II the rank of an individual in a population is the non-domination level of the individual in all the population.

4.8 Linguistic labeling

The next step in the modeling process is to associate of fuzzy sets with linguistic labels that are easily interpretable. To do this, we propose using algorithm (18) for linguistic labeling. As inputs this algorithm requires the classifier fuzzy sets $\widetilde{A}_{ij} = (a_{ij}, \sigma_{ij})$, $i = 1, \ldots, p$, $j = 1, \ldots, M$ and the variable domains $[u_i, l_i]$, $i = 1, \ldots, p$. The output of the algorithm is the set of linguistic labels assigned to each fuzzy set. For each variable x_i, $i = 1, \ldots, p$, the number of linguistic labels N_i is equal to the maximum number of fuzzy sets, identified by $L_{min} = \frac{u_i - l_i}{S_i}$, if the number of different fuzzy sets for that variable is greater than one; if there is only one different fuzzy set for that variable, the number of linguistic labels N_i for that variable is equal to 1.

When there is only one different fuzzy set for the i variable and the value N_i is equal to 1, the condition $x_i \; is \; \widetilde{A}_{ij}, \; j = 1, \ldots, M$, denoted as *don't care condition* in the literature [63], does not affect the computation of the activation degree. In other words, the variable x_i is not taken into account for the classifier.

The labeling algorithm identifies a set of N_i linguistic labels, namely L_{ik}, and its center values C_{ik}, $k = 1, \ldots, N_i$. Eventually, each fuzzy set $\widetilde{A}_{ij}$ is associated with the linguistic label L_{ij} which minimizes the distance between the centers C_{ik} and a_{ij}, $k = 1, \ldots, N_i$.

Algorithm 18 Linguistic labeling algorithm

Require: $\widetilde{A}_{ij} = (a_{ij}, \sigma_{ij}), i = 1, \ldots, p, j = 1, \ldots, M$

Require: l_i,u_i ($l_i < u_i$), $i = 1, \ldots, p$ {Low and Upper limits for i real input variable}

for $i = 1$ to p **do**
  if There is only one different fuzzy set for i real input variable **then**
    $N_i \leftarrow 1$
  else
    $S_i \leftarrow \min_{\substack{j=1,\ldots,M \\ k=1,\ldots,M \\ \widetilde{A}_{ij} \neq \widetilde{A}_{ik}}} |a_{ij} - a_{ik}|$
    $N_i \leftarrow \max\left(2, \left\lceil \frac{u_i - l_i}{S_i} - 0.5 \right\rceil\right)$
  end if
  switch (N_i)
  case 1**:** $L_i \leftarrow \{DCC\}$ {Don't care condition}
  case 2**:** $L_i \leftarrow \{L, H\}$
  case 3**:** $L_i \leftarrow \{L, M, H\}$
  case 4**:** $L_i \leftarrow \{L, ML, MH, H\}$
  case 5**:** $L_i \leftarrow \{L, ML, M, MH, H\}$
  case 6**:** $L_i \leftarrow \{VL, L, ML, MH, H, VH\}$
  case 7**:** $L_i \leftarrow \{VL, L, ML, M, MH, H, VH\}$
  end switch
  for $k = 1$ to N_i **do**
    $C_{ik} \leftarrow l_i + (k - 0.5)\frac{u_i - l_i}{N_i}$
  end for
  for $j = 1$ to M **do**
    $l \leftarrow \arg_k \min_{k=1}^{N_i} |a_{ij} - C_{ik}|$
    $LABEL_{ij} \leftarrow L_{il}$
  end for
end for
return $LABEL_{ij}, i = 1, \ldots, p, j = 1, \ldots, M$

4.9 Decision-making process

Let $S = \{s_1, \ldots, s_D\}$ be the set of non-dominated solutions so that $\mathcal{CR}(s_i) \geq CR_{min}$, $(i = 1, \ldots, D)$, where CR_{min} is the minimum classification rate acceptable by the decision maker. In the decision-making process (algorithm 19), the most accurate solution is chosen. If this solution is easily interpretable (transparent and compact), then the solution is shown as output. In any other case, the chosen solution is rejected and the process is repeated until a satisfactory solution is found. In this last case, a new run of the MOEA is required with a greater minimum variance parameter.

Algorithm 19 Decision-making algorithm

Require: $S = \{s_1, \ldots, s_D\}$ {Set of non-dominated solutions}
Require: CR_{min} {Minimum acceptable classification rate}
Require: M_{max} {Maximum acceptable number of rules}
Remove from S all solutions s_i such that $\mathcal{CR}(s_i) < CR_{min}$
while $\mathcal{S} \neq \emptyset$ **do**
    Select s_i with greatest $\mathcal{CR}(s_i)$ {Most accurate solution s_i}
    if $\mathcal{NR}(s_i) \leq M_{max}$ and $LABEL(s_i) \neq \emptyset$ {s_i sufficiently interpretable} **then**
        return s_i
    else
        Remove s_i from S
    end if
end while
return EMPTY SOLUTION {There is no satisfactory solution}

5 Experiments and results

In this section, two data sets have been used to carry out the experiments. One of them, the Medical data set, is the main goal of the work, and it is described in section 5.2. The other one, the Iris data set, is a typical data set for classification problems, described in section 5.1. Three different sets of experiments and results are shown. The first set (section 5.3) compares the performance of the three presented MOEA: niched pre-selection, NSGA-II and ENORA. Experiments involved over 100 runs for each MOEA using as test problems the Iris data set and a medical data set. A statistical analysis of the results for the hypervolume metric is shown. The following set of experiments (section 5.4) compares the performance of the classifiers obtained with the algorithms using a multi-objective cross-validation technique proposed by the authors. Minimum, maximum and medium classification rate values are shown for solutions

obtained with a different number of rules for each algorithm. In the case of the Medical data set, minimum, maximum and average sensitivity and specificity values are also shown. Sensitivity and specificity are performance measures for binary classifiers. Sensitivity (true positive rate) measures the proportion of actual positives which are correctly identified. In medicine, sensitivity of a test is the probability of a positive test, given that the patient is ill. On the other hand, specificity is defined as the proportion of actual negatives which are correctly identified. In medicine, specificity of a test is defined as the probability of a negative test, given that the patient is not ill. A detailed explanation about sensitivity and specificity in diagnostic test may be found in [89].

For the binary classifiers obtained in this paper for the medical data set, we show average sensitivity and average specificity values, which define the receiver operating characteristic (ROC) space [90]. In classifiers used for survival models in medicine and clinical medicine, ROC space is a fundamental tool [91, 92]. ROC space is defined by *f*alse positive rate, equivalent to $sensitivity$, and *t*rue positive rate, equivalent to $1 - specificity$ as x and y axes respectively. Therefore, each prediction result for a binary classifier represents one point in the ROC space. The best classifier obtains the point $(0, 1)$ of the ROC space, this is, 100% sensitivity (no false negatives) and 100% specificity (no false positives). This is called a *perfect classification*. Besides, the classifiers' performance is compared with other non-evolutionary machine learning techniques, such as artificial neural networks, probabilistic models, decision trees and analogy-based approaches. Finally, in section 5.5, the third set of experiments analyzes a solution obtained with the proposed methodology shown in section 3.

5.1 Iris data set

The Iris data set is perhaps the best known data set to be found in data classification literature. Fisher's paper [93] is a classic in the field and is still referenced frequently. The multivariate data set contains 3 classes of 50 instances each (150 instances), where each class refers to a type of iris plant. One class is linearly separable from the other two, which are not linearly separable from each other. This is an exceedingly simple domain with 4 real attributes (no missing values) where the predicted attribute is the class of iris plant.

Although Iris data set is unrelated to medicine, it has been chosen in this paper because it is one of the best studied database to be found in classification literature and there are multiple authors which have reported different classification algorithms and the solutions obtained.

For example, in [94], solutions from 10 different techniques, with classifications rates between 95.57% and 97.33% are compared. Given that our proposal is valid for any classification problem, not only in a medical domain, we understand it is interesting to show the results obtained for Iris data set in order to make it possible comparisons with any other classification algorithm. Table 5 summarizes the data set characteristics.

Name	**Description**	**Type**	**Attributes**	**Limits**
Sepal length	Sepal length in cm.	Real	x_1	$4.3 - 7.9$
Sepal width	Sepal width in cm.	Real	x_2	$2.0 - 4.4$
Petal length	Petal length in cm.	Real	x_3	$1.0 - 6.9$
Petal width	Petal width in cm.	Real	x_4	$0.1 - 2.5$
Type of flower	Setosa (1), Versicolor (2), Virginica (3)	Discrete	y	$1\ ,\ 2\ ,\ 3$

Table 5: Attribute information for Iris data set.

5.2 Medical data set

In order to test the suitability of our proposal in the medical domain, we evaluate the proposed fuzzy classification methodology in the intensive care burn unit domain (ICBU). ICBUs are hospital services that provides health care to severe burn patients. The patients' data set used were collected from the Health Information System of the ICBU from 1999 to 2002. For the study, the clinical team selected 99 records from this data set, taking into account demographic and etiological diversity criteria. The physicians also summarized the patients' records according to the most clinically relevant evidence for establishing survival. Table 6 shows the 18 patients' parameters considered in this study [4].

Once the parameter set has been fixed, it is necessary to establish the importance of each parameter for survival. For this goal, we used the Mutual Information feature selection method [95], based on Shannon's Entropy [96]. The weights depicted in table 6 show the results obtained and, for this study, the parameters with a weighting of 0 are SAPS II, weight, age and renal co-morbidity. We are therefore left with a problem of $p = 2$ real inputs and $q = 11$ categorized inputs. The two real inputs considered are within the range $[0, 85]$, while all the categorized inputs (and output) are boolean type.

Name	Description	Type	Weight	Attribute	Limits
Total	Total burnt surface %	Real	0.164613	x_1	$0 - 85$
Prof	Deeply burnt surface %	Real	0.115335	x_2	$0 - 85$
SAPS II	Severity score	Real	0	$-$	
Weight	Patient's weight	Real	0	$-$	
Age	Patient's age	Real	0	$-$	
Sex	Patient's sex	Boolean	0.069403	w_1	1 , 2
Inh	Use of inhibitors	Boolean	0.067112	w_2	1 , 2
Bacteremia	Presence of bacteria in blood	Boolean	0.023324	w_3	1 , 2
Pneumonia	Presence of pulmonary infection	Boolean	0.207799	w_4	1 , 2
Wound-Infect	Infection by surgical wound	Boolean	0.034936	w_5	1 , 2
Co-Card	Previous cardiopathy	Boolean	0.008141	w_6	1 , 2
Co-Respir	Previous respiratory problems	Boolean	0.007035	w_7	1 , 2
Co-Liver	Previous liver problems	Boolean	0.024223	w_8	1 , 2
HBP	High blood pressure	Boolean	0.001650	w_9	1 , 2
Diabetes	Diabetic patient	Boolean	0.000541	w_{10}	1 , 2
aids-drugs	HVI drug consumption	Boolean	0.025781	w_{11}	1 , 2
Co-Renal	Previous kidney problems	Boolean	0	$-$	
Death	Prognosis of death	Boolean	$-$	y	1 , 2

Table 6: Patients' parameters considered: name, description, type, weight, associated input attribute and limits. For the boolean variables, the values 1 and 2 correspond to false and true boolean values, respectively.

5.3 Comparison of the algorithms

In this section, the niched pre-selection algorithms, NSGA-II and ENORA, are compared and the results are shown. The aim of this set of experiments was to identify best performing algorithm.

The algorithms were run 100 times with the parameters shown in table 7 for both Iris and Medical data sets. These parameters have been set according to the following criteria:

- Size of population $N = 100$ is a typically used value in EA community.
- Number of evaluations $NoEvaluations = 10^5$ has been set after a previous experimental process. In this process, a number of evaluations around this value indicates that the algorithms may reach an appropriate convergence.

- A minimum number of rules $M_{min} = 3$ and a maximum number of rules $M_{max} = 10$ are values usually reported in literature for the Iris data set. For the case of the Medical data set, values $M_{min} = 10$ and $M_{max} = 20$ have been supplied by a physician.

- $\gamma_1 = 30$ value ensures that interpretable models, with a number of linguistic labels less or equal than 7, are obtained with a maximum overlapping of 10% ($g_s = 0.1$) of the fuzzy sets. We understand that a greater number of linguistic labels or a greater overlapping value make a model hardly interpretable by an human. This maximum number of 7 linguistic labels together with a maximum overlapping of 10% are usually accepted by the scientific community.

- $\gamma_2 = 2$ value ensures that at least 47.72% of any gaussian fuzzy set will be within the variable domain. The rest of fuzzy sets are not considered.

- $p_v = 0.1$ has been set by a previous experimental process in which 100 different values uniformly distributed in $[0.01, 1.00]$ have been checked.

- $g_s = 0.1$ indicates a maximum overlapping of the fuzzy sets of a 10%. This value is usually accepted in the scientific community in order to get interpretable models. This value together with the maximum number of linguistic labels 7 imply the choice of $\gamma_1 = 30$, as explained above.

- The parameter NC (number of children) for the niched pre-selection algorithm is related with the selection pressure of the algorithm. High values of NC imply a high selection pressure, whereas low values of NC imply a low selection pressure. Similarly to the choice of other parameters, NC has been set by a previous experimental process to obtain an adequate selection pressure, neither too low nor too high.

- The parameters NS_{min} and NS_{max} establish the minimum and maximum number of individuals for each niche, respectively. Given that the number of niches is $n = M_{max} - M_{min} + 1$ and the population size is $N = 100$, values of $NS_{min} = 5$ and $NS_{max} = 35$ are appropriated values for the problems Iris ($n = 8$) and Medical ($n = 11$) in order to preserve a good diversity regarding to the number of rules of the models.

To compare the algorithms, we used the hypervolume indicator (ν), which calculates the fraction of objective space not dominated by any of the solutions obtained by the algorithm

$N = 100$
$No\ Evaluations = 10^5$
$M_{min} = 3(Iris), 10(Medical)$
$M_{max} = 10(Iris), 20(Medical)$
$\gamma_1 = 30$
$\gamma_2 = 2$
$p_v = 0.1$
$g_s = 0.1$
$NC = 10$ (Number of Children - niched pre-selection)
$NS_{min} = 5$ (Minimum size of niche - niched pre-selection)
$NS_{max} = 35$ (Maximum size of niche - niched pre-selection)

Table 7: Parameters used to run the algorithms.

[29, 97, 98]. Although other MOEA performance metrics can be used [29], we have chosen the hypervolume metric for this work. The hypervolume metric measures, simultaneously, both diversity and optimality of the non-dominated solutions. The hyper volume metric does not require the use of an optimal population, which is not available for our test problems. Other metrics, such as *error ratio*, *generational distance*, *maximum Pareto-optimal front error*, *spread*, *maximum spread*, or *chi-square-like deviation*, need an optimal population, which makes impossible their calculation for our test problems. Additionally, other metrics as *spacing*, only measure the uniformity of the non-dominated solutions and do not take into account neither the extent of spread nor the optimality. Figures 2 and 3 depict the evolution of medium hypervolume of 100 runs for classifying Iris and medical data sets.

Tables 8 and 9 show the statistical values and boxplots over the hypervolume metric, ν, for Iris and Medical data sets, respectively.

The confidence intervals of 95% for the mean obtained with the t-test show that, for the Iris data set, niched pre-selection provided better values than ENORA and NSGA-II. The niched pre-selection algorithm therefore performed better than NSGA-II and ENORA, according to the hypervolume indicator. The t-test is robust with samples of more than 30 individuals, and so the results are significant, leading us to conclude that the differences between the hypervolume values obtained with the algorithms are statistically significant. Nevertheless, the statistics for the Medical data set are not conclusive and there is no statistically significant difference between the algorithms, although the medium, minimum and maximum values obtained by ENORA are

	Niched pre-selection	NSGA-II	ENORA
Minimum	0.0010	0.0005	0.0005
Maximum	0.0079	0.0269	0.0136
Mean	0.0066	0.0102	0.0070
S.D.	0.0015	0.0044	0.0026
C.I. Low	0.0063	0.0093	0.0065
C.I. High	0.0069	0.0111	0.0076

S.D = Standard deviation of mean

C.I. = Confidence interval for the mean (95%)

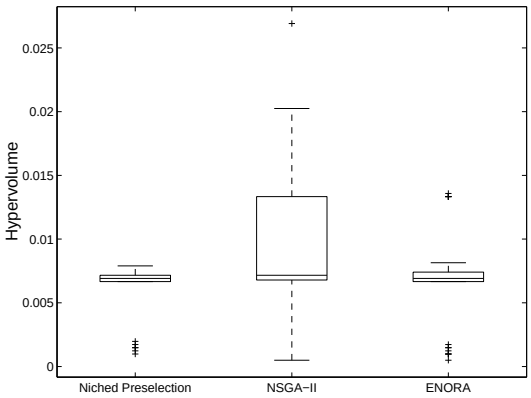


Table 8: Statistics and boxplots for the hypervolume obtained with 100 runs of the algorithms for classifying Iris data set.

better than the values obtained by the other algorithms. The box plots confirm these conclusions.

The statistical analysis shows that, for the type of multi-objective problem being considered, a Pareto search based on the partition of the search space into radial slots is more efficient than general search strategies based solely on diversity functions, such as NSGA-II, or based on diversity schemes involving the explicit formation of niches, such as niched pre-selection.

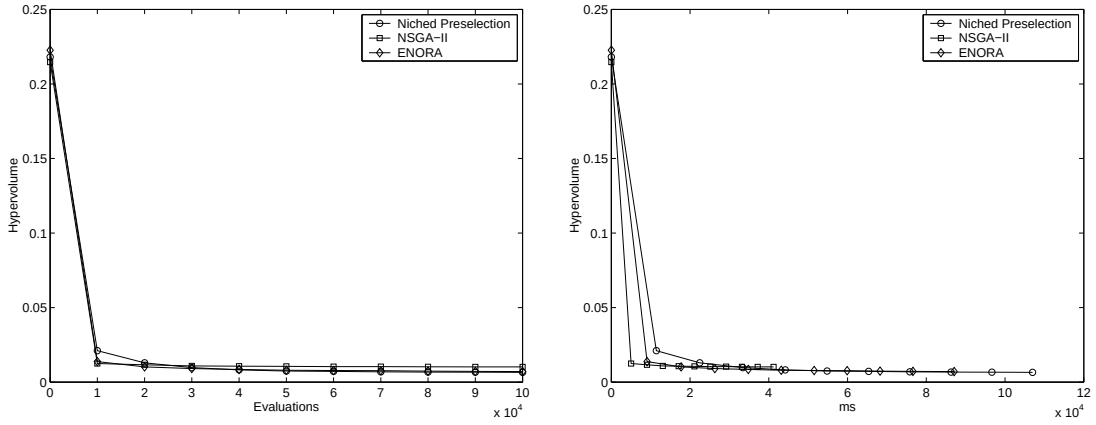


Figure 2: Evolution of medium hypervolume of 100 runs of the algorithms for classifying Iris data set.

	Niched pre-selection	NSGA-II	ENORA
Minimum	0.2679	0.2663	0.2531
Maximum	0.3307	0.3541	0.3381
Mean	0.3100	0.2992	0.2821
S.D.	0.0081	0.0176	0.0154
C.I. Low	0.3084	0.2957	0.2791
C.I. High	0.3116	0.3027	0.2852

S.D = Standard deviation of mean

C.I. = Confidence interval for the mean (95%)

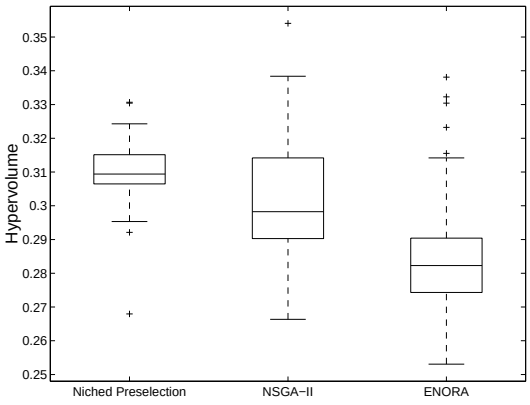


Table 9: Statistics and boxplots for the hypervolume obtained with 100 runs of the algorithms for classifying Medical data set.

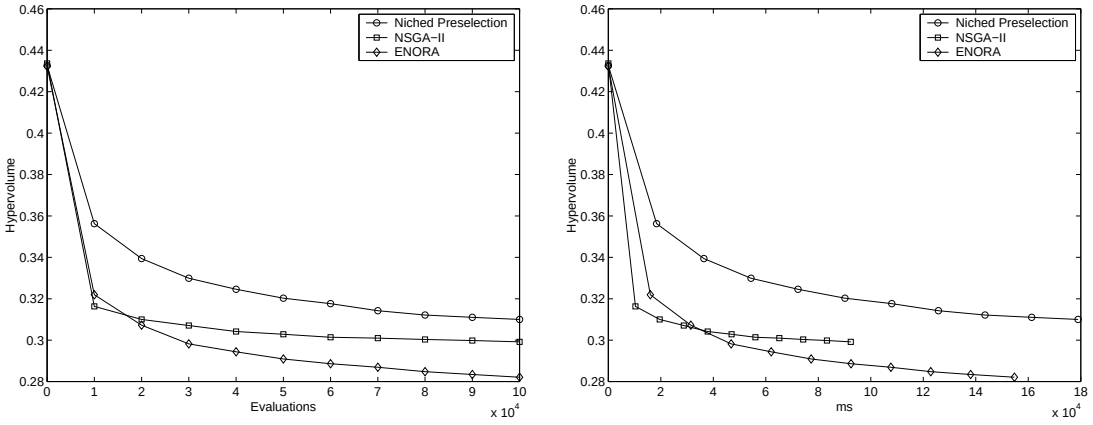


Figure 3: Evolution of medium hypervolume of 100 runs of the algorithms for classifying Medical data set.

5.4 Comparison with other non-evolutionary machine learning techniques: multi-objective cross-validation

In this section, the aim is to validate the classification systems obtained with the proposed algorithms and to compare our approach with other non-evolutionary machine learning techniques. In this case, cross-validation is required.

5.4.1 Multi-objective cross-validation

Cross-validation [99] is a technique for assessing how the results of a given statistical analysis can be generalized to an independent data set. It is mainly used in settings where the goal is prediction and to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of the data set into two complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other (called the validation set or testing set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

A common type of cross-validation is k-fold cross-validation, in which the original sample is randomly partitioned into k subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation only once. 10-fold cross-validation is commonly used, but, in general, k remains an unfixed parameter [100].

In multi-objective classification systems, the cross-validation process is obviously more complex because the set of solutions to be validated is greater than in simple classification systems. The solutions of the optimization model (3) is a set of non-dominated solutions, each with a different number of rules. In this section, we propose an adaptation of the k-fold cross-validation ($k = 10$) for multi-objective classification problems as in (3). The algorithm 20 shows the proposed technique. This algorithm requires as input K partitions of the data set D (we use $K = 10$). For each partition P_k, a MOEA is run $NoRuns$ (we use $NoRuns = 100$) considering $D - P_k$ as input data set (training data). The MOEA output, for training data $D - P_k$ in the run l, is a set of non-dominated solutions S_k^l. For each set of solutions S_k^l, classification rate, sensitivity and specificity values are calculated using test data P_k by means of the functions $Evaluate\_CR$, $Evaluate\_Sensitivity$ and $Evaluate\_Specificity$, respectively. Classification rate is evaluated as introduced in section 3.3, while specificity and sensitivity are calculated as defined in section 5. Finally, for each rule number i, for $i = M_{min}$ to M_{max}, the average classification rate of the obtained solutions (classifiers) with a number of rules equal to i is showed

as output. The average specificity and sensitivity are similarly calculated and also included as output.

Tables 10 and 11 show the results of the multi-objective k-fold cross-validation algorithm for iris and medical data set respectively. Additionally, minimum and maximum values of metrics for each number of rules are shown. As summary of the multi-objective cross-validation process, we show, for each algorithm, the medium values over the models obtained for each metric, including the average number of rules.

Algorithm 20 Multi-objective k-fold cross-validation algorithm

Require: Data Set D; $P_1, P_2, \ldots, P_K$ {K partitions of D}

Require: $NoRuns$ {Number of runs}

Require: M_{min},M_{max}, $(0 < M_{min} \leq M_{max})$ {Minimum and maximum number of rules}

for $k = 1$ to K **do**

  for $l = 1$ to $NoRuns$ **do**

    $S_k^l = \{S_k^{li}, i = M_{min}, \ldots, M_{max}\} \leftarrow MOEA\,(D - P_k)$

    $CR_k^l = \{CR_k^{li}, i = M_{min}, \ldots, M_{max}\} \leftarrow Evaluate\_CR\,(S_k^l, P_k)$

    $Se_k^l = \{Se_k^{li}, i = M_{min}, \ldots, M_{max}\} \leftarrow Evaluate\_Sensitivity\,(S_k^l, P_k)$

    $Sp_k^l = \{Sp_k^{li}, i = M_{min}, \ldots, M_{max}\} \leftarrow Evaluate\_Specificity\,(S_k^l, P_k)$

  end for

end for

$$CR_{avg}^i \leftarrow \frac{1}{K \cdot NoRuns} \sum_{k=1}^{K} \sum_{l=1}^{NoRuns} CR_k^{li},\ i = M_{min}, \ldots, M_{max}$$

$$Se_{avg}^i \leftarrow \frac{1}{K \cdot NoRuns} \sum_{k=1}^{K} \sum_{l=1}^{NoRuns} Sensitivity_k^{li},\ i = M_{min}, \ldots, M_{max}$$

$$Sp_{avg}^i \leftarrow \frac{1}{K \cdot NoRuns} \sum_{k=1}^{K} \sum_{l=1}^{NoRuns} Specificy_k^{li},\ i = M_{min}, \ldots, M_{max}$$

return $CR_{avg}^i, Se_{avg}^i, Sp_{avg}^i, i = M_{min}, \ldots, M_{max}$

5.4.2 Comparison with other non-evolutionary machine learning techniques

Previous studies on mortality scoring and ICBU survival estimation using AI approaches have considered a wealth of methods, with special attention paid to artificial neuronal networks [37] and probabilistic and analogy-based techniques [9, 39, 101]. Comparing the effectiveness of different AI techniques to solve general medical problem (such as mortality prediction) is a complex and still unresolved issue for the scientific community. However, in order to clearly show the advantages and disadvantages of the fuzzy classification methodology and algorithms proposed, we consider two main dimensions: the results statistics (e.g. classification rate, specificity and sensitivity) and the complexity of the model and its comprehensibility by physicians. The following approaches have been studied:

M	CR_{min}	CR_{max}	CR_{avg}
Niched pre-selection			
3	0.9333	1.0000	0.9916
4	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000
Mean 4.0000	0.9778	1.0000	0.9972
NSGA-II			
3	0.8667	1.0000	0.9623
4	1.0000	1.0000	1.0000
Mean 3.5000	0.9333	1.0000	0.9811
ENORA			
3	0.8667	1.0000	0.9648
4	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000
Mean 5.5000	0.9778	1.0000	0.9941

Table 10: Minimum, maximum and average classification rate $(CR_{min}, CR_{max}, CR_{avg})$, obtained with 100 runs of algorithms and 10-fold for Iris data set.

M	CR_{min}	CR_{max}	CR_{avg}	Se_{min}	Se_{max}	Se_{avg}	Sp_{min}	Sp_{max}	Sp_{avg}
Niched pre-selection									
10	0.5000	1.0000	0.8143	0.0000	1.0000	0.7334	0.5000	1.0000	0.8806
11	0.8000	1.0000	0.9400	0.7143	1.0000	0.9429	0.8571	1.0000	0.9714
12	0.9000	1.0000	0.9333	0.6667	1.0000	0.8222	1.0000	1.0000	1.0000
13	0.9000	0.9000	0.9000	1.0000	1.0000	1.0000	0.8571	0.8571	0.8571
14	0.7000	1.0000	0.8667	0.7143	1.0000	0.9048	0.6667	1.0000	0.8413
15	0.9000	0.9000	0.9000	0.6667	0.8571	0.7619	1.0000	1.0000	1.0000
18	0.8000	0.8000	0.8000	0.7143	0.7143	0.7143	1.0000	1.0000	1.0000
Mean 13.2857	0.7857	0.9429	0.8792	0.6395	0.9388	0.8399	0.8401	0.9796	0.9358
NSGA-II									
10	0.5000	1.0000	0.7245	0.0000	1.0000	0.6671	0.4286	1.0000	0.7882
11	0.8000	0.8000	0.8000	0.5000	1.0000	0.7500	0.7143	1.0000	0.8571
Mean 10.5000	0.6500	0.9000	0.7622	0.2500	1.0000	0.7085	0.5714	1.0000	0.8227
ENORA									
10	0.7000	1.0000	0.8693	0.2500	1.0000	0.8170	0.3333	1.0000	0.8969
13	0.9000	0.9000	0.9000	1.0000	1.0000	1.0000	0.8571	0.8571	0.8571
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0.9000	1.0000	0.9500	0.8571	1.0000	0.9286	1.0000	1.0000	1.0000
Mean 14.2500	0.8750	0.9750	0.9298	0.7768	1.0000	0.9364	0.7976	0.9643	0.9385

Table 11: Minimum, maximum and average classification rate ($CR_{min}, CR_{max}, CR_{avg}$), sensitivity ($Se_{min}, Se_{max}, Se_{avg}$) and specificity ($Sp_{min}, Sp_{max}, Sp_{avg}$) obtained with 100 runs of algorithms and 10-fold for Medical data set.

- Artificial neuronal networks: commonly used in the literature. In this work we have considered the classical multilayer perceptron considering 10 hidden layers [102].

- Naive Bayes: a probabilistic model which assumes that attributes independently contribute to the probability of survival (e.g. $p\left(mortality|Total, Prof, \ldots\right)$) [103].

- Decision trees: well-known decision support techniques and easily interpretable by physicians. In particular, the following classical algorithm considered: J.48 [104], altering decision tree (AD tree) [105] and best first decision tree (BF tree) [106, 107].

- Analogy-reasoning approaches: in particular, the case-based reasoning methodology (CBR), which we considered in previous works using this clinical data set [101]. This CBR process considers standard similarity functions for the retrieval step and a rule-based approach for the adaptation step.

Table 12 summarizes the results of the above mentioned techniques using the ICBU data set.

$\boldsymbol{Method}$	$\boldsymbol{CR_{avg}}$	$\boldsymbol{Sp_{avg}}$	$\boldsymbol{Se_{avg}}$	$\boldsymbol{Complexity}$	$\boldsymbol{Time\,(s.)}$
J.48 [104]	0.6969	0.7586	0.6097	3 nodes	0.3
AD [105]	0.7273	0.8103	0.6097	31 nodes	0.3
BF [107]	0.7373	0.7758	0.6829	8 nodes	0.8
NBayes [103]	0.8282	0.8621	0.7804	18 variables	0.01
CBR [101]	0.7530	**0.9720**	0.6940	90 cases + 14 rules	7.28
ANN [102]	0.8282	0.8620	0.7804	10 layers	6.6
Niched pre-selection	0.8792	0.9358	0.8399	13.3 rules (avg.)	$1786.07(avg.)$
ENORA	**0.9298**	0.9385	**0.9364**	14.2 rules (avg.)	$1528.61(avg.)$
NSGA-II	0.7622	0.8227	0.7085	10.5 rules (avg.)	$984.01(avg.)$

Table 12: Summary of results obtained with non-evolutionary AI approaches and MOEA using 10-fold cross-validation and 100 runs for the evolutionary algorithms.

The results obtained highlight the main advantages of the evolutionary proposal: a balance between the need for accuracy and a model that is easily understood. In general, the greater complexity of the non-evolutionary models implies better statistical results. For example, the ANN approach achieves an average 82% of classification, but but involves 10-layer neuronal network, a model that cannot be easily interpreted by physicians. On the one hand, ENORA and niched pre-selection outperform the non-evolutionary approaches, for all statistical results

(Classification Rate, Specificity and Sensitivity), but, on average, the number of rules considered guarantees simple validation by intensivists. In this sense, some other approaches considered (e.g. decision tree algorithms) provide simple models. For example, J.48 gives a 3-node tree or CBR uses only 14 rules to adapt the solutions. It is worth mentioning the high specificity of the CBR approach, however the system is only specialized to identify *exitus* since this specificity is to the detriment of the accuracy and sensitivity. However, their classification rates are considerably lower than those obtained with niched pre-selection or ENORA. Regarding the analysis of computational time of the evolutionary and non-evolutionary algorithms, it is worth mentioning that they belong to different orders of magnitude. While non-evolutionary techniques require a maximum of 6.6 seconds to perform cross-validation with $K = 10$ in a HP Proliant DL145 G2 AMD Opteron Processor 246 2GHz/1MB, evolutionary approaches requires, in the same machine, at least 0.27 hours to perform cross-validation with $K = 10$ and $NoRuns = 100$. Nevertheless, this is not a disadvantage in practice, since this kind of process is performed offline. Moreover, note that non-evolutionary approaches require only one execution in the cross-validation process, while evolutionary algorithms require a greater number of executions, usually $NoRuns = 100$, due to they are probabilistic algorithms.

5.5 Analysis of the solution

The aim of this section is to qualitatively analyze one solution of the problem obtained with the proposed algorithms. The solution is chosen by the decision-making process described in section (4.9) from all the sets of non-dominated solutions generated by the multi-objective cross-over validation process in the three algorithms. The solutions chosen for the Iris and Medical Data Sets are shown in tables 14 and 15 respectively.

The chosen solutions for the Iris and Medical Data Sets were obtained by the ENORA algorithm. As regards the number of rules and the different fuzzy sets for each real variable, it may be said that the models are compact. The similarity of the fuzzy sets satisfies the imposed constraint to the optimization model (3). The linguistic label algorithm (18) found an acceptable number of linguistic labels for both the Iris and Medical data sets. Therefore, the chosen models fulfil the interpretability criteria (compactness and similarity) imposed by the decision-maker. As for accuracy, the classification rate for Iris and Medical data set is maximum. Additionally for the Medical data set, the sensitivity and specific values are also maximum. Therefore, the chosen solutions for both the Iris and Medical data sets are accurate and easily interpretable.

Data set	**Iris**				**Medical**	
Algorithm	Niched pre-selection				ENORA	
Number of rules	4				10	
Number of fuzzy sets	9				8	
Variables	x_1	x_2	x_3	x_4	x_1	x_2
Number of fuzzy sets for each variable	1	2	3	3	4	4
Number of linguistic labels for each variable	–	2	6	3	6	5
Similarity	0.0185				0.0921	
Classification rate	1.0000				0.9293	
Specificity	–				0.9310	
Sensitivity	–				0.9268	

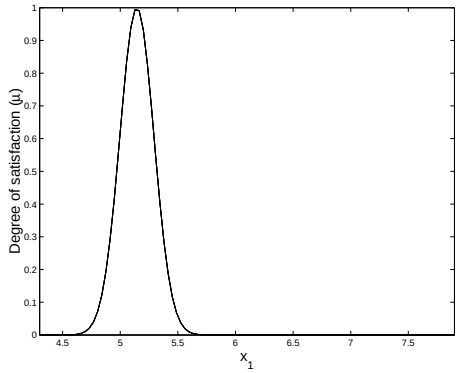
Table 13: Accuracy and interpretability values for the chosen solutions obtained with niched pre-selection and ENORA.

Table 13 shows the accuracy and interpretability values.

6 Discussion, conclusions and future trends

This article presents a novel methodology for the fuzzy classification of data sets with real and categorized input variables and multiple categorized outputs, by means of multi-objective evolutionary fuzzy systems, using a theme of great relevance in clinical practice – the classification of mortality through infection in severe burn patients. We propose a multi-objective constrained optimization model for fuzzy classification, taking into account the criteria accuracy, transparency and compactness. Two Pareto-based elitist multi-objective evolutionary algorithms are proposed (niched pre-selection and ENORA) and compared with the well known NSGA-II using statistical tests over the hypervolume values. These algorithms use variable-length representation and specialized variation operators over this representation. We also propose a multi-objective cross-validation technique to validate the fuzzy classifiers obtained by these algorithms. Besides, a linguistic labeling algorithm is proposed to associate linguistic labels with fuzzy sets. A decision-making process is also described to choose an a posteriori solution from the Pareto set. The results obtained are better than those obtained with other techniques commonly used in medicine, with the added advantage that the proposed technique identifies alternative solutions.

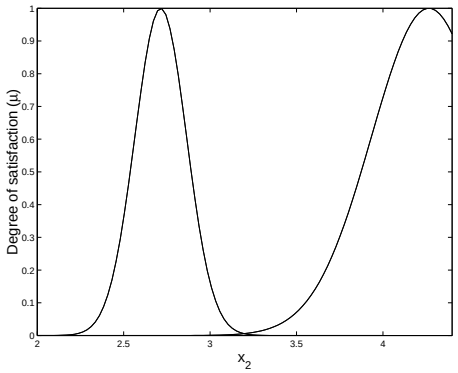
x_1	x_2	x_3	x_4	y
DCC	L	VL	L	1
DCC	H	L	M	2
DCC	H	VH	L	3
DCC	L	L	H	3



$\mathcal{S} = 0.000000$

1 linguistic label

Do not Care Condition DCC $(5.14; 0.15)$

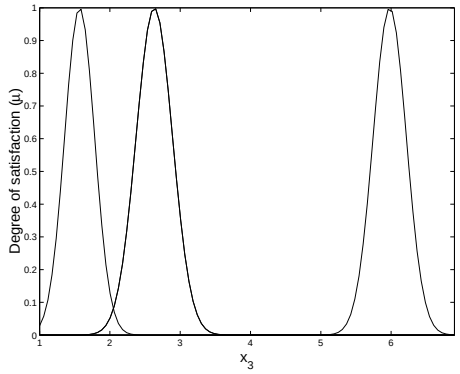


$\mathcal{S} = 0.003645$

2 linguistic labels

Low L $(2.71; 0.15)$

High H $(4.27; 0.33)$



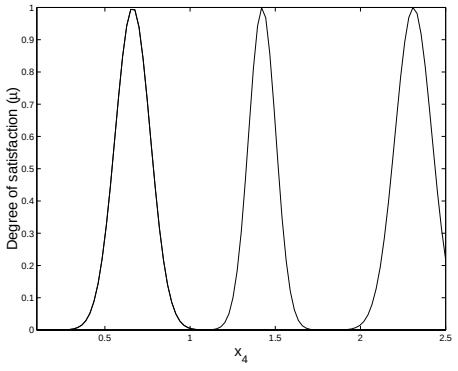
$\mathcal{S} = 0.018532$

6 linguistic labels

Very Low VL $(1.57; 0.21)$

Low L $(2.63; 0.26)$

Very High VH $(5.98; 0.24)$



$\mathcal{S} = 0.000011$

3 linguistic labels

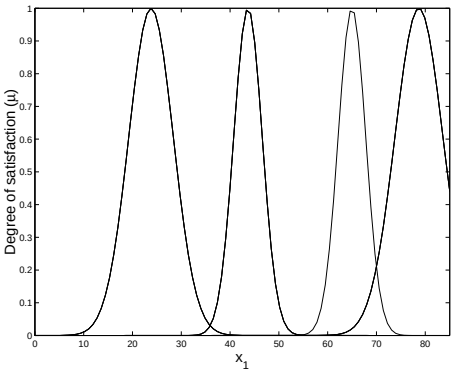
Low L $(0.66; 0.10)$

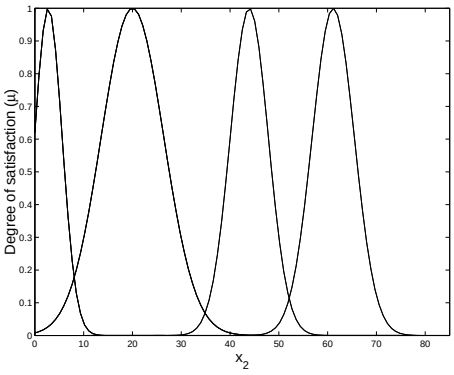
Medium M $(1.42; 0.08)$

High H $(2.31; 0.11)$

Table 14: Fuzzy model with 4 rules and 9 fuzzy sets for classifying Iris data set obtained with niched pre-selection. Classification rate=1.000000, similarity=0.018532.

x_1	x_2	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	y
H	M	2	1	2	2	2	2	2	2	2	2	1	1
L	ML	1	2	1	2	2	1	1	1	1	1	2	1
MH	L	1	2	2	2	1	1	2	1	1	2	1	1
MH	L	1	1	1	1	1	2	1	1	2	2	1	2
L	ML	2	2	2	2	1	2	2	1	2	1	2	2
VH	MH	1	2	2	2	1	2	1	1	2	1	1	2
VH	ML	1	1	2	1	2	2	2	1	2	1	2	2
MH	MH	2	2	2	1	1	1	1	1	1	1	2	2
VH	L	2	2	1	1	2	1	2	2	2	2	1	2
L	M	2	2	1	2	2	1	1	2	2	1	1	2





$\mathcal{S} = 0.092131$

6 linguistic labels

Low	L	$(23.81; 4.57)$
Moderately High	MH	$(43.67; 2.93)$
High	H	$(64.97; 2.83)$
Very High	VH	$(78.70; 4.92)$

$\mathcal{S} = 0.077850$

5 linguistic labels

Low	L	$(2.77; 2.83)$
Moderately Low	ML	$(20.00; 6.42)$
Medium	M	$(43.94; 3.91)$
Moderately High	MH	$(61.14; 4.32)$

Table 15: Fuzzy model with 10 rules and 8 fuzzy sets for classifying Medical data set obtained with ENORA. Classification rate=0.929293, specificity = 0.931034, sensitivity = 0.926829, similarity=0.092131.

The following clinical conclusions about the fuzzy rule set presented in table 15, can be considered:

- Physicians find these rules easily to interpret since the antecedent is a simple conjunction of variables and the consequent is a single dichotomic variable.

- The knowledge base consists of 10 fuzzy rules: 3 rules suggest the survival of the patient and 7 rules suggest a potential $exitus$. The number of rules is reasonable from the clinical practice point of view.

- Physicians find these rules useful in order to support the treatment of patients according to the survival expectancy, providing an aggressive or more conservative therapy. For example, the $5th$, $6th$ and $7th$ rules support decisions regarding patients suffering from heart complications, a high blood pressure and bacteremia at the same time. The $2nd$ and $3rd$ rules show that no deep burnt patients with no heart/liver complications and normal blood pressure belong to a survival profile.

- Specificity values obtained by our proposal are especially remarkable for physicians. In this test, a positive outcome means that the system estimates and exitus and aggressive measures must be taken.

- Unlike traditional severity scores (e.g. Baux, PBI, ABSI), age and weight variables are neglected in our rule set (see table 6). From the statistical point of view, these variables are not correlated to the survival of the patient in the medical database. Despite physicians expected a higher analogy between our proposal and the traditional severity scores, our rules deal with infections and co-morbidity information. In fact, in the medical literature there are some authors that suggest that simple measurements (weight, height, sex) outweigh other significant variables [7, 16].

The following advantages and benefits compared with existing solutions can be mentioned:

- Constrained multi-objective optimization allows the simultaneous search for accurate and interpretable fuzzy classifiers in a global way. Other works in the literature perform the search for accurate and interpretable classifiers in separate processes with a consequent loss of information between both processes.

- Our proposal allows the classification of data sets containing both real and categorized input variables, identifying a reasoning method and a rule weight assignment which combines both types of input variable. Most works in the literature only allow classification of data sets with real input variables. Besides, our proposal allows the classification of multiple categorized outputs, whereas other works only consider binary classifications.

- Our proposal uses Pareto-based elitist MOEAs which generate a set of non-dominated solutions from which a decision maker may choose, in an a posteriori process, the most satisfying solution according to current preferences. In this way, a change in decision-maker preferences does not require new runs of the algorithm as the decision-maker simply perform a new choice among the set of solutions. Other works in the literature propose non-Pareto MOEAs which only return a single solution according to a priori decision-maker preferences and while runs of the algorithm are required if decision-makers change their preferences.

- In our proposal, the degree of interpretability is parameterized. The parameter g_s constraints the overlapping degree of the fuzzy sets while the minimum and maximum number of rules of the classifiers is bounded by the parameters M_{min} and M_{max}. Other approaches establish a priori fixed levels of overlapping between the fuzzy sets. Besides, many works just consider a fixed number of rules for the classifiers.

- Our proposal uses real parameter optimization for the fuzzy set search while other works which use combinatorial optimization. A deterministic linguistic labeling algorithm is used when the real parameter optimization is finished in order to assign one linguistic label to each fuzzy set. The advantage of our proposal is the learning of high accurate fuzzy classifiers with an acceptable degree of interpretability, and a smaller number of evaluations is required to reach convergence in the presence of high dimensional data sets.

- We use a Pareto-based elitist MOEA with variable-length representation which allows the search for classifiers with a different number of rules. Many of the works in the literature just search for classifiers with a fixed number of rules.

- We use fuzzy sets with Gaussian membership functions for our classifiers, whereas other works use fuzzy sets with triangular or trapezoidal membership functions, which have

been proved to be less flexible than gaussian in fuzzy modeling.

- Our proposal identifies the granularity of the real input variables and is therefore perse a feature-selection technique, given that it allows "don't care condition" variables to be identified.

- In our proposal, the classifiers are validated with a multi-objective cross-validation methodology. Other studies use classical cross-validation.

- Compared to other MOEAs, niched pre-selection and ENORA algorithms, as proposed in this paper, maintain diversity as regards the number of rules. Other MOEAs, such as NSGA-II, are less effective in this aspect.

As future trends, we suggest the following:

- Asymmetric Gaussian membership functions. In our proposal we use, without losing generality, symmetric gaussian membership functions. By using asymmetric gaussian membership functions, which are more flexible than symmetric ones, classifiers may provide better classification rates without harming interpretability.

- Three objective optimization constrained model. An optimization model that maximizes sensitivity and specificity and minimizes the number of rules could be considered.

- Application of the methodology to other medical and non-medical data sets from the University of California Irvine (UCI) Machine Learning Repository.

Acknowledgement

We would like to thank Dr. J.A. Lorente and Dr. F. Palacios at the ICBU of the Getafe University Hospital (Madrid, Spain) for their fine contribution to this work.

This work was partially funded by the Spanish Ministry of Science and Innovation+European FEDER+PlanE funds under the project TIN2009-14372-C03-01 and by the Seneca Research Foundation of the Region of Murcia under project 15277/PI/10.

References

[1] K. Rafla and E.E. Tredget. Infection control in the burn unit. *Burns*, 37(1):5 – 15, 2011.

[2] American Burn Association. Burn incidence and treatment in the united states: 2011 fact sheet, 2011.

[3] J.P. Barret, P. Gómez, I. Solano, M. González-Dorrego, and Crisol F.J. Epidemiology and mortality of adult burns in catalonia. *Burns*, 25(4):325 – 329, 1999.

[4] R. Galeiras, J.A. Lorente, S. Pértega, A. Vallejo, V. Tomicic, M.A. De La Cal, S. Pita, E. Cerda, and A. Esteban. A model for predicting mortality among critically ill burn victims. *Burns*, 35(2):201 – 209, 2009.

[5] X. Jie and C. Baoren. Mortality rates among 5321 patients with burns admitted to a burn unit in china: 1980-1998. *Burns*, 29(3):239 – 245, 2003.

[6] M.F. Khadim, A. Rashid, B. Fogarty, and K. Khan. Mortality estimates in the elderly burn patients: The northern ireland experience. *Burns*, 35(1):107 – 113, 2009.

[7] N.M. Sheppard, S. Hemington-Gorse, O.P. Shelley, B. Philp, and P. Dziewulski. Prognostic scoring systems in burns: A review. *Burns*, 37(8):1288 – 1295, 2011.

[8] J.R. Legall, S. Lemeshow, and F. Saulnier. A new simplified acute physiology score (saps-ii) based on a european north-american multicenter study. *Jama-Journal of the American Medical Association*, 270(24):2957–2963, 1993.

[9] T. Toma, A. Abu-Hanna, and R.J. Bosman. Discovery and inclusion of sofa score episodes in mortality prediction. *Journal of Biomedical Informatics*, 40(6):649–660, 2007.

[10] V. McGill, A. Kowal-Vern, and R.L. Gamelli. Outcome for older burn patients. *Archives of Surgery*, 135(3):320–325, MAR 2000.

[11] S. Baux. *Contribution à l'étude du traitement local des brûlures thermiques étendues*. PhD thesis, Paris, 1961.

[12] H. Nakae and H. Wada. Characteristics of burn patients transported by ambulance to treatment facilities in akita prefecture, japan. *Burns*, 28(1):73 – 79, 2002.

[13] C.J. Clark, W.H. Reid, W.H. Gilmour, and D. Campbell. Mortality probability in victims of fire trauma - revised equation to include inhalation injury. *British Medical Journal*, 292(6531):1303–1305, MAY 17 1986.

[14] Y. Godwin and S.H. Wood. Major burns in cape town: a modified burns score for patient triage. *Burns*, 24(1):58 – 63, 1998.

[15] T. Osler, L.G. Glance, and D.W. Hosmer. Simplified estimates of the probability of death after burn injuries: Extending and updating the baux score. *Journal of trauma-injury infection and critical care*, 68(3):690–697, 2010.

[16] W.S. Ho, S.Y. Ying, and A. Burd. Outcome analysis of 286 severely burned patients: retrospective study. *Hong Kong Med Journal*, 8(4):235–239, 2002.

[17] J. Tobiasen, J.M. Hiebert, and R.F. Edlich. The abbreviated burn severity index. *Annals of Emergency Medicine*, 11(5):260–262, 1982.

[18] B.S. Atiyeh, S.W. Gunn, and S.N. Hayek. State of the art in burn treatment. *World Journal of Surgery*, 29(2):131–148, 2005.

[19] J. Fitzwater, G.F. Purdue, J.L. Hunt, and O'Keefe G.E. The risk factors and time course of sepsis and organ dysfunction after burn trauma. *Journal of Trauma-Injury Infection and Critical Care*, 54(5):959–966, 2003.

[20] F. Herrera. Genetic Fuzzy Systems: Status, Critical Considerations and Future Directions. *International Journal of Computational Intelligence Research*, 1, 2005.

[21] F. Herrera. Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1:27–46, 2008.

[22] R. Babuska. *Fuzzy modeling and identification*. PhD thesis, Technische Universiteit Delft, 1996.

[23] R. Babuska and H. B. Verbruggen. Fuzzy Set Methods for Local Modelling and Identification. In Murray R. Smith and T. A. Johansen, editors, *Multiple Model Approaches to Modeling and Control*, pages 75–100. Taylor and Francis, London, UK, 1997.

[24] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena. *Interpretability issues in fuzzy modeling*. Springer, Berlin, Gernay, 2003.

[25] C. Mencar. Interpretability of fuzzy information granules. In Andrzej Bargiela and Witold Pedrycz, editors, *Human-Centric Information Processing Through Granular*

Modelling, volume 182 of *Studies in Computational Intelligence*, pages 95–118. Springer Berlin Heidelberg, Berlin, Germany, 2009.

[26] D. Nauck, J. Gebhardt, and R. Kruse. Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine*, 16:149–169, 1999.

[27] C.A. Coello. Evolutionary multi-objective optimization: a historical view of the field. *IEEE Computacional Intelligence Magazine*, 1(1):28 – 36, 2006.

[28] C.A. Coello, Veldhuizen D.V., and Lamont G.B. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic/Plenum publishers, New York, NY, USA, 2002.

[29] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, London, UK, 2001.

[30] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 8(2):173 – 195, 2000.

[31] D.C. Becalick and T.J. Coats. Comparison of artificial intelligence techniques with UK-TRISS for estimating probability of survival after trauma. *Journal of trauma-injury infection and critical care*, 51(1):123–133, JUL 2001.

[32] T.G. Buchman, K.L. Kubos, A.J. Seidler, and M.J. Siegforth. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive-care unit. *Critical Care Medicine*, 22(5):750–762, MAY 1994.

[33] V. Gautier, H. Redier, J.L. Pujol, J. Bousquet, H. Proudhon, C. Michel, J.P. Daures, F.B. Michel, and P. Godard. Comparison of an expert system with other clinical scores for the evaluation of severity of asthma. *European Respiratory Journal*, 9(1):58–64, JAN 1996.

[34] M.D. McGonigal, J. Cole, C.W. Schwab, D.R. Kauder, M.F. Rotondo, and P.B. Angood. A new approach to probability of survival scoring for trauma quality assurance. *Journal of trauma-injury infection and critical care*, 34(6):863–870, JUN 1993.

[35] L.J. Scheetz, J. Zhang, and J. Kolassa. Classification tree modeling to identify severe and moderate vehicular injuries in young and middle-aged adults. *Artificial Intelligence in Medicine*, 45(1):1–10, JAN 2009.

[36] P.J.G. Lisboa. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks*, 15(1):11–39, JAN 2002.

[37] A. Nimgaonkar, D.R. Karnad, S. Sudarshan, L. Ohno-Machado, and I. Kohane. Prediction of mortality in an Indian intensive care unit - Comparison between APACHE II and artificial neural networks. *Intensive Care Medicine*, 30(2):248–253, FEB 2004.

[38] M. Frize and R. Walker. Clinical decision-support systems for intensive care units using case-based reasoning. *Medical Engineering & Physics*, 22(9):671–677, NOV 2000.

[39] L. Minne, A. Abu-Hanna, and De Jonge. E. Evaluation of sofa-based models for predicting mortality in the icu: A systematic review. *Critical Care*, 12(6), 2008.

[40] J.M. Juárez, M. Campos, J. Palma, and R. Marín. T-care: temporal case retrieval system. *Expert Systems*, 28(4, SI):324–338, 2011.

[41] L. Vanneschi, A. Farinaccio, M. Giacobini, G. Mauri, M. Antoniotti, and P. Provero. Identification of Individualized Feature Combinations for Survival Prediction in Breast Cancer: A Comparison of Machine Learning Techniques. In Pizzuti, C. and Ritchie, M.D. and Giacobini, M., editor, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Proceedings*, volume 6023 of *Lecture Notes in Computer Science*, pages 110–121, 2010.

[42] C.L. Chan and H.W. Ting. Constructing a novel mortality prediction model with bayes theorem and genetic algorithm. *Expert Systems with Applications*, 38(7):7924 – 7928, 2011.

[43] E.H. Shortliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier/North Holland, New York, NY, USA, 1976.

[44] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338 – 353, 1965.

[45] F. Steimann. On the use and usefulness of fuzzy sets in medical ai. *Artificial Intelligence in Medicine*, 21:131 – 7, 2001.

[46] Z. Cui. A novel medical image dynamic fuzzy classification model based on ridgelet transform. *Journal of Software*, 5(5):458 – 456, 2010.

[47] N.S. Iyer, A. Kandel, and M. Schneider. Feature-based fuzzy classification for interpretation of mammograms. *Fuzzy Sets and Systems*, 114:271 – 280, 2000.

[48] S. Alayón, R. Robertson, S.K. Warfield, and J. Ruiz-Alzola. A fuzzy system for helping medical diagnosis of malformations of cortical development. *Journal of Biomedical Informatics*, 40:221 – 235, 2007.

[49] I. Gadaras and L. Mikhailov. An interpretable fuzzy rule-based classification methodology for medical diagnosis. *Artificial Intelligence in Medicine*, 47:25 – 41, 2009.

[50] M. Setnes, R. Babuska, and H.B. Verbruggen. Rule-based modeling: precision and transparency. *Trans. Sys. Man Cyber Part C*, 28(1):165–169, feb 1998.

[51] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.

[52] C.A. Peña Reyes and M. Sipper. Evolutionary computation in medicine: an overview. *Artificial Intelligence in Medicine*, 19(1):1–23, 2000.

[53] O. Cordón, F. Herrera, and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems*, 9:2001, 2000.

[54] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules from classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(3):260–270, 1995.

[55] Gómez-Skarmeta A.F. and F. Jiménez. Fuzzy modeling with hybrid systems. *Fuzzy Sets and Systems*, 104:199–208, 1999.

[56] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions Fuzzy Systems*, 3(3):269 – 279, 2005.

[57] J.M. Alonso and L. Magdalena. Special issue on interpretable fuzzy systems. *Information Sciences*, 181(20):4331 – 4339, 2011.

[58] J. Casillas, O. Cordón, M.J. Del Jesús, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Transactions Fuzzy Systems*, 13(1):13 – 29, 2005.

[59] J. Casillas, F. Herrera, R. Pérez, M.J. Del Jesús, and P. Villar. Special issue on genetic fuzzy systems and the interpretabilityaccuracy tradeoff. *International Journal of Approximate Reasoning*, 44(1):1 – 3, 2007.

[60] J. Yen, L. Wang, and C.W. Gillespie. Improving the interpretability of TSK fuzzy models by combining global learning and global learning. *IEEE Transactions on Fuzzy Systems*, 6(4):530–537, Nov 1998.

[61] C.A. Coello and Lamont G.B. *Applications of multiobjective evolutionary algorithms*, volume 1 of *Advances in Natural Computation*. World Scientific, Singapore, 2004.

[62] A.F. Gómez-Skarmeta, F. Jiménez, and J. Ibáñez. Pareto-optimality in fuzzy modeling. In *In Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing*, pages 694–700, 1998.

[63] H. Ishibuchi, T. Murata, and I.B. Turksen. Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets and Systems*, 89(2):135 – 150, 1997.

[64] T. Suzuki, T. Furuhashi, S. Matsushita, and H. Tsutsui. Efficient fuzzy modeling under multiple criteria by using genetic algorithm. In *Proccedings of the IEEE International Conference on Systems, Man, and Cybernetics On 1999. IEEE SMC '99 Conference Proceedings*, volume 5 of *SMC '99*, pages 314–319, Washington, DC, USA, 1999. IEEE Computer Society.

[65] H. Ishibuchi, T. Nakashima, and T. Murata. Three-objective genetics-based machine learning for linguistic rule extraction. *Information Sciences*, 136(1-4):109 – 133, 2001. Recent Advances in Genetic Fuzzy Systems.

[66] Y. Jin, editor. *Multi-Objective Machine Learning*, volume 16 of *Studies in Computational Intelligence*. Springer, Warsaw, Poland, 2006.

[67] F. Jiménez, A.F. Gómez-Skarmeta, H. Roubos, and R. Babuska. Accurate, transparent, and compact fuzzy models for function approximation and dynamic modeling through multi-objective evolutionary optimization. In Eckart Zitzler, Lothar Thiele, Kalyanmoy Deb, CarlosArtemio Coello Coello, and David Corne, editors, *Evolutionary Multi-Criterion Optimization*, volume 1993 of *Lecture Notes in Computer Science*, pages 653–667. Springer Berlin Heidelberg, Berlin, Gernany, 2001.

[68] F. Jiménez, A.F. Gómez-Skarmeta, H. Roubos, and R. Babuska. A multi-objective evolutionary algorithm for fuzzy modeling. In *Proceedings of the IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, volume 2, pages 1222–1228, 2001.

[69] O. Cordón, F. Herrera, M.J. del Jesús, and P. Villar. A multiobjective genetic algorithm for feature selection and granularity learning in fyzzy-rule based classifiction systems. In *Proceedings of the IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, volume 3, pages 1253–1258, 20021.

[70] H. Ishibuchi and T. Yamamoto. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Systems*, 141(1):59 – 88, 2004. ¡ce:title¿Genetic Fuzzy Systems: New Developments¡/ce:title¿.

[71] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4 – 31, 2007. ¡ce:title¿Genetic Fuzzy Systems and the Interpretability Accuracy Trade-off¡/ce:title¿.

[72] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni. A pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems. *Soft Computing*, 11, 2007.

[73] A.F. Gómez-Skarmeta, F. Jiménez, and G. Sánchez. Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms. *International Journal of Intelligent Systems*, 22(9):943–969, 2007.

[74] I. Krivy. Optimization of fuzzy models using evolutionary algorithms. Technical Report 115, Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, 2007.

[75] M.J. Gacto, R. Alcalá, and F. Herrera. Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. *Soft Computing*, 13(5):419–436, December 2009.

[76] P. Ducange, B. Lazzerini, and F. Marcelloni. Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets. *Soft Computing*, 14(7):713–728, May 2010.

[77] R. Alcaraz, M.J. Gacto, and F. Herrera. A fast and scalable multiobjective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems. *IEEE T. Fuzzy Systems*, 19(4):666–681, 2011.

[78] M. Cruz-Ramírez, C. Hervás-Martínez, J.C. Fernández-Caballero, J. Briceño, and M. de la Mata. Multi-objective evolutionary algorithm for donor-recipient decision system in liver transplants. *European Journal of Operational Research*, 222:317–327, 2012.

[79] Y. Jin and B. Sendhoff. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38:397 – 415, 2008.

[80] H. Ishibuchi. Multiobjective genetic fuzzy systems: Review and future research directions. In *FUZZ-IEEE'07*, pages 1–6, 2007.

[81] V. Kreinovich, C. Quintana, and L. Reznik. Gaussian membership functions are most adequate in representing uncertainty in measurements. In *Proceedings of the NAFIPS'92*, pages 618–624, Puerto Vallarta, 1992. North American Fuzzy Information Processing Society Conference.

[82] M. Setnes, R. Babuska, U. Kayamak, and van Nauta. Similarity Measures in Fuzzy Rule Base Simplification. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 28:376–386, 1998.

[83] F. Jiménez, A.F. Gómez-Skarmeta, G. Sánchez, and K. Deb. An evolutionary algorithm for constrained multi-objective optimization. In *Proceedings of the Evolutionary Com-*

putation on 2002. CEC '02. Proceedings of the 2002 Congress, volume 2 of *CEC '02*, pages 1133–1138, Washington, DC, USA, 2002. IEEE Computer Society.

[84] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs*. Springer-Verlag, London, UK, 3rd edition, 1996.

[85] M. Srinivas and L.M. Patnaik. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4):656–667, 1994.

[86] D.J. Cavicchio. *Adaptive Search Using Simulated Evolution*. Phd thesis, University of Michigan, 1970.

[87] I. Rechenberg. *Evolutionsstrategie: optimierung technischer systeme nach prinzipien der biologischen evolution*. Frommann-Holzboog, Stuttgart, Germany, 1973.

[88] H.P. Schwefel. *Numerical Optimization of Computer Models*. John Wiley & Sons, Inc., New York, NY, USA, 1981.

[89] D.G. Altman and J.M. Bland. Diagnostic tests. 1: Sensitivity and specificity. *BMJ*, 308, 1994.

[90] T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874, 1006.

[91] P.J. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61:92–105, 2005.

[92] M.H. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.

[93] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2):179–188, 1936.

[94] S.M. Chen and Y.D. Fang. A new approach for handling the iris data classification problem. *International Journal of Applied Science and Engineering*, 3(1):39–49, 2005.

[95] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[96] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[97] M. Laumanns, E. Zitzler, and L. Thiele. On the effects of archiving, elitism, and density based selection in evolutionary multi-objective optimization. In Eckart Zitzler, Lothar Thiele, Kalyanmoy Deb, CarlosArtemio Coello Coello, and David Corne, editors, *Evolutionary Multi-Criterion Optimization*, volume 1993 of *Lecture Notes in Computer Science*, pages 181–196. Springer Berlin Heidelberg, Berlin, Germany, 2001.

[98] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7:117–132, 2002.

[99] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, volume 2(12), pages 1137–1143. Morgan Kaufmann, 1995.

[100] S. Geisser. *Predictive inference: an introduction*. Chapman & Hall, New York, NY, USA, 1993.

[101] J.M. Juárez, M. Campos, A. Gomariz, J. Palma, and R. Marín. A reuse-based cbr system evaluation in critical medical scenarios. In *ICTAI*, pages 261–268, 2009.

[102] David E. Rumelhart, James L. McClelland, and CORPORATE PDP Research Group, editors. *Parallel distributed processing: explorations in the microstructure of cognition*, volume 1: foundations. MIT Press, Cambridge, MA, USA, 1986.

[103] L.I. Kuncheva. On the optimality of naive bayes with dependent binary features. *Pattern Recognition Letters*, 27(7):830 – 837, 2006.

[104] R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1993.

[105] Y. Freund and L. Mason. The alternating decision tree learning algorithm. In *Proceeding of the Sixteenth International Conference on Machine Learning*, pages 124–133, Bled, S., 1999.

[106] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression : A statistical view of boosting. *Annals of statistics*, 28(2):337–407, 2000.

[107] H. Shi. Best-first decision tree learning. Master's thesis, University of Waikato, Hamilton, N.Z., 2007. COMP594.