Soft Computing manuscript No.
(will be inserted by the editor)

Complexity Reduction and Interpretability Improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection

Marta Galende-Hernández · Gregorio I. Sainz-Palmero · M. J. Fuente-Aparicio

Received: date / Accepted: date

Abstract The aim of this paper is to develop a general post-processing methodology to reduce the complexity of data-driven linguistic fuzzy models, in order to reach simpler fuzzy models preserving enough accuracy and better fuzzy linguistic performance with respect to their initial values.

This post-processing approach is based on rule selection via the formulation of a bi-objective problem with one objective focusing on accuracy and the other on interpretability. The latter is defined via the aggregation of several interpretability measures, based on the concepts of similarity and complexity of fuzzy systems and rules. In this way, a measure of the fuzzy model interpretability is given.

Two neuro-fuzzy systems for providing initial fuzzy models, FasArt (Fuzzy Adaptive System ART based) and NefProx (Neuro-Fuzzy Function Approximation) and several case studies, data sets from KEEL Project Repository, are used to check this approach. Both fuzzy and neuro-fuzzy systems generate Mamdani-type fuzzy rule-based systems, each with its own particularities and complexities from the point of view of the fuzzy sets and the rule generation. Based on these systems and data sets, several fuzzy models are generated to check the performance of the proposal under different restrictions of complexity and fuzziness.

Keywords Fuzzy Modeling · Accuracy · Interpretability · Complexity · Genetic Algorithms

M. Galende-Hernández · G. Sainz-Palmero
CARTIF Centro Tecnológico. Parque Tecnológico de Boecillo, parcela 205, 47151 Boecillo (Valladolid), Spain
Tel.: +34-983546504
Fax: +34-983546521
E-mail: margal@cartif.es

G. Sainz-Palmero
E-mail: gresai@cartif.es

G. Sainz-Palmero · M.J. Fuente-Aparicio
Department of Systems Engineering and Control, School of Industrial Engineering, University of Valladolid. Paseo del Cauce s/n, 47011 Valladolid, Spain
E-mail: gresai@eis.uva.es

M.J. Fuente-Aparicio
E-mail: maria@autom.uva.es

1 Introduction

Fuzzy logic based systems have proved their usefulness in a large number of applications (Konar 2005; Karray and De Silva 2004; Bonissoene et al. 1999). In the scientific literature, it is possible to find methodologies, algorithms and applications based on fuzzy logic theory or in combination with other approaches: neural networks, genetic algorithms, etc. Fuzzy logic has been widely used for modeling, control, patterns recognition, computer vision, signal processing, etc. In particular fuzzy modeling approaches are very common in the engineering domain to develop black and grey box models with real-world applications, for which a large number of algorithms can be found in technical and scientific literature.

Initially, two well known modeling approaches to generate fuzzy rules are described (Herrera 2008; Casillas et al. 2003a,b; Cordón et al. 2001):

1. *Precise Fuzzy Modeling*, whose main goal is to minimize the error. In general, the models generated have a good accuracy but low level of interpretability. This modeling is popular with data-driven knowledge but expert knowledge can be considered too.
2. *Linguistic Fuzzy Modeling*, whose main goal is to have a good level of interpretability but poor accuracy. Here, knowledge from experts and from data guide the modeling process.

Both modeling approaches have drawbacks concerning accuracy or interpretability. Therefore, one interesting goal is to achieve a good balance, or compromise solution, between accuracy and interpretability, obtaining a fuzzy model with adequate accuracy and a good level of explanation (Casillas et al. 2003a,b). This accuracy-interpretability trade-off has been worked from different points of view:

- Linguistic fuzzy modeling with improved accuracy, extending the model design or changing the rule structure to make it more flexible.
- Precise fuzzy modeling with improved interpretability by reducing the complexity of the model.
- Fuzzy Modeling that pays attention to both concepts simultaneously.

The trade-off question is an open one, what is the way in which the fuzzy systems are more interpretable and accurate enough?: In (Gacto and Alcalá 2011. In press.; Alonso et al. 2009; Zhou and Gan 2008; Mencar and Fanelli 2008) some reviews about interpretability, and the way in which this can be achieved, can be found. Sometimes, they appear associated with the concepts of complexity and explanation capability (Ishibuchi et al. 2009a), which can be considered as indirect measures to evaluate the interpretability. In some works, for instance (Setnes and Babuška 2001; Roubos and Setnes 2001; Jin 2000; Yen and Wang 1999; Setnes et al. 1998), the reduction of the complexity system can imply a better interpretability of the fuzzy system. In any case, the interpretability of fuzzy systems is still a point of discussion amongst researchers (Gacto and Alcalá 2011. In press.; Ishibuchi et al. 2009b; Ishibuchi and Nojima 2009).

One possible approach is based on genetic algorithms (GAs) used to generate *Fuzzy Rule-Based Systems* (FRBSs). Then they are called *Genetic Fuzzy Systems (GFS)* (Herrera 2008; Cordón et al. 2001). In this case one alternative is to use Multi-Objective Genetic Algorithms (MOEAs) to obtain a good trade-off between accuracy and interpretability, since both concepts are contradictory (Gacto et al. 2010; Pulkkinen and Koivisto 2010; Botta et al. 2009; Alcalá et al. 2009; Gacto et al. 2009; Pulkkinen and Koivisto 2008; Cococcioni et al. 2007; Ishibuchi and Nojima 2007; Alcalá et al. 2007b; González et al. 2007; Ishibuchi and Yamamoto 2004; Jimenez et al. 2003; Suzuki and Furuhashi 2003).

An interesting approach is to use MOEAs for genetic rule selection. Initial models are usually made up for a high number of fuzzy rules, then a rule subset can be selected to represent the system in a more compact way with a better trade-off in accuracy and interpretability . If irrelevant, redundant, erroneous and conflictive rules are removed, then the rule set is more compact and more interpretable, even more accurate. The first contribution to this approach is in (Ishibuchi et al. 1997).

This paper is focused on the interpretability improvement of well-known algorithms for fuzzy modeling through a *complexity reduction based on an accuracy-interpretability trade-off*. This aim is carried out through a bi-objective (accuracy and interpretability) genetic approach via rule selection, and a simple and understandable set of indices concerning accuracy-interpretability.

This proposal takes advantage of two well-known fuzzy modeling algorithms, FasArt (Fuzzy Adaptive System ART based) (Cano Izquierdo et al. 2001; Sainz Palmero et al. 2000) and NefProx (Neuro-Fuzzy Function Approximation) (Nauck and Kruse 1999), which are very popular in technical and engineering domains, improving their fuzzy performance through the reduction of their complexity in order to achieve a better accuracy-interpretability trade-off using MOEAs. In this way, the fuzzy rules describing the behavior of the modeled problem are accessible in a more interpretable way, and thus extra knowledge and performance from these fuzzy models are obtained.

This procedure has been applied to nine real-world regression problems. Several fuzzy models with a good accuracy are generated by the two neuro-fuzzy modeling algorithms. They are used to check the proposal under different restrictions of complexity and fuzziness: two models (compact model and complex model) are generated for each case study and fuzzy modeling algorithm. Each has its own performance in the accuracy-interpretability trade-off. Both types of models allow the approach proposed to be checked in two different fuzzy contexts.

The paper is organized as follows: first, in Section 2, a brief description of alternative points of view about fuzzy modeling, interpretability and accuracy are given. Then, in Section 3, several complexity and interpretability measures are presented as a part of an index that aggregates them in a single measure. In Section 4, the methodology used in this work is described. Some experimental studies are carried out and the main results obtained are discussed in Section 5. Finally, in Section 6, the most interesting conclusions obtained from this work are set out. Additionally, Appendix A includes some numeric results obtained in the experiments.

2 Preliminaries: Fuzzy Modeling Interpretability and Neuro-Fuzzy Systems

This section first introduces the fuzzy modeling and the accuracy-interpretability trade-off problem. Then, the neuro-fuzzy systems considered in this work are commented.

2.1 Fuzzy Modeling: accuracy vs. interpretability

Fuzzy modeling implies several contradictory points of view, such as accuracy and interpretability. This section introduces the accuracy-interpretability trade-off problem and the MOEAs as a tool for managing this trade-off.

2.1.1 Accuracy vs. Interpretability

Since one of the most relevant features of a fuzzy system must be its capacity to explain the system behaviour in an understandable way, a good balance between accuracy and interpretability must be achieved in the fuzzy modeling.

One of the initial tasks needed to find a good accuracy-interpretability trade-off is to define how these factors can be measured (Gacto and Alcalá 2011. In press.), tested or checked:

- The way in which fuzzy modeling deals with *accuracy*, or the capacity to faithfully represent the real system, is based on error.
- But when the *interpretability* idea is involved, then there is no a concise, clear and unique way about the meaning of the interpretability concept and its formulation.

A review of the most representative works and points of view on interpretability concepts in the specialized literature can be found in (Mencar and Fanelli 2008). In (Zhou and Gan 2008), a review of interpretability in fuzzy system modeling is given. Here, the authors consider two levels of interpretability in order to analyze this concept in the fuzzy models:

- *Low-level Interpretability*, the fuzzy set level is involved in the analysis.
- *High-level Interpretability*, fuzzy rule level is considered.

Another complementary point of view on interpretability of FRBSs is described in (Alonso et al. 2009). In this paper, the framework is based on the previous reference (Zhou and Gan 2008) with extra elements and concepts: the authors extend the Low/High Level Interpretability concepts to new levels named *Description (System Structure Readability)* and *Explanation (System Comprehension)*.

Finally, a review on the used measures to assess the interpretability of linguistic FRBSs can be found in (Gacto and Alcalá 2011. In press.), where a particular taxonomy based on the structure of linguistic based approaches is given.

According to (Gacto and Alcalá 2011. In press.; Gacto et al. 2010), there are two main kinds of interpretability measures for linguistic FRBSs depending on the type of interpretability:

1. *Complexity-based interpretability*, in which the reduction of the complexity systems can imply a better interpretability of the fuzzy system. Some measures used in these cases are the number of rules, variables, labels per rule, etc.
2. *Semantic-based interpretability*, in which the semantic associated with the membership functions is the priority. The semantic can be seen as either the semantic integrity or the other properties of membership functions (distinguishability, coverage, fuzzy ordering, etc.).

Most of the available works from the scientific literature introduce the interpretability of the fuzzy systems based on the interpretability of the rules, their variables, the fuzzy partitions and the membership functions. In fact, in Mamdani-type FRBSs are usually the most interpretable ones.

In short, some approaches that deal with the generation of fuzzy systems based on an adequate accuracy-interpretability trade-off are:

- Algorithms taking into account the idea of accuracy-interpretability during the generation of the fuzzy system. (Alonso and Magdalena 2010; Cpalka 2009; Ishibuchi and Nojima 2007; González et al. 2007; Alcalá et al. 2007a, 2006; Mikut et al. 2005; Delgado et al. 2003; Guillaume and Charnomordic 2003; Jimenez et al. 2003; Suzuki and Furuhashi 2003; Fiordaliso 2003; Espinosa and Vandewalle 2000)
- Interpretability of the fuzzy systems is improved or maintained through a postprocessing stage. Here, two approaches are found:
  - Similarity measures of the fuzzy systems, rules, etc. are used in the approach mentioned in (Zong-Yi et al. 2008; Setnes 2003; Jin 2000; Setnes et al. 1998).
  - Orthogonal Transformations, Here the interpretability is improved by a complexity reduction ruled by orthogonal transformations (Destercke et al. 2007; Zong-Yi et al. 2005; Setnes 2003; Setnes and Babuška 2001; Yen and Wang 1999).

2.1.2 Genetic Fuzzy Systems: MOEAs as a tool for managing the trade-off

A GFS is basically a fuzzy system augmented by a learning process based on evolutionary computation, which includes genetic algorithms, genetic programming, and evolutionary strategies, among other evolutionary algorithms (EAs). A general taxonomy of GFS is introduced in (Herrera 2008) where GAs could be used in two different ways to generate fuzzy systems: tuning and learning. In the scientific literature, it is possible to find papers and contributions that use MOEAs to improve the accuracy-interpretability trade-off in these two ways [1]:

- Genetic Tuning: to improve the FRBS performance without changing the existing rule base. (Gacto et al. 2010, 2009; Alcalá et al. 2007b)
- Genetic Learning: to learn fuzzy rules or other components in the FRBS, including *genetic rule selection*. (Alcalá et al. 2009; Ishibuchi and Nojima 2007; Cococcioni et al. 2007)

Within this taxonomy an alternative is to use MOEAs to select a subset of cooperative rules from a set of candi-

[1] A complete list of papers in this area is available from the web page http://www.iet.unipi.it/m.cococcioni/emofrbss.html

date fuzzy rules. Here, the objective is to obtain a more reduced rule set, improving its original performance, in this case the accuracy and interpretability. One of the key aspects, amongst the different proposals based on rule selection, is the way in which interpretability is measured.

Several alternatives have been considered to measure interpretability in genetic rule selection:

- The most classic and first contribution in this area can be found in (Ishibuchi et al. 1995). The authors apply a mono-objective fitness function to maximize the number of patterns classified correctly, minimizing the number of rules. In (Ishibuchi et al. 1997), a multi-objective genetic algorithm is used in the same way, and in (Ishibuchi et al. 2001), a third objective is included to minimize the length of the rules. Other contributions in the same scheme are (Ishibuchi and Yamamoto 2004; Nojima and Ishibuchi 2009; Alcalá et al. 2011).
- In (Gacto et al. 2010), a new semantic-based interpretability index called GM3M is used to do genetic tuning. The index is based on membership functions displacement, symmetry and area similarity measures. The authors use a multi-objective genetic algorithm to do selection and tuning simultaneously: minimizing the complexity of the system through the number of rules, measuring the accuracy through the error and maximizing the interpretability through GM3M.
- In (Galende et al. 2008, 2009), the authors apply the ideas commented previously from other points of view, using mono-objetive algorithms and an incremental process to measure the interpretability.

2.2 Neuro-Fuzzy Systems

Neuro Fuzzy systems are a very popular approach to generate FRBSs, taking advantage of the learning capacity of Artificial Neural Networks (ANN) and the explanatory capacity of Fuzzy Logic. In this work, two different neuro-fuzzy systems are used: FasArt (Cano Izquierdo et al. 2001) and NefProx (Nauck and Kruse 1999). Both use Mamdani-type rules to obtain models with high precision, but the way in which each generates the fuzzy model is different, so their performance is different too. If the taxonomy for FBRSs described in (Herrera 2008) is taken into account, Nefprox can be considered a linguistic model and FasArt an approximate model. Another classification can be done if (Casillas et al. 2003a,b) is considered: FasArt and NefProx are Mamdani-type FRBSs for precise modeling.

2.2.1 Neuro-Fuzzy System FasArt

The FasArt model (Cano Izquierdo et al. 2001; Sainz Palmero et al. 2000) is a neuro fuzzy system based on the Adaptive Resonance Theory (ART).

FasArt introduces an equivalence between the activation function of each FasArt neuron and a membership function. In this way, FasArt is equivalent to a Mamdani-type FRBSs with: Fuzzification by single point, Inference by product, and Defuzzification by average of fuzzy set centers. A full description of this model can be found in (Cano Izquierdo et al. 2001) and (Sainz Palmero et al. 2000).

The FasArt system has been used in several previous works (Sainz Palmero et al. 2005; Sainz et al. 2004) for modeling, fault detection, pattern recognition, etc. with reasonable results when its accuracy as a fuzzy model is involved; but when other aspects, such as rule interpretability, are important, then some problems appear: proliferation of rules, fuzzy sets, etc., so this system is an adequate instance for checking this proposal, taking advantage of the knowledge learnt and stored by FasArt for each problem involved. This aspect is important in analyzing the results and its comparison with other algorithms. Here the fuzzy sets or the number of rules are not defined by the user, but by its own algorithm in a non supervised way, in comparison with other fuzzy modeling algorithms, i.e. NefProx.

Most of these aspects are common for models based on ART Theory, and they have been treated in different works (Gómez-Sánchez et al. 2002; Parrado-Hernández et al. 2003).

2.2.2 Neuro-Fuzzy Function Approximation

NefProx[2] (Nauck and Kruse 1999) is a neuro-fuzzy algorithm based on supervised learning for the function approximation. The user fixed the parameters and the learning algorithm generates fuzzy rules from data, minimizing the error.

In this paper NefProx, is used to generate Mamdani-type fuzzy rules with triangular membership functions, Inference by max-min and Defuzzification by mean of maximum. Previously, the user fixes the size of the rule base depending on the coverage of training data and based on an initial fuzzy partition. Here, the fuzzy partitions are uniformly distributed, so the interpretability of these elements is higher in comparison with the previous algorithm.

3 A Proposal for Aggregating Complexity and Interpretability Indices in a Common Measure

In this section a measure for preserving the interpretability is proposed. This measure is defined via the aggregation of different simple indices based on complexity and similarity concepts.

In general, in this context, the complexity reduction is a way to improve the interpretability of the fuzzy systems, and thus a good level of interpretability implies a lower level of

[2] http://fuzzy.cs.uni-magdeburg.de/nefprox/

complexity. When the complexity is reduced, then the interpretability, or at least some aspects of it, is improved.

In general Mamdani-type rules are the most interpretable ones, but this is not enough to achieve an adequate degree of explanation of the contained knowledge.

If a good level of interpretability is desired, which implies a lower complexity, then the fuzzy sets, rules, partitions, structure model, operators, etc. of the fuzzy model must reasonably show this performance. If not, as usual in (precise) fuzzy modeling, these model elements should be improved to achieve a better performance focused on accuracy and interpretability/complexity trade-off.

In this paper, the interpretability is based on the complexity but also on the distinguishability concept (Chen and Linkens 2004). The interpretability measures proposed for this goal are based on the similarity of fuzzy sets and rules.

The index and measures proposed in this work are conceptually linked with the formal framework described in (Zhou and Gan 2008; Alonso et al. 2009; Gacto and Alcalá 2011. In press.) from a practical engineering point of view. This proposal is open to remove metrics/measures or to adding new indices to set up the function according to the desired performance needed for the fuzzy model.

The concepts for measuring the interpretability through the reduction of the complexity are:

- Compactness (Eq. 1), a lower *number of rules* can be adequate for a lower complexity and better interpretability level of the model rule set.

$$Compactness = Number\ of\ rules \tag{1}$$

- *Similarity* amongst rules (Eqs. 2, 3). This index must be minimized to improve the rule distinguishability.
  First, the similarity between two rules is calculated using eq. 2.

$$\begin{aligned} &S_k(R_i,R_j) = \frac{\sum_{i,j} R_{ik}(x) \wedge R_{jk}(x)}{\sum_{i,j} R_{ik}(x) \vee R_{jk}(x)} = \frac{\sum_{i,j} \min(R_{ik},R_{jk})}{\sum_{i,j} \max(R_{ik},R_{jk})} \\ &\forall 1 \le i < j \le RuleNumber \\ &\forall 1 \le k \le AntecedentNumber \end{aligned} \tag{2}$$

  Then the global similarity of the fuzzy rule set is calculated by the arithmetic mean (Eq. 3) of antecedent similarities (Eq. 2):

$$\begin{aligned} &Similarity = F_{i,j}(F_k(S_k(R_i,R_j))) \\ &F \Rightarrow ArithmeticMean \\ &\forall 1 \le i < j \le RuleNumber \\ &\forall 1 \le k \le AntecedentNumber \end{aligned} \tag{3}$$

  Other authors have used this measure (Eq. 2) successfully to generate fuzzy rules with low complexity (Setnes et al. 1998; Casillas et al. 2003b; Setnes 2003; Jimenez et al. 2003). Nevertheless, some other options can be used to measure similarity (Jin et al. 1999; Jin 2000).

- *Redundancy* (Eq. 4), based on the previous equations, the whole redundancy of the fuzzy rule set is calculated through the number of pairs of rules (*Card*) whose redundancy is higher than a threshold, β_R ($0 < \beta_R < 1$), for antecedents (S_{kA}) and consequents (S_{kC}). Redundant rules must be avoided.

$$\begin{aligned} &Redundacy = \frac{Card(S_{kA}(R_i,R_j) > \beta_R\ AND\ S_{kC}(R_i,R_j) > \beta_R)}{(RuleNumber-1)!} \\ &\forall 1 \le i < j \le RuleNumber \\ &\forall 1 \le kA \le AntecedentNumber \\ &\forall 1 \le kC \le ConsequentNumber \end{aligned} \tag{4}$$

- *Consistency*, avoiding *incoherent* rules (Eq. 5) the system is more understandable. A threshold of incoherency, β_I, is defined as $\beta_I = 1 - \beta_R$ to measure the "no similarity" amongst consequents.

$$\begin{aligned} &Incoherency = \frac{Card(S_{kA}(R_i,R_j) > \beta_R\ AND\ S_{kC}(R_i,R_j) < \beta_I)}{(RuleNumber-1)!} \\ &\forall 1 \le i < j \le RuleNumber \\ &\forall 1 \le kA \le AntecedentNumber \\ &\forall 1 \le kC \le ConsequentNumber \end{aligned} \tag{5}$$

- Completeness or *No-Coverage*(Eq. 6). Complete fuzzy partition involves minimizing no coverage. The no coverage is calculated as the "no coverage" arithmetic mean for each variable.

$$\begin{aligned} &NoCoverage = ArithmeticMean(NoCoverPartition_k) \\ &NoCoverPartition_k = \frac{NoCoverPoints}{TotalPoint} \\ &NoCoverPoints\ if\ Activation\ Level <= \beta_C \\ &\forall 1 \le k \le AntecedentNumber + ConsequentNumber \end{aligned} \tag{6}$$

In every case, a low value of these measurements has a positive influence on reducing complexity, thus improving interpretability. First of all, these metrics are balanced using λ_j, then they are combined using the *Arithmetic Mean* to estimate a global value for the interpretability, mainly based on complexity concepts, of the model (Eq 7). λ_j is used to weight the metrics and it can take values between 0 and 1, where 1 indicates that the metric is highly relevant in $Inter_C$ and 0 indicates that the metric is ignored. Moreover, each metric is normalized ($_{nor}$) to a common range.

$$\begin{aligned} &Inter_C = ArithmeticMean(\lambda_{nr} * RuleNumber_{nor}, \\ &\quad \lambda_s * Similarity_{nor}, \lambda_r * Redundancy_{nor}, \\ &\quad \lambda_i * Incoherency_{nor}, \lambda_{nc} * NoCoverage_{nor}) \\ &\lambda_j \in (0,1) \end{aligned} \tag{7}$$

In this work the normalization used has been (Eq. 8):

$$Index_{nor} = 1 - \frac{Index_{Original} - Index_{Current}}{Index_{Original}} \tag{8}$$

where *Index* refers to the particular metric that is being normalized, i.e., to *RuleNumber*, *Similarity*, etc. Other normalization options were used, but this one was the best for the genetic approach used in this work. Here, the values of the original fuzzy model are involved in the normalization.

4 Methodology: A proposal to improve the trade-off between accuracy and interpretability

In this paper, the proposed methodology is focused on improving the (precise) fuzzy modeling from the point of view of the interpretability, preserving a good level of accuracy. Here, an approximate and a linguistic fuzzy rule generation algorithm are taken into consideration and then, based on this proposal, a better accuracy-interpretability trade-off is reached.

This goal is reached using a general post-processing fuzzy rule selection through a bi-objective genetic approach and a simple, and easily interpretable, set of measures of interpretability and accuracy.

In this way, the interpretability, or explaining capacities, of the base precise (approximate and linguistic) fuzzy models is improved, so the complexity is reduced, preserving or even improving the model's accuracy.

This methodology in two stages is described in the next sections, describing in detail the MOEA applied in the post-processing stage for this rule selection.

4.1 A methodology in two stages

There are two stages in the methodological proposal introduces in this work:

1. First some base, and precise, fuzzy models are generated containing a good set of candidate fuzzy rules.
2. Then, a MOEA is applied to carry out a rule selection for improving interpretability while most of their accuracy is preserved.

In the first stage, base models are generated based on two well-known neuro-fuzzy algorithms: FasArt and NefProx, the first is an approximate FBRS algorithm and the second is a linguistic one but both try to obtain fuzzy models as accurate as possible, and they do not pay attention on other aspects. Two type of base models have been generated, each has its own performance in accuracy, interpretability and fuzzy nature, in order to test the proposal under several contexts. This stage is open to any rule-based algorithm described in the specialized literature.

The second stage of the methodology, the post-processing fuzzy rule selection through a bi-objective genetic approach, is explained in detail in the next subsections.

4.2 Multi-Objective Evolutionary Algorithm for Rule Selection

The second stage of this proposal is based on a bi-objective ($Inter_C$ and Acc) genetic approach. This is implemented through the well-known NSGA-II algorithm (Deb et al. 2002). Here, any other optimization technique could be taken into consideration.

In the next subsections, the fitness functions are formulated and the rest of the parameters needed to run the NSGA-II algorithm are described.

4.2.1 Objectives

The fitness functions are shown in Eq. 9. Here, the model performance to be reached can be considered through the addition of a new index and its relevancy by user.

$$\begin{aligned} &max(Accuracy) = min(Error)\\ &max(Interpretability) = min(Inter_C) = \\ &\quad = min(AritmeticMean(\lambda_j * InterpretabilityIndex_j)) \end{aligned} \tag{9}$$

The bi-objective genetic algorithm must reach a fuzzy model with lower complexity and better accuracy-interpretability trade-off by:

- Maximizing the accuracy: this is evaluated by minimizing Mean Squared Error (MSE) (Eq 10). This is the most usual way for this goal.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_i')^2 \tag{10}$$

- Maximizing the interpretability of the fuzzy model: this is guided by interpretability concepts, minimizing the $Inter_C$ index defined in Section 3. In this way, a more compact and interpretable model can be obtained through a rule selection, avoiding non-relevant, redundant and incoherent rules, incomplete fuzzy partitions, etc. In order to be able to use the indices proposed in Section 3, some thresholds must be tuned by the user. Considering (Setnes 2003; Roubos and Setnes 2001), in this work the thresholds used are:
  - $\beta_R = 0.8$ for redundancy.
  - $\beta_I = 1 - \beta_R = 0.2$ for incoherency.
  - $\beta_C = 0$ as activation level for coverage.
  - $\lambda_i = 1$ since for this work it is considered that all individual indices in $Inter_C$ have the same importance.

4.2.2 Coding Scheme, Populations and Genetic Operators

In order to run NSGA-II for the rule selection, other relevant MOEA aspects has to be considered. These are described next:

- Individuals are coded by *binary-coding*: $S = s_1 s_2 \ldots s_N$ (N is the number of initial rules), where $s_q = 0$ shows that the rule R_q is not included, while $s_q = 1$ shows the rule is present.
- Genes take the value 1 for all of the individuals of the *initial population* in order to achieve a progressive extraction of the worst rules.

Table 1 NSGA-II Parameters

Genetic operator	
Selection	Binary Tournament
Crossover	HUX P_c=0.9
Mutation	Classical P_m=0.7
Other options	
Population size	100
Evaluations	50000

– *Genetic operators* selected according to the final objective (see Table 1):
  – Binary tournament for *selection*.
  – HUX (Eshelman 1991) is used to *crossover* with probability P_c. The HUX crossover exactly interchanges the mid of the alleles that are different in the parents (the genes to be crossed are randomly selected among those that are different in the parents). This operator ensures the maximum distance of the offspring to their parents (exploration).
  – Classical *mutation* with probability P_m. This operator changes a gene value at random, sets to zero a gene with probability P_m and sets to one with probability $1-P_m$. This operator was proposed for rule selection in (Ishibuchi et al. 1997) and it promotes the elimination of the rules since all individuals the initial population contained all candidates rules.
– In addition, if one individual (subset of candidates rules) do not cover some examples previously covered then both fitness objectives are *penalized*. Then these solutions go (at least) to the second non-dominated front.
– The *stopping criterion* is the number of evaluations.

We have used the *implementation* of the NSGA-II algorithm obtained from Kanpur Genetic Algorithms Laboratory webpage [3], adapting some genetic operators and the evaluation of the fitness function. Table 1 shows the parameters used to run NSGA-II. NSGA-II.

5 Experimental Study

In order to check the performance of the proposal introduced in this work, nine real-world data sets from the KEEL Project (Alcalá-Fdez et al. 2009, 2011) [4] have been used:

1. Plastic Strength (PLA): 3 variables, 1650 records.
2. Quake (QUA): 4 variables, 2178 records.
3. Electrical Maintenance (ELE): 5 variables, 1056 records.
4. Abalone (ABA): 9 variables, 4177 records.
5. Stock prices (STP): 10 variables, 950 records.
6. Weather Ankara (WAN): 10 variables, 1609 records.
7. Weather Izmir (WIZ): 10 variables, 1461 records.
8. Mortgage (MOR): 16 variables, 1049 records.
9. Treasury (TRE): 16 variables, 1049 patterns.

First of all, the base fuzzy models are generated by the neuro-fuzzy algorithms, then the multi-objective rule selection is carried out, generating a Pareto Front for each dataset and for each trial.

For all the experiments, a 5-fold cross validation model is adopted (each fold contained 20% of the records) using four folds for training and one for testing. For each of the possible five different partitions (train/test), both stages of the algorithm were run 6 times, considering each time a different seed for the random-number generator. Therefore, we consider the average results of 30 runs.

Only three representative models (according to the objectives) from the Pareto front are considered:

1. The most interpretable model: Best $Inter_C$.
2. The most accurate model: Best Acc.
3. The median accuracy-interpretability model: Median $Acc-Inter_C$.

This procedure for comparison was proposed and used in (Gacto et al. 2010) and (Alcalá et al. 2009). Finally, in order to know the statistical significance of the results, the mean values (over 30 runs) are calculated on the three representative points and non-parametric statistical tests (Demšar 2006; García and Herrera 2008; García et al. 2009a,b) are carried out [5].

The experimental framework is organized as follow:

– The parameters used to generate the fuzzy models by FasArt and NefProx, and their performance, are described in Section 5.1. In fact, two models, one more compact and one more complex, are generated for each algorithm and data set. Here the error is the only criterion to obtain these models.
– Results obtained by compact fuzzy models and Non-parametric Wilcoxon's signed-rank tests (Zar 1999; Sheskin 2003) are shown in Section 5.2.1.
– Section 5.2.2, with the same scheme that in Section 5.2.1, shows the results and the non-parametric statistical tests obtained by complex fuzzy models.
– Finally, some global conclusions are commented in Section 5.2.3.

5.1 Base Models: FasArt and NefProx

Several fuzzy models are generated based on FasArt and NefProx algorithms for each data set considered. Both algorithms allow to achieve a good accuracy, taking only in consideration the model error during the learning. Furthermore,

[3] http://www.iitk.ac.in/kangal/codes.shtml
[4] http://sci2s.ugr.es/keel/datasets.php
[5] http://sci2s.ugr.es/sicidm

Table 2 FasArt and NefProx Codification and Parameters of the Fuzzy Models

	FasArt	NefProx
Compact	Models-1	Models-3
	$\rho_A = \rho_B = 0.3$	$msf = 5$
	$\gamma_A = \gamma_B = 10$	$maxR = -1$
Complex	Models-2	Models-4
	$\rho_A = \rho_B = 0.9$	$msf = 7$
	$\gamma_A = \gamma_B = 10$	$maxR = -1$

Table 3 Performance of the FasArt Fuzzy Models

Model	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-1	3.483	3.621	48.6	0.238	0.001	0.016	99.9
PLA-2	2.783	2.821	96.6	0.162	43e-5	0.006	100
QUA-1	0.050	0.054	119.8	0.220	9e-5	0.002	98.0
QUA-2	0.046	0.050	243.8	0.265	20e-5	0.002	98.1
ELE-1	117867	158820	92.6	0.225	0.004	0.001	96.3
ELE-2	56584	100229	129.8	0.266	0.003	0.001	96.9
ABA-1	6.872	7.683	122.8	0.277	32e-5	0.001	98.6
ABA-2	5.033	6.247	298.0	0.331	41e-5	0.001	100
STP-1	2.091	2.270	101.8	0.195	0	16e-5	100
STP-2	0.426	0.698	163.6	0.185	3e-5	16e-5	100
WIZ-1	5.452	16.555	221.6	0.335	3e-5	0	99.9
WIZ-2	1.788	21.934	466.4	0.360	0.6e-5	0	100
WAN-1	9.813	21.970	231.8	0.304	0	0	99.9
WAN-2	2.593	28.312	537.6	0.312	0.3e-5	0.1e-5	99.9
MOR-1	1.041	1.258	52.6	0.299	0	0	99.9
MOR-2	0.085	0.352	92.2	0.284	63e-5	8e-5	100
TRE-1	0.908	1.339	49.6	0.292	50e-5	0	99.8
TRE-2	0.150	0.552	76.6	0.289	8e-5	13e-5	100

two types of models for FasArt and NefProx are carried out with different complexity and fuzzy performance but good accuracy. In order to identify the models used in the experimental study, they are numbered as shown in Table 2: 1 for compact models generated by FasArt, 2 for complex models generated by FasArt, 3 for compact models generated by NefProx and 4 for complex models generated by NefProx. The FasArt and NefProx parameters used to generate these base models are shown in the same Table 2:

- $\rho_A = \rho_B$ is the vigilance parameter used by FasArt.
- $\gamma_A = \gamma_B$ is the fuzzification rate in FasArt.
- msf is the number of fuzzy sets for input and output variables used by NefProx.
- $maxR$ is the maximum number of rules generated by NefProx: -1 means all rules found in the data are used (no evaluation), 0 means automatic evaluation (use percent = 0.75).

Tables 3 and 4 summarize the main performance aspects of these base fuzzy models generated applying the methodology described previously. The indexes shown in the tables are the mean squared error for training (MSE_{tra}) and test (MSE_{tst}), the rule number (RN), the similarity (S), the redundancy (R), the incoherency (I) and the percentage of completeness ($C(\%)$).

Models 1 and 3, on FasArt and NefProx respectively, are more compact obtaining a good accuracy with low number of rules. The similarity amongst rules could be lower, there is little redundancy and incoherency and the fuzzy partitions are complete. The other models (2 and 4) are more complex, achieving a high and better accuracy with additional rules. The NefProx complex models have less similarity, redundancy and incoherency than compact models while FasArt complex models have sometimes less similarity, redundancy and incoherency than compact models and other times more with similar completeness in all cases. These models are representative of the precise fuzzy modeling, whose main objective is to obtain a system as accurate as possible.

Table 4 Performance of the NefProx Fuzzy Models

Model	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-3	3.208	3.222	17.0	0.210	0	0	100
PLA-4	2.606	2.636	31.0	0.161	0	0	100
QUA-3	0.039	0.041	55.8	0.272	0	0	100
QUA-4	0.035	0.037	98.0	0.229	0	0	100
ELE-3	620411	622331	79.6	0.286	0	0	99.4
ELE-4	556228	598472	100.2	0.227	0	0	99.4
ABA-3	6.653	7.231	272.2	0.433	0.012	0.007	99.3
ABA-4	5.636	6.370	500.0	0.350	0.006	0.003	95.5
STP-3	2.248	2.493	303.4	0.285	0.017	0.007	100
STP-4	1.307	1.727	433.8	0.215	0.009	0.003	100
WIZ-3	9.958	13.837	500.0	0.503	0.024	0.009	99.7
WIZ-4	10.103	17.471	500.0	0.442	0.011	0.004	100
WAN-3	12.227	15.920	500.0	0.449	0.017	0.006	98.3
WAN-4	21.836	33.620	500.0	0.390	0.008	0.003	98.5
MOR-3	0.716	0.729	170.0	0.359	0.046	0.011	100
MOR-4	0.337	0.510	301.6	0.255	0.026	0.006	99.9
TRE-3	1.029	1.087	170.6	0.350	0.047	0.011	100
TRE-4	0.491	0.673	305.4	0.243	0.028	0.004	100

In Table 5 this performance is compared with the results obtained by the Wang & Mendel algorithm (Wang and Mendel 1992) shown in (Gacto et al. 2010). In general, the FasArt and NefProx models show a higher accuracy but more rules, and, in general, poorer interpretability as it is usual for precise and approximate models.

Then, these fuzzy models are improved in order to reach better fuzzy models using a more adequate accuracy-interpretability trade-off through the proposal described in this work.

5.2 Improved Fuzzy Models: Results and Analysis

The results and their analysis are organized according to the results obtained for compact base fuzzy models and the results for the complex base fuzzy models. In both cases, the measurements of *Acc* and $Inter_C$ for the improved models are presented, and some *Non-Parametric Statistical Tests* are carried out in order to check the statistical significance of these results. Finally, a global analysis of the results is introduced.

Here, the Wilcoxon test is run taking in consideration the three characteristic models of the Pareto front generated by FasArt and NefProx for each compact and complex base models. This test is used for detecting significant differences between two sample means: it is analogous to the paired t-

Table 5 Wang & Mendel vs Neuro-Fuzzy Systems. The results shown for Wang & Mendel have been obtained from (Gacto et al. 2010)

Compact	Wang & Mendel (Wang and Mendel 1992)			FasArt (Model-1)			NefProx (Model-3)		
Models	RN	MSE_{tra}	MSE_{tst}	RN	MSE_{tra}	MSE_{tst}	RN	MSE_{tra}	MSE_{tst}
PLA	14.8	6.868	7.114	48.6	**3.4834**	**3.6205**	17	**3.208**	**3.222**
QUA	53.6	0.0516	0.0534	119.8	**0.0500**	0.0540	55.8	**0.039**	**0.041**
ELE	65	115212	115868	92.6	117866.65	158819.67	79.6	620411.26	622331.12
ABA	68	16.814	16.844	122.8	**6.8719**	**7.6829**	272.2	**6.653**	**7.231**
STP	122.8	18.148	18.084	**101.8**	**2.0909**	**2.2696**	303.4	**2.248**	**2.493**
WIZ	104.8	13.888	14.736	221.6	**5.4522**	16.5549	500	**9.958**	**13.837**
WAN	156	32.126	32.786	231.8	**9.8128**	**21.9696**	500	**12.227**	**15.920**
MOR	77.6	1.97	1.946	**52.6**	**1.0414**	**1.2580**	170	**0.716**	**0.729**
TRE	75	3.272	3.262	**49.6**	**0.9083**	**1.3392**	170.6	**1.029**	**1.087**
Complex	Wang & Mendel (Wang and Mendel 1992)			FasArt (Model-2)			NefProx (Model-4)		
Models	NR	MSE_{tra}	MSE_{tst}	NR	MSE_{tra}	MSE_{tst}	NR	MSE_{tra}	MSE_{tst}
PLA	14.8	6.868	7.114	96.6	**2.7825**	**2.8207**	31	**2.606**	**2.636**
QUA	53.6	0.0516	0.0534	243.8	**0.0455**	**0.0501**	98	**0.035**	**0.037**
ELE	65	115212	115868	129.8	**56584.25**	**100229.04**	100.2	556227.86	598472.35
ABA	68	16.814	16.844	298	**5.0332**	**6.2466**	500	**5.636**	**6.370**
STP	122.8	18.148	18.084	163.6	**0.4265**	**0.6981**	433.8	**1.307**	**1.727**
WIZ	104.8	13.888	14.736	466.4	**1.7878**	21.9338	500	**10.103**	17.471
WAN	156	32.126	32.786	537.6	**2.5934**	**28.3117**	500	**21.836**	33.620
MOR	77.6	1.97	1.946	92.2	**0.0855**	**0.3524**	301.6	**0.337**	**0.510**
TRE	75	3.272	3.262	76.6	**0.1501**	**0.5524**	305.4	**0.491**	**0.673**

test in non-parametric statistical procedures. In general, the test asks about (H_0): do two samples come from populations with the same distributions? and is based on ranks of the differences between pairs of data.

In order to have well-defined differences in MSE and NR a normalized differences DIFF (using Eq. 11) are adopted, where Mean(x) represent either the MSE or the NR means that are obtained by the x algorithm. This difference expresses the improvement in percentage of the reference algorithm (Gacto et al. 2010; Alcalá et al. 2009). For the rest of simple indices (S,R,I,NC and $Inter_C$) it is not necessary.

$$DIFF = \frac{Mean(Other) - Mean(ReferenceAlgorithm)}{Mean(Other)} \tag{11}$$

5.2.1 Compact Models

This section introduces the results obtained for the compact fuzzy models (named Models-1 for FasArt and Models-3 for NefProx).

Table 6 shows the averaged results obtained in the three characteristic models of the Pareto Front considered in this work over 30 runs for each case study considered. Specifically, the table shows the mean of the proposed index $Inter_C$ and mean squared error for the test, MSE_{tst}, for each one of these three models taken into account: Best $Inter_C$, Median $Acc - Inter_C$ and Best Acc. The first line shows the base model (I), while the second line shows the performance of the final improved model (F).

It is possible easily to see that $Inter_C$ is better, being able to improve up to 71.53%. The only exception is the Mor-1 case, in Best Acc model where the interpretability is reduced: $Inter_C$ increases from 0.480 to 1.548. MSE_{tst} es preserved in the same order of magnitude for all cases.

In Appendix A, Table 12 shows the mean of each individual measurements: the mean squared error for training (MSE_{tra}) and test (MSE_{tst}), the mean rule number (RN), the mean similarity (S), the mean redundancy (R), the mean incoherency (I) and the mean percentage of completeness ($C(\%)$).

The Wilcoxon test for the FasArt Compact Models (Table 7) accepts that:

– Best $Inter_C$ model; here the interpretability index is improved and the accuracy is slightly worse. This is usual: the contradictory dilemma Accuracy vs. Interpretability and Complexity. Although in some cases the accuracy is worse, its value can be acceptable because it acceptably stays low.
– Median $Acc - Inter_C$ models have an accuracy similar to the base models, and the interpretability index is improved. The accuracy has been preserved without relevant loss of precision.
– Best Acc models have accuracy and interpretability indexes that are not statistically different. Here, the no variation in the interpretability can be considered paradoxical if the Rule Number, Similarity, etc... are individually analyzed in Table 12. This result is a consequence, basically, of the way in which FasArt generates fuzzy partitions, and the equality of relevance given to each metric in the index defined: rule number vs. completeness.

The Wilcoxon test results obtained by the NefProx compact models (Table 8) shows:

Table 6 Performance of the Improved Compact Fuzzy Models

FasArt Models	Best $Inter_C$ MSE_{tst}	$Inter_C$	Median $Acc - Inter_C$ MSE_{tst}	$Inter_C$	Best Acc MSE_{tst}	$Inter_C$
PLA-1(I)	**3.621**	0.800	3.621	0.800	3.621	0.800
PLA-1(F)	3.718	**0.228**	**3.073**	**0.260**	**2.688**	**0.472**
QUA-1(I)	0.054	0.840	0.054	0.840	0.054	0.840
QUA-1(F)	**0.042**	**0.448**	**0.039**	**0.483**	**0.038**	**0.546**
ELE-1(I)	**158820**	1.000	**158820**	1.000	**158820**	**1.000**
ELE-1(F)	166133	**0.884**	159943	**0.916**	160842	1.017
ABA-1(I)	7.683	1.000	7.683	1.000	7.683	1.000
ABA-1(F)	**6.881**	**0.494**	**5.902**	**0.565**	**5.788**	**0.633**
STP-1(I)	**2.270**	0.520	**2.270**	0.520	**2.270**	**0.520**
STP-1(F)	2.815	**0.346**	2.745	**0.368**	2.405	0.537
WIZ-1(I)	**16.555**	0.520	**16.555**	0.520	**16.555**	0.520
WIZ-1(F)	17.571	**0.366**	17.401	**0.376**	16.701	**0.387**
WAN-1(I)	**21.970**	0.560	**21.970**	0.560	**21.970**	0.560
WAN-1(F)	23.233	**0.439**	22.970	**0.451**	22.567	**0.465**
MOR-1(I)	**1.258**	0.480	1.258	0.480	1.258	**0.480**
MOR-1(F)	1.266	**0.411**	**1.216**	**0.469**	**1.178**	1.548
TRE-1(I)	**1.339**	0.680	**1.339**	0.680	1.339	0.680
TRE-1(F)	1.463	**0.498**	1.351	**0.512**	**1.335**	**0.652**
NefProx Models	**Best $Inter_C$ MSE_{tst}**	**$Inter_C$**	**Median $Acc - Inter_C$ MSE_{tst}**	**$Inter_C$**	**Best Acc MSE_{tst}**	**$Inter_C$**
PLA-3(I)	**3.222**	0.400	**3.222**	0.400	**3.222**	**0.400**
PLA-3(F)	4.391	**0.287**	3.572	**0.344**	**3.222**	0.400
QUA-3(I)	**0.041**	0.400	0.041	0.400	0.041	0.400
QUA-3(F)	0.042	**0.303**	**0.041**	**0.327**	**0.041**	**0.363**
ELE-3(I)	**622331**	0.600	622331	0.600	622331	0.600
ELE-3(F)	700774	**0.474**	**620292**	**0.505**	**614691**	**0.551**
ABA-3(I)	7.231	0.920	7.231	0.920	7.231	0.920
ABA-3(F)	**6.377**	**0.642**	**6.271**	**0.680**	**6.205**	**0.725**
STP-3(I)	2.493	0.800	2.493	0.800	2.493	0.800
STP-3(F)	**2.339**	**0.528**	**2.164**	**0.557**	**2.055**	**0.588**
WIZ-3(I)	13.837	0.840	13.837	0.840	13.837	0.840
WIZ-3(F)	**13.271**	**0.630**	**12.998**	**0.644**	**12.804**	**0.667**
WAN-3(I)	**15.920**	1.000	**15.920**	1.000	**15.920**	1.000
WAN-3(F)	16.661	**0.808**	16.161	**0.827**	16.034	**0.852**
MOR-3(I)	**0.729**	0.800	0.729	0.800	0.729	0.800
MOR-3(F)	0.975	**0.475**	**0.660**	**0.529**	**0.632**	**0.615**
TRE-3(I)	**1.087**	0.800	1.087	0.800	1.087	0.800
TRE-3(F)	1.350	**0.482**	**0.978**	**0.543**	**0.942**	**0.608**

– Best $Inter_C$ models have improved the interpretability index (better interpretability so lower complexity) but the accuracy is a little more reduced. If the error values are observed (i.e. in the worst case the ECM_{tst} increased from 3.222 to 4.391) these remain low enough.
– Median $Acc - Inter_C$ models have similar precision of the base models and the interpretability index is increased, so their complexity have been reduced.
– Best Acc models improve both accuracy and interpretability index. All the individual indices MSE_{tst}, NR, S, R, I are reduced while NC is maintained.

5.2.2 Complex Models

This section presents the results obtained when complex base fuzzy models (named Models-2 for FasArt and Models-4 for NefProx) are involved.

The Table 9 shows the averaged results obtained for the Best $Inter_C$, Median $Acc - Inter_C$ and Best Acc models in the Pareto front for each data set: in fact the table shows the mean of the proposed index $Inter_C$ and the mean squared error for the test MSE_{tst}. The first line shows the base model (I), while the second line shows the performance of the final improved model(F).

Table 7 Wilcoxon test for the FasArt Compact Fuzzy Models: original model (R+) and improved model (R-)

Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
Best $Inter_C$				
MSE_{tst}	15.0	30.0	Accepted	0.374
NR	45.0	0.0	Rejected	0.008
S	45.0	0.0	Rejected	0.008
R	41.0	2.0	Rejected	0.028
I	40.0	5.0	Rejected	0.043
NC	5.0	40.0	Rejected	0.018
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	26.0	19.0	Accepted	0.678
NR	45.0	0.0	Rejected	0.008
S	45.0	0.0	Rejected	0.008
R	41.0	2.0	Rejected	0.028
I	40.0	5.0	Rejected	0.043
NC	2.0	41.0	Rejected	0.012
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	31.0	14.0	Accepted	0.314
NR	45.0	0.0	Rejected	0.008
S	40.0	5.0	Rejected	0.038
R	33.0	10.0	Accepted	0.249
I	29.0	16.0	Accepted	0.225
NC	1.5	43.5	Rejected	0.008
$Inter_C$	33.0	12.0	Accepted	0.214

The $Inter_C$ improvement is from 2.02% to 71.52%, while for MSE_{tst} is from -0.82% to $\% - 26.04$. In the worst case the interpretability increase from 0.560 to 1.123 (Tre-2, Best Acc model) and the accuracy increase from 0.552 to 0.725 (Tre-2, Best $Inter_C$ model).

In Appendix A, Table 13 shows the mean of the individual measures: the mean squared error for training (MSE_{tra}) and test (MSE_{tst}), the mean of rule number (RN), the mean similarity (S), the mean redundancy (R), the mean incoherency (I) and the mean percentage of completeness ($C(\%)$).

The Wilcoxon test for the FasArt Complex Models (Table 10) shows:

– Best $Inter_C$ models have improved the interpretability index with worse accuracy. Taking into consideration the high precision of the original models, this precision loss can be acceptable.
– Median $Acc - Inter_C$ models have improved the interpretability index while preserving the accuracy of the base models.
– Best Acc models are not statistically different, for neither the accuracy nor the interpretability index. Here the no variation of interpretability models is similar to the case commented previously.

Table 8 Wilcoxon test for the NefProx Compact Fuzzy Models: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	11.0	34.0	Accepted	0.173
NR	45.0	0.0	Rejected	0.008
S	44.0	1.0	Rejected	0.011
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	18.0	27.0	Accepted	0.655
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	35.0	10.0	Accepted	0.139
NR	45.0	0.0	Rejected	0.008
S	42.0	3.0	Rejected	0.021
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	18.0	27.0	Accepted	0.655
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	42.0	2.0	Rejected	0.017
NR	44.0	0.0	Rejected	0.012
S	38.0	6.0	Rejected	0.066
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	18.0	27.0	Accepted	0.655
$Inter_C$	44.0	0.0	Rejected	0.011

For NefProx Complex Models the results of Wilcoxon test (Table 11) shows:

– Best $Inter_C$ models, the accuracy is worse in some cases but this is balanced by the complexity reduction, so better interpretability.
– Median $Acc - Inter_C$ models have similar error with better interpretability (fewer rules, similarity, redundancy and incoherency with equal completeness).
– Best Acc models, the interpretability index is improved as previous models and the accuracy remains small.

5.2.3 Global Analysis

In general, the results obtained show a reasonable interpretability improvement of the (precise) fuzzy models obtained by well-known fuzzy algorithms whose main and original goal is to obtain models as accuracy as possible. In general, this interpretability improvement, that implies a complexity reduction, have been reached with an acceptable loss of accuracy and even, in some cases, the accuracy has been increased or preserved in similar ratios to the base models. In fact, a better interpretability has not to imply a loss of accuracy (see Tables 6 and 9). This better accuracy-interpretability trade-off based on complexity measures has been reached using a genetic approach focused on the final objectives.

Table 9 Performance of the Improved Complex Fuzzy Models

FasArt	Best $Inter_C$		Median $Acc - Inter_C$		Best Acc	
Models	MSE_{tst}	$Inter_C$	MSE_{tst}	$Inter_C$	MSE_{tst}	$Inter_C$
PLA-2(I)	**2.821**	0.760	2.821	0.760	2.821	0.760
PLA-2(F)	3.172	**0.216**	**2.567**	**0.248**	**2.374**	**0.590**
QUA-2(I)	0.050	1.000	0.050	1.000	0.050	1.000
QUA-2(F)	**0.041**	**0.492**	**0.038**	**0.521**	**0.038**	**0.581**
ELE-2(I)	100229	1.000	100229	1.000	100229	1.000
ELE-2(F)	**97416**	**0.740**	**87004**	**0.788**	**85157**	**0.869**
ABA-2(I)	6.247	0.880	6.247	0.880	6.247	0.880
ABA-2(F)	**5.385**	**0.431**	**5.273**	**0.470**	**5.240**	**0.535**
STP-2(I)	**0.698**	0.680	**0.698**	0.680	0.698	**0.680**
STP-2(F)	0.755	**0.353**	0.749	**0.549**	**0.686**	0.721
WIZ-2(I)	**21.934**	0.440	**21.934**	0.440	**21.934**	0.440
WIZ-2(F)	23.251	**0.360**	23.182	**0.365**	23.067	**0.377**
WAN-2(I)	**28.312**	0.680	**28.312**	0.680	**28.312**	0.680
WAN-2(F)	31.191	**0.510**	30.416	**0.516**	30.070	**0.580**
MOR-2(I)	**0.352**	0.680	**0.352**	0.680	**0.352**	0.680
MOR-2(F)	0.447	**0.391**	0.381	**0.415**	0.373	**0.642**
TRE-2(I)	**0.552**	0.560	0.552	0.560	0.552	**0.560**
TRE-2(F)	0.725	**0.354**	**0.548**	**0.379**	**0.534**	1.123
NefProx	Best $Inter_C$		Median $Acc - Inter_C$		Best Acc	
Models	MSE_{tst}	$Inter_C$	MSE_{tst}	$Inter_C$	MSE_{tst}	$Inter_C$
PLA-4(I)	**2.636**	0.400	**2.636**	0.400	**2.636**	0.400
PLA-4(F)	3.286	**0.296**	2.824	**0.335**	2.655	**0.392**
QUA-4(I)	**0.037**	0.400	0.037	0.400	0.037	0.400
QUA-4(F)	0.037	**0.285**	**0.037**	**0.302**	**0.037**	**0.331**
ELE-4(I)	**598472**	0.600	**598472**	0.600	598472	0.600
ELE-4(F)	612939	**0.502**	598735	**0.520**	**593548**	**0.543**
ABA-4(I)	6.370	1.000	6.370	1.000	6.370	1.000
ABA-4(F)	**6.128**	**0.782**	**6.093**	**0.799**	**6.019**	**0.827**
STP-4(I)	**1.727**	0.800	1.727	0.800	1.727	0.800
STP-4(F)	1.812	**0.560**	**1.693**	**0.580**	**1.664**	**0.611**
WIZ-4(I)	**17.471**	0.840	**17.471**	0.840	**17.471**	0.840
WIZ-4(F)	19.099	**0.603**	18.829	**0.624**	18.860	**0.653**
WAN-4(I)	**33.620**	0.880	**33.620**	0.880	**33.620**	0.880
WAN-4(F)	35.156	**0.658**	35.003	**0.681**	34.684	**0.744**
MOR-4(I)	0.510	0.880	0.510	0.880	0.510	0.880
MOR-4(F)	**0.415**	**0.620**	**0.382**	**0.647**	**0.377**	**0.681**
TRE-4(I)	0.673	0.800	0.673	0.800	0.673	0.800
TRE-4(F)	**0.611**	**0.528**	**0.584**	**0.551**	**0.572**	**0.585**

The performance evaluation permits the following analysis based on the Best Acc, Best $Inter_C$ and Median $Acc - Inter_C$ models from Pareto Fronts by the improvement:

– For the Best $Inter_C$ models the interpretability is improved (and complexity is reduced) but the deterioration of the precision is higher than in the other models (Median $Acc - Inter_C$ and Best Acc). Here, the model error is more significant, but taking into account the fact that the base models usually have a high level of precision, the improved models can remain accurate enough even with this precision loss.
– The Median $Acc - Inter_C$ models reduce their complexity, improving the metrics considered for the $Inter_C$ index, without a significative loss of accuracy. In fact, the interpretability improvement is achieved while preserving the accuracy level.
– The improvement for the Best Acc models is higher in the case of the NefProx compact models, but in NefProx complex models this improvement is also reached. This does not happen with FasArt models where both aspect are similar in base and improved models.

Table 10 Wilcoxon test for the FasArt Complex Fuzzy Models: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	14.0	31.0	Accepted	0.314
NR	45.0	0.0	Rejected	0.008
S	45.0	0.0	Rejected	0.008
R	45.0	0.0	Rejected	0.008
I	44.0	0.0	Rejected	0.012
NC	5.0	40.0	Rejected	0.028
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	31.0	14.0	Accepted	0.314
NR	45.0	0.0	Rejected	0.008
S	45.0	0.0	Rejected	0.008
R	42.0	3.0	Rejected	0.021
I	44.0	0.0	Rejected	0.012
NC	5.0	40.0	Rejected	0.028
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	33.0	12.0	Accepted	0.214
NR	45.0	0.0	Rejected	0.008
S	38.0	7.0	Rejected	0.066
R	26.0	19.0	Accepted	0.678
I	29.0	15.0	Accepted	0.401
NC	5.0	40.0	Rejected	0.028
$Inter_C$	34.0	11.0	Accepted	0.173

Table 11 Wilcoxon test for the NefProx Complex Fuzzy Models: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	17.0	28.0	Accepted	0.515
NR	45.0	0.0	Rejected	0.008
S	44.0	1.0	Rejected	0.011
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	6.0	36.0	Rejected	0.028
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	27.0	18.0	Accepted	0.594
NR	45.0	0.0	Rejected	0.008
S	44.0	1.0	Rejected	0.011
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	12.0	29.0	Rejected	0.080
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	33.0	12.0	Accepted	0.214
NR	45.0	0.0	Rejected	0.008
S	38.0	7.0	Rejected	0.066
R	41.0	2.0	Rejected	0.028
I	41.0	2.0	Rejected	0.028
NC	12.0	29.0	Rejected	0.080
$Inter_C$	45.0	0.0	Rejected	0.008

If the final models are compared, even though the different focus of both approaches, with the Wang & Mendel algorithm shown in (Gacto et al. 2010) again, it is possible to observe that the accuracy is better and the trade-off accuracy-interpretability has been highly improved for these precise models, but although a great reduction of the number of rules have been obtained, the number of rules in the improved FasArt and NefProx models remains higher.

On the other hand, there is a significant difference between FasArt and NefProx algorithms, when the interpretability of the fuzzy partitions is involved from the classical point of view. This aspect is guaranteed, although optimizable, in NefProx by the user, such as in this work. In FasArt that is not guaranteed, mainly in complex models, and some extra post-processing focused on the fuzzy sets, that are generated by clustering, is recommendable.

In the previous analysis the different influence of the individual complexity measures is remarkable, and this aspect must be taken into account by the user: i.e. Rule Number vs. Completeness: Must they have the same relevance?, this should be tuned by the user in accordance with his/her way of understanding the complexity and interpretability. In this work, this is influenced by the fuzzy modeling algorithms considered. Here the measures described previously have different relevancy for FasArt and NefProx.

6 Conclusions

This work introduces a simple methodology to improve fuzzy models obtained by well-known fuzzy modeling algorithms. The improvement is based on the complexity reduction, redundancy reduction, consistency, completeness, etc. of these models via rule selection. This is based on a bi-objective genetic approach guided by Accuracy and Interpretability measures to obtain a good trade-off between both contradictory performance.

A well-known set of useful measures about interpretability and accuracy concepts are implemented and adequately aggregated to check this proposal. They are based on the compactness, similarity, redundancy, consistency and completeness of the fuzzy sets and rules.

The checking of this proposal is carried out using nine case studies of the KEEL project, and two fuzzy modeling algorithms with good accuracy: FasArt and NefProx. Each one of these approaches generated two type of models, each one with a different performance in complexity and other aspects in order to test two different contexts.

The experimental results have shown a reasonable success: the complexity reduction, so a interpretability improvement, is reached for all models. This better interpretability has not to imply a mandatory loss of accuracy. When this happens the loss of accuracy has been moderate and acceptable, even in no few cases, the accuracy have been preserved even improved.

As further work, this approach will be applied to a real problem, as a biotechnological process focused on waste water treatment, with a double objective: to obtain good fuzzy models and knowledge through the interpretable rules of these models describing the complex relationships ruling these treatments.

Acknowledgements The authors would like to thank Francisco Herrera for his valuable and useful comments and support in the preparation of this manuscript. This work was supported by the Spanish Ministry of Science and Innovation under grants no. CIT-460000-2009-46 and DPI2009-14410-C02-02.

References

R. Alcalá, J. Alcalá-Fdez, J. Casillas, O. Cordón, and F. Herrera. Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling. *Soft Computing*, 10(9):717 – 734, 2006.

R. Alcalá, J. Alcalá-Fdez, F. Herrera, and J. Otero. Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation. *International Journal of Approximate Reasoning. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off.*, 44:45 – 64, 2007a.

R. Alcalá, M. J. Gacto, F. Herrera, and J. Alcalá-Fdez. A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5):539 – 557, 2007b.

R Alcalá, P. Ducange, F. Herrera, B Lazzerini, and F. Marcelloni. A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems. *IEEE Transactions on Fuzzy Systems*, 17(5):1106 – 1122, October 2009.

R. Alcalá, Y. Nojima, F. Herrera, and H. Ishibuchi. Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. *Soft Computing*, in press, 2011.

J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernndez, and F. Herrera. KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 13(3):307 – 318, Febraury 2009.

J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17:2-3:255–287, 2011.

J. M. Alonso and L. Magdalena. HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, Online First, June 2010.

J.M. Alonso, L. Magdalena, and G. González-Rodríguez. Looking for a good fuzzy system interpretability index: An experimental approach. *International Journal of Approximate Reasoning*, 51(1): 115 – 134, December 2009.

P.P. Bonissoene, Y.-T. Chen, K. Goebel, and P.S. Khedkar. Hybrid soft computing systems: industrial and commercial applications. *Proceedings of the IEEE*, 87(9):1641–1667, September 1999.

A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437 – 449, 2009.

J.M. Cano Izquierdo, Y.A. Dimitriadis, E. Gómez Sánchez, and J. López Coronado. Learnning from noisy information in FasArt and Fasback neuro-fuzzy systems. *Neural Networks*, 14(4-5):407–425, May 2001.

J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Accuracy Improvements in Linguistic Fuzzy Modelling*, volume 129 of *Studies in Fuzziness and SoftComputing*. Springer-Verlag, Berlin Heidelberg, 2003a.

J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Interpretability Issues in Fuzzy Modeling*, volume 128 of *Studies in Fuzziness and SoftComputing*. Springer-Verlag, Berlin Heildelberg, 2003b.

M.Y. Chen and D.A. Linkens. Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems*, 142 (2):243–265, March 2004.

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*, 11:1013 – 1031, 2007.

O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, volume 19 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, Singapore., 2001.

K. Cpalka. A new method for design and reduction of neuro-fuzzy classification systems. *IEEE Transactions in Neural Networks*, 20 (4):701 – 714, April 2009.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transacions on Evolutionary Computation*, 6(2):182–197, April 2002.

M.R. Delgado, F. Von Zuben, and F. Gomide. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter Hierarchical Genetic Fuzzy Systems: Accuracy, Interpretability and Design Autonomy, pages 379–405. Springer-Verlag, Berlin Heildelberg, 2003.

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

S. Destercke, S. Guillaume, and B. Charnomordic. Building an interpretable fuzzy rule base from data using orthogonal least squares-application to a depollution problem. *Fuzzy Sets and Systems*, 158 (18):2078 – 2094, September 2007.

J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591 – 600, October 2000.

A. Fiordaliso. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter About de trade-off between acuracy and interpretability of Takagi-Sugeno models in the context of nonlinear time series forecasting, pages 406–430. Springer-Verlag, Berlin Heildelberg, 2003.

M.J. Gacto and F. Alcalá, R.and Herrera. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Information Sciences*, 2011. In press.

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 - A Fusion of Foundations, Methodologies and Applications*, 13(5): 419 – 436, March 2009.

M.J. Gacto, R. Alcalá, and F. Herrera. Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 18(3):515 – 531, 2010.

M. Galende, G.I. Sainz, M.J. Fuente, and A. Herreros. Interpretability-accuracy improvement in a neuro-fuzzy ART based model of a DC motor. In *Proceedings of the 17th IFAC World Congress*, pages 7034 – 7039, Seoul, Korea, 6-11 July 2008.

M. Galende, G.I. Sainz, and M.J. Fuente. Accuracy-interpretability balancing in fuzzy models based on multiobjective genetic algortihm. In *Proceedings of European Control Conference 2009 (ECC'09)*, pages 3915 – 3920, Budapest, Hungary, 23 – 26 August 2009.

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

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

S. García, D. Molina, M. Lozano, and F. Herrera. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the CEC2005 special session on real parameter optimization. *J Heuristics*, 15:617–644, 2009b.

E. Gómez-Sánchez, Y.A. Dimitriadis, J.M. Cano-Izquierdo, and J. López-Coronado. μARTMAP: use of mutual information for category reduction in Fuzzy ARTMAP. *IEEE Transactions on Neural Networks*, 13(1):58 – 69, January 2002.

J. González, I. Rojas, H. Pomares, L.J. Herrera, A. Guillén, J.M. Palomares, and F. Rojas. Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms. *International Journal of Approximate Reasoning. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off.*, 44:32 – 44, 2007.

S. Guillaume and B. Charnomordic. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data, pages 148–175. Springer-Verlag, Berlin Heildelberg, 2003.

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

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. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off.*, 44(1):4 – 31, January 2007.

H. Ishibuchi and Y. Nojima. Discussions on interpretability of fuzzy systems using simple examples. *Proc. of 13th IFSA World Congress and 6th Conference of EUSFLAT*, pages 1649–1654, 2009.

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, January 2004.

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

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

H. Ishibuchi, T. Nakashima, and T. Murata. Three-objective genetics-based machine learning for linguistic rule extraction. *Information Sciences*, 136(1-4):109 – 133, August 2001.

H. Ishibuchi, Y. Kaisho, and Y. Nojima. Complexity, interpretability and explanation capability of fuzzy rule-based classifiers. In *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*, pages 1730 – 1735, 20-24 August 2009a.

H. Ishibuchi, Y. Nakashima, and Y. Nojima. Search ability of evolutionary multiobjective optimization algorithms for multiobjective fuzzy genetics-based machine learning. In *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*, pages 1724 – 1729, 20-24 August 2009b.

F. Jimenez, A. F. Gómez-Skarmeta, G. Sanchez, H. Roubos, and R. Babuška. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter Accurate, Transparent and Compact Fuzzy Models by Multi-Objective Evolutionary Algorithms, pages 431–451. Springer-Verlag, Berlin Heildelberg, 2003.

Y. Jin. Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement. *IEEE Transactions on Fuzzy Systems*, 8(2):212–221, April 2000.

Y. Jin, W. Von Seelen, and B. Sendhoff. On generating FC^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 29(6): 829–845, December 1999.

F.O. Karray and C. de De Silva. *Soft Computing and Intelligent Systems Design. Theory, Tools and Applications*. Addison Wesley, 2004.

A Konar. *Computational Intelligence: Principles, techniques and applications*. Springer-Verlag, Berlin, 2005.

C. Mencar and A. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585 – 4618, December 2008.

R. Mikut, J. Jäkel, and L. Gröll. Interpretability issues in data-based learning of fuzzy systems. *Fuzzy Sets and Systems*, 150(2):179 – 197, March 2005.

D. Nauck and R. Kruse. Neuro-fuzzy systems for function approximation. *Fuzzy Sets and Systems*, 101(2):261–271, January 1999.

Yusuke Nojima and Hisao Ishibuchi. Incorporation of user preference into multi-objective genetic fuzzy rule selection for pattern classifi cation problems. *Artificial Life and Robotics*, 14(3):418 – 421, 2009.

E. Parrado-Hernández, E. Gómez-Sánchez, and Y.A. Dimitriadis. Study of distributed learning as a solution to category proliferation in fuzzy artmap based neural systems. *Neural Networks*, 16 (7):1039 – 1057, September 2003.

P. Pulkkinen and H. Koivisto. Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. *International Journal of Approximate Reasoning*, 48(2):526 – 543, 2008.

P. Pulkkinen and H. Koivisto. A dynamically constrained multiobjective genetic fuzzy system for regression problems. *IEEE Transactions on Fuzzy Systems*, 18(1):161 – 177, 2010.

H. Roubos and M. Setnes. Compact and transparent fuzzy models and classfiers through iterative complexity reduction. *IEEE Transactions on Fuzzy Systems*, 9(4):516–524, August 2001.

G.I. Sainz, M.J. Fuente, and P. Vega. Recurrent neuro-fuzzy modelling of a wastewater treatment plant. *European Journal of Control*, 10: 83–95, 2004.

G.I. Sainz Palmero, Y.A. Dimitriadis, J.M. Cano Izquierdo, E. Gómez Sánchez, and E. Parrado Hernández. ART based model set for pattern recognition: FasArt family. In H. Bunke and A. Kandel, editors, *Neuro-fuzzy pattern recognition*, chapter 1, pages 147–177. World Scientific Pub. Co., December 2000.

G.I. Sainz Palmero, J. Juez Santamaria, E.J. Moya de la Torre, and J.R. Perán González. Fault detection and fuzzy rule extraction in AC motors by a neuro-fuzzy ART-based system. *Engineering Applications of Artificial Intelligence*, 18:867–874, 2005.

M. Setnes. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter Simplification and Reduction of Fuzzy Rules, pages 278–302. Springer-Verlag, Berlin Heildelberg, 2003.

M. Setnes and R. Babuška. Rule base reduction: Some comments on the use of orthogonal transforms. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 31(2): 199 – 206, May 2001.

M. Setnes, R. Babuška, U. Kaymak, and H.R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 28 (3):376 – 386, June 1998.

David J. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, 2003.

T. Suzuki and T. Furuhashi. *Interpretability Issues in Fuzzy Modelling*, volume 128 of *Studies in Fuzziness and SoftComputing*, chapter Conciseness of Fuzzy Models, pages 569–586. Springer-Verlag, Berlin Heildelberg, 2003.

L.-X. Wang and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6):1414–1427, November/December 1992.

J. Yen and L. Wang. Simplifying fuzzy rule-based models using orthogonal transformation methods. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 29(1):13–24, February 1999.

J.H. Zar. *Biostatistical Analysis*. Prentice Hall, 1999.

S.-M. Zhou and J.Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159:3091 – 3131, 2008.

X. Zong-Yi, J. Li-Min, Z. Yong, H. Wei-Li, and Q. Yong. A case study of data-driven interpretable fuzzy modeling. *Acta Automatica Sinica*, 31(6):815 – 824, November 2005.

X. Zong-Yi, Z. Yong, H. Yuan-Long, and C. Guo-Qiang. Multi-objective fuzzy modeling using NSGA-II. In *IEEE Conference on Cybernetics and Intelligent Systems*, pages 119 – 124, 21-24 September 2008.

A Tables of results

Table 12 Performance of the improved compact fuzzy models

FasArt	Best $Inter_C$						NefProx	Best $Inter_C$						
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-1(I)	3.483	**3.621**	48.60	0.2379	0.0011	0.0156	**99.90**	PLA-3(I)	3.208	**3.222**	17.00	0.2098	0	0	**100**
PLA-1(F)	3.742	3.718	**18.80**	**0.1314**	**0**	**0**	99.90	PLA-3(F)	4.350	4.391	**10.60**	**0.1701**	-	-	**100**
QUA-1(I)	0.050	0.054	119.80	0.2202	0.0001	0.0023	**98.03**	QUA-3(I)	0.039	**0.041**	55.80	**0.2723**	0	0	**100**
QUA-1(F)	0.039	**0.042**	**56.57**	**0.1692**	**0**	**0**	98.03	QUA-3(F)	0.041	0.042	**28.40**	0.2742	-	-	**100**
ELE-1(I)	117867	**158820**	92.60	0.2251	0.0043	0.0010	**96.34**	ELE-3(I)	620411	**622331**	79.60	0.2860	0	0	**99.44**
ELE-1(F)	124646	166133	**87.20**	**0.2178**	**0.0038**	**0.0005**	96.34	ELE-3(F)	659619	700774	**37.20**	**0.2575**	-	-	**99.44**
ABA-1(I)	6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61	ABA-3(I)	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25
ABA-1(F)	6.187	**6.881**	**67.80**	**0.2536**	**0**	**0**	**98.61**	ABA-3(F)	5.783	**6.377**	**153.40**	**0.4199**	**0.0071**	**0.0033**	**99.25**
STP-1(I)	2.091	**2.270**	101.80	0.1952	0	0.0002	**100**	STP-3(I)	2.248	2.493	303.40	0.2848	0.0174	0.0066	**100**
STP-1(F)	2.239	2.815	**64.60**	**0.1756**	-	**0.0001**	99.15	STP-3(F)	1.881	**2.339**	**174.47**	**0.2766**	**0.0126**	**0.0025**	**100**
WIZ-1(I)	5.452	**16.555**	221.60	0.3354	3e-5	0	**99.88**	WIZ-3(I)	9.958	13.837	500.00	0.5031	0.0237	0.0092	**99.67**
WIZ-1(F)	5.341	17.571	**115.97**	**0.3041**	**0**	-	99.67	WIZ-3(F)	8.988	**13.271**	**357.80**	**0.4847**	**0.0187**	**0.0045**	**99.67**
WAN-1(I)	9.813	**21.970**	231.80	0.3039	0	0	**99.92**	WAN-3(I)	12.227	**15.920**	500.00	0.4492	0.0174	0.0060	**98.33**
WAN-1(F)	9.686	23.233	**109.80**	**0.2794**	-	-	99.87	WAN-3(F)	11.627	16.661	**369.60**	**0.4346**	**0.0140**	**0.0032**	**98.33**
MOR-1(I)	1.041	**1.258**	52.60	0.2986	0	0	**99.93**	MOR-3(I)	0.716	**0.729**	170.00	0.3587	0.0460	0.0112	**100**
MOR-1(F)	1.007	1.266	**36.80**	**0.2847**	-	-	99.60	MOR-3(F)	0.890	0.975	**76.60**	**0.3510**	**0.0316**	**0.0029**	**100**
TRE-1(I)	0.908	**1.339**	49.60	0.2920	0.0005	0	**99.83**	TRE-3(I)	1.029	**1.087**	170.60	0.3498	0.0465	0.0106	**100**
TRE-1(F)	1.030	1.463	**36.00**	**0.2816**	**0**	-	99.74	TRE-3(F)	1.260	1.350	**76.80**	**0.3440**	**0.0293**	**0.0037**	99.55
FasArt	Median $Acc - Inter_C$							NefProx	Median $Acc - Inter_C$						
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-1(I)	3.483	3.621	48.60	0.2379	0.0011	0.0156	**99.90**	PLA-3(I)	3.208	**3.222**	17.00	0.2098	0	0	**100**
PLA-1(F)	2.818	**3.073**	**22.00**	**0.1535**	**0**	**0**	99.90	PLA-3(F)	3.559	3.572	**13.00**	**0.2008**	-	-	**100**
QUA-1(I)	0.050	0.054	119.80	0.2202	0.0001	0.0023	**98.03**	QUA-3(I)	0.039	0.041	55.80	**0.2723**	0	0	**100**
QUA-1(F)	0.035	**0.039**	**64.13**	**0.1760**	**0**	**0.0002**	98.02	QUA-3(F)	0.039	**0.041**	**34.00**	0.2790	-	-	**100**
ELE-1(I)	117867	**158820**	92.60	0.2251	0.0043	0.0010	**96.34**	ELE-3(I)	620411	622331	79.60	0.2860	0	0	**99.44**
ELE-1(F)	118905	159943	**87.60**	**0.2198**	**0.0042**	**0.0007**	96.34	ELE-3(F)	616480	**620292**	**47.20**	**0.2659**	-	-	**99.44**
ABA-1(I)	6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61	ABA-3(I)	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25
ABA-1(F)	5.185	**5.902**	**80.00**	**0.2614**	**0.0001**	**5e-5**	**98.61**	ABA-3(F)	5.627	**6.271**	**161.20**	**0.4222**	**0.0079**	**0.0039**	**99.25**
STP-1(I)	2.091	**2.270**	101.80	0.1952	0	0.0002	**100**	STP-3(I)	2.248	2.493	303.40	0.2848	0.0174	0.0066	**100**
STP-1(F)	2.100	2.745	**69.20**	**0.1802**	-	**0.0001**	98.97	STP-3(F)	1.671	**2.164**	**181.40**	**0.2770**	**0.0129**	**0.0032**	**100**
WIZ-1(I)	5.452	**16.555**	221.60	0.3354	3e-5	0	**99.88**	WIZ-3(I)	9.958	13.837	500.00	0.5031	0.0237	0.0092	**99.67**
WIZ-1(F)	5.096	17.401	**124.40**	**0.3083**	**0**	-	99.67	WIZ-3(F)	8.189	**12.998**	**363.00**	**0.4860**	**0.0188**	**0.0049**	**99.67**
WAN-1(I)	9.813	**21.970**	231.80	0.3039	0	0	**99.92**	WAN-3(I)	12.227	**15.920**	500.00	0.4492	0.0174	0.0060	**98.33**
WAN-1(F)	9.298	22.970	**119.20**	**0.2858**	-	-	99.87	WAN-3(F)	10.791	16.161	**376.00**	**0.4359**	**0.0146**	**0.0035**	**98.33**
MOR-1(I)	1.041	1.258	52.60	0.2986	0	0	**99.93**	MOR-3(I)	0.716	0.729	170.00	0.3587	0.0460	0.0112	**100**
MOR-1(F)	0.945	**1.216**	**38.80**	**0.2961**	-	-	98.91	MOR-3(F)	0.618	**0.660**	**82.80**	**0.3521**	**0.0321**	**0.0053**	**100**
TRE-1(I)	0.908	**1.339**	49.60	0.2920	0.0005	0	**99.83**	TRE-3(I)	1.029	1.087	170.60	0.3498	0.0465	0.0106	**100**
TRE-1(F)	0.917	1.351	**39.00**	**0.2833**	**0**	-	99.74	TRE-3(F)	0.848	**0.978**	**84.20**	**0.3447**	**0.0291**	**0.0065**	99.77
FasArt	Best Acc							NefProx	Best Acc						
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-1(I)	3.483	3.621	48.60	0.2379	0.0011	0.0156	**99.90**	PLA-3(I)	3.208	**3.222**	**17.00**	**0.2098**	0	0	**100**
PLA-1(F)	2.549	**2.688**	**27.20**	**0.2207**	**0.0006**	**0.0037**	98.63	PLA-3(F)	3.208	**3.222**	**17.00**	0.2098	-	-	**100**
QUA-1(I)	0.050	0.054	119.80	0.2202	0.0001	0.0023	**98.03**	QUA-3(I)	0.039	0.041	55.80	**0.2723**	0	0	**100**
QUA-1(F)	0.034	**0.038**	**79.90**	**0.1925**	**0**	**0.0004**	97.97	QUA-3(F)	0.038	**0.041**	**43.80**	0.2799	-	-	**100**
ELE-1(I)	117867	**158820**	92.60	0.2251	**0.0043**	**0.0010**	**96.34**	ELE-3(I)	620411	622331	79.60	0.2860	0	0	**99.44**
ELE-1(F)	116196	160842	**88.60**	**0.2208**	0.0045	0.0011	96.34	ELE-3(F)	612113	**614691**	**62.20**	**0.2791**	-	-	**99.44**
ABA-1(I)	6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61	ABA-3(I)	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25
ABA-1(F)	5.079	**5.788**	**91.20**	**0.2630**	**0.0002**	**0.0001**	**98.61**	ABA-3(F)	5.579	**6.205**	**174.40**	**0.4232**	**0.0089**	**0.0045**	**99.25**
STP-1(I)	2.091	**2.270**	101.80	**0.1952**	0	**0.0002**	**100**	STP-3(I)	2.248	2.493	303.40	0.2848	0.0174	0.0066	**100**
STP-1(F)	2.037	2.405	**78.20**	0.1965	-	0.0002	99.12	STP-3(F)	1.616	**2.055**	**191.00**	**0.2773**	**0.0137**	**0.0037**	**100**
WIZ-1(I)	5.452	**16.555**	221.60	0.3354	3e-5	0	**99.88**	WIZ-3(I)	9.958	13.837	500.00	0.5031	0.0237	0.0092	**99.67**
WIZ-1(F)	5.021	16.701	**134.53**	**0.3129**	**0**	-	99.81	WIZ-3(F)	8.004	**12.804**	**375.40**	**0.4880**	**0.0199**	**0.0053**	**99.67**
WAN-1(I)	9.813	**21.970**	231.80	0.3039	0	0	**99.92**	WAN-3(I)	12.227	**15.920**	500.00	0.4492	0.0174	0.0060	**98.33**
WAN-1(F)	9.151	22.567	**131.80**	**0.2903**	-	-	99.87	WAN-3(F)	10.654	16.034	**388.80**	**0.4380**	**0.0153**	**0.0037**	**98.33**
MOR-1(I)	1.041	1.258	52.60	**0.2986**	0	0	**99.93**	MOR-3(I)	0.716	0.729	170.00	0.3587	0.0460	0.0112	**100**
MOR-1(F)	0.924	**1.178**	**41.40**	0.3075	-	-	98.05	MOR-3(F)	0.582	**0.632**	**98.00**	**0.3556**	**0.0366**	**0.0079**	**100**
TRE-1(I)	0.908	1.339	49.60	0.2920	0.0005	0	**99.83**	TRE-3(I)	1.029	1.087	170.60	0.3498	0.0465	0.0106	**100**
TRE-1(F)	0.898	**1.335**	**41.40**	**0.2877**	**0.0005**	-	99.72	TRE-3(F)	0.816	**0.942**	**97.80**	**0.3446**	**0.0331**	**0.0081**	99.77

Table 13 Performance of the improved complex fuzzy models

FasArt			Best $Inter_C$					NefProx			Best $Inter_C$				
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-2(I)	2.783	**2.821**	96.60	0.1622	0.0004	0.0058	**100**	PLA-4(I)	2.606	**2.636**	31.00	0.1610	0	0	**100**
PLA-2(F)	3.206	3.172	**36.20**	**0.1147**	**0**	**0**	**100**	PLA-4(F)	3.284	3.286	**18.00**	**0.1423**	-	-	98.68
QUA-2(I)	0.046	0.050	243.80	0.2649	0.0002	0.0022	**98.13**	QUA-4(I)	0.035	**0.037**	98.00	0.2286	0	0	**100**
QUA-2(F)	0.036	**0.041**	**145.80**	**0.2240**	**0**	**3e-5**	98.12	QUA-4(F)	0.035	0.037	**46.00**	**0.2180**	-	-	**100**
ELE-2(I)	56584	100229	129.80	0.2661	0.0032	0.0010	**96.85**	ELE-4(I)	556228	**598472**	100.20	0.2266	0	0	**99.35**
ELE-2(F)	54089	**97416**	**111.80**	**0.2541**	**0.0006**	**0.0007**	96.84	ELE-4(F)	577220	612939	**59.40**	**0.2082**	-	-	**99.35**
ABA-2(I)	5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96	ABA-4(I)	5.636	6.370	500.00	0.3496	0.0062	0.0030	**95.52**
ABA-2(F)	4.276	**5.385**	**208.20**	**0.3048**	**3e-5**	**0.0001**	**99.96**	ABA-4(F)	5.247	**6.128**	**356.40**	**0.3355**	**0.0046**	**0.0015**	95.52
STP-2(I)	0.426	**0.698**	163.60	0.1854	3e-5	0.0002	**100**	STP-4(I)	1.307	**1.727**	433.80	0.2145	0.0093	0.0026	**100**
STP-2(F)	0.451	0.755	**133.60**	**0.1749**	**0**	**0**	99.45	STP-4(F)	1.213	1.812	**296.00**	**0.2098**	**0.0070**	**0.0010**	**100**
WIZ-2(I)	1.788	**21.934**	466.40	0.3603	0.6e-5	0	**100**	WIZ-4(I)	10.103	**17.471**	500.00	0.4418	0.0106	0.0042	**99.97**
WIZ-2(F)	1.743	23.251	**358.00**	**0.3472**	**0.2e-5**	-	**100**	WIZ-4(F)	9.819	19.099	**373.60**	**0.4257**	**0.0069**	**0.0019**	**99.97**
WAN-2(I)	2.593	**28.312**	537.60	0.3119	0.3e-5	0.1e-5	**99.93**	WAN-4(I)	21.836	**33.620**	500.00	0.3899	0.0080	0.0032	**98.50**
WAN-2(F)	2.505	31.191	**415.40**	**0.3053**	**0**	**0**	**99.93**	WAN-4(F)	21.239	35.156	**388.20**	**0.3763**	**0.0050**	**0.0017**	98.50
MOR-2(I)	0.085	**0.352**	92.20	0.2844	0.0006	0.0001	**99.99**	MOR-4(I)	0.337	0.510	301.60	0.2548	0.0260	0.0062	**99.88**
MOR-2(F)	0.097	0.447	**64.13**	**0.2759**	**0**	**4e-5**	99.82	MOR-4(F)	0.226	**0.415**	**171.60**	**0.2510**	**0.0217**	**0.0019**	99.70
TRE-2(I)	0.150	**0.552**	76.60	0.2893	0.0001	0.0001	**99.99**	TRE-4(I)	0.491	0.673	305.40	**0.2426**	0.0275	0.0044	**100**
TRE-2(F)	0.208	0.725	**46.73**	**0.2760**	**0**	**0**	99.46	TRE-4(F)	0.333	**0.611**	**171.40**	0.2434	**0.0220**	**0.0012**	99.69
FasArt			Median $Acc-Inter_C$					NefProx			Median $Acc-Inter_C$				
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-2(I)	2.783	2.821	96.60	0.1622	0.0004	0.0058	**100**	PLA-4(I)	2.606	**2.636**	31.00	0.1610	0	0	**100**
PLA-2(F)	2.409	**2.567**	**45.80**	**0.1240**	**0**	**0**	**100**	PLA-4(F)	2.733	2.824	**23.00**	**0.1499**	-	-	**100**
QUA-2(I)	0.046	0.050	243.80	0.2649	0.0002	0.0022	**98.13**	QUA-4(I)	0.035	0.037	98.00	0.2286	0	0	**100**
QUA-2(F)	0.033	**0.038**	**154.80**	**0.2315**	**0**	**0.0002**	98.12	QUA-4(F)	0.035	**0.037**	**52.20**	**0.2236**	-	-	**100**
ELE-2(I)	56584	100229	129.80	0.2661	0.0032	0.0010	**96.85**	ELE-4(I)	556228	**598472**	100.20	0.2266	0	0	**99.35**
ELE-2(F)	41899	**87004**	**113.60**	**0.2561**	**0.0012**	**0.0007**	96.84	ELE-4(F)	555512	598735	**67.60**	**0.2093**	-	-	**99.35**
ABA-2(I)	5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96	ABA-4(I)	5.636	6.370	500.00	0.3496	0.0062	0.0030	**95.52**
ABA-2(F)	4.087	**5.273**	**216.20**	**0.3075**	**5e-5**	**0.0002**	**99.96**	ABA-4(F)	5.151	**6.093**	**362.60**	**0.3362**	**0.0047**	**0.0016**	95.52
STP-2(I)	0.426	**0.698**	163.60	0.1854	**3e-5**	0.0002	**100**	STP-4(I)	1.307	1.727	433.80	0.2145	0.0093	0.0026	**100**
STP-2(F)	0.409	0.749	**147.40**	**0.1806**	3e-5	**0.0001**	99.74	STP-4(F)	1.090	**1.693**	**300.20**	**0.2100**	**0.0071**	**0.0012**	**100**
WIZ-2(I)	1.788	**21.934**	466.40	0.3603	0.6e-5	0	**100**	WIZ-4(I)	10.103	**17.471**	500.00	0.4418	0.0106	0.0042	**99.97**
WIZ-2(F)	1.655	23.182	**365.87**	**0.3488**	**0.2e-5**	-	**100**	WIZ-4(F)	9.497	18.829	**380.60**	**0.4270**	**0.0072**	**0.0021**	**99.97**
WAN-2(I)	2.593	**28.312**	537.60	0.3119	0.3e-5	0.1e-5	**99.93**	WAN-4(I)	21.836	**33.620**	500.00	0.3899	0.0080	0.0032	**98.50**
WAN-2(F)	2.374	30.416	**428.60**	**0.3062**	**0**	**0**	**99.93**	WAN-4(F)	20.772	35.003	**392.80**	**0.3779**	**0.0052**	**0.0019**	97.22
MOR-2(I)	0.085	**0.352**	92.20	0.2844	0.0006	0.0001	**99.99**	MOR-4(I)	0.337	0.510	301.60	0.2548	0.0260	0.0062	99.88
MOR-2(F)	0.081	0.381	**69.33**	**0.2806**	**0**	**0.0001**	99.80	MOR-4(F)	0.196	**0.382**	**179.00**	**0.2514**	**0.0217**	**0.0026**	**99.88**
TRE-2(I)	0.150	0.552	76.60	0.2893	0.0001	0.0001	**99.99**	TRE-4(I)	0.491	0.673	305.40	**0.2426**	0.0275	0.0044	**100**
TRE-2(F)	0.151	**0.548**	**54.93**	**0.2822**	**0**	**0**	99.58	TRE-4(F)	0.284	**0.584**	**176.60**	0.2428	**0.0217**	**0.0017**	99.69
FasArt			Best Acc					NefProx			Best Acc				
Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$	Models	MSE_{tra}	MSE_{tst}	RN	S	R	I	$C(\%)$
PLA-2(I)	2.783	2.821	96.60	0.1622	**0.0004**	0.0058	**100**	PLA-4(I)	2.606	**2.636**	31.00	**0.1610**	0	0	**100**
PLA-2(F)	2.249	**2.374**	**55.40**	**0.1452**	0.0007	**0.0013**	**100**	PLA-4(F)	2.606	2.655	**29.60**	0.1618	-	-	**100**
QUA-2(I)	0.046	0.050	243.80	0.2649	0.0002	0.0022	**98.13**	QUA-4(I)	0.035	0.037	98.00	**0.2286**	0	0	**100**
QUA-2(F)	0.033	**0.038**	**163.60**	**0.2355**	**3e-5**	**0.0004**	98.12	QUA-4(F)	0.035	**0.037**	**62.80**	0.2313	-	-	**100**
ELE-2(I)	56584	100229	129.80	0.2661	0.0032	**0.0010**	**96.85**	ELE-4(I)	556228	598472	100.20	0.2266	0	0	**99.35**
ELE-2(F)	39839	**85157**	**115.20**	**0.2540**	**0.0014**	0.0011	96.84	ELE-4(F)	550608	**593548**	**77.60**	**0.2127**	-	-	**99.35**
ABA-2(I)	5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96	ABA-4(I)	5.636	6.370	500.00	0.3496	0.0062	0.0030	**95.52**
ABA-2(F)	4.028	**5.240**	**223.00**	**0.3096**	**0.0001**	**0.0003**	**99.96**	ABA-4(F)	5.105	**6.019**	**374.80**	**0.3388**	**0.0051**	**0.0018**	95.52
STP-2(I)	0.426	0.698	163.60	**0.1854**	**3e-5**	**0.0002**	**100**	STP-4(I)	1.307	1.727	433.80	0.2145	0.0093	0.0026	**100**
STP-2(F)	0.396	**0.686**	**149.60**	0.1871	4e-5	0.0002	99.86	STP-4(F)	1.048	**1.664**	**312.20**	**0.2106**	**0.0073**	**0.0015**	**100**
WIZ-2(I)	1.788	**21.934**	466.40	0.3603	0.6e-5	0	**100**	WIZ-4(I)	10.103	**17.471**	500.00	0.4418	0.0106	0.0042	**99.97**
WIZ-2(F)	1.614	23.067	**380.53**	**0.3513**	**0.3e-5**	-	**100**	WIZ-4(F)	9.400	18.860	**392.80**	**0.4295**	**0.0078**	**0.0024**	**99.97**
WAN-2(I)	2.593	**28.312**	537.60	0.3119	0.3e-5	0.1e-5	**99.93**	WAN-4(I)	21.836	**33.620**	500.00	0.3899	0.0080	0.0032	**98.50**
WAN-2(F)	2.326	30.070	**447.00**	**0.3074**	**0.2e-5**	**0**	**99.93**	WAN-4(F)	20.639	34.684	**402.20**	**0.3798**	**0.0055**	**0.0020**	96.46
MOR-2(I)	0.085	**0.352**	92.20	**0.2844**	**0.0006**	0.0001	**99.99**	MOR-4(I)	0.337	0.510	301.60	0.2548	0.0260	0.0062	99.88
MOR-2(F)	0.074	0.373	**79.47**	0.2848	0.0007	**0.0001**	99.94	MOR-4(F)	0.191	**0.377**	**191.80**	**0.2528**	**0.0224**	**0.0032**	**99.88**
TRE-2(I)	0.150	0.552	76.60	**0.2893**	**0.0001**	**0.0001**	**99.99**	TRE-4(I)	0.491	0.673	305.40	**0.2426**	0.0275	0.0044	**100**
TRE-2(F)	0.138	**0.534**	**65.93**	0.2940	0.0001	0.0002	99.76	TRE-4(F)	0.275	**0.572**	**187.40**	0.2430	**0.0219**	**0.0023**	99.84