

## Bagged ensemble of Fuzzy C-Means classifiers for nuclear transient identification

Piero Baraldi, Roozbeh Razavi-Far, Enrico Zio

► **To cite this version:**

Piero Baraldi, Roozbeh Razavi-Far, Enrico Zio. Bagged ensemble of Fuzzy C-Means classifiers for nuclear transient identification. *Annals of Nuclear Energy*, Elsevier Masson, 2011, 38 (5), pp.1161-1171. <10.1016/j.anucene.2010.12.009>. <hal-00609529>

**HAL Id: hal-00609529**

**<https://hal-supelec.archives-ouvertes.fr/hal-00609529>**

Submitted on 27 Jul 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bagged Ensemble of Fuzzy C Means Classifiers for Nuclear Transient Identification

**Piero Baraldi, Roozbeh Razavi-Far, Enrico Zio**

*Department of Energy - Nuclear Engineering Division, Polytechnic of Milan, Via Ponzio 34/3, 20133 Milano, Italy*

**Abstract:** This paper presents an ensemble-based scheme for nuclear transient identification. The approach adopted to construct the ensemble of classifiers is bagging; the novelty consists in using supervised Fuzzy C-Means (FCM) classifiers as base classifiers of the ensemble. The performance of the proposed classification scheme has been verified by comparison with a single supervised, evolutionary-optimized FCM classifier with respect of the task of classifying artificial datasets. The results obtained indicate that in the cases of datasets of large or very small sizes and/or complex decision boundaries, the bagging ensembles can improve classification accuracy. Then, the approach has been applied to the identification of simulated transients in the feedwater system of a Boiling Water Reactor (BWR).

**Keywords:** Classification; Fuzzy C-Means (FCM) Clustering; Ensemble; Bagging; Transient Identification; BWR Nuclear Power Plant.

## 1. Introduction

The increasing degree of automation and growing demand for performance, efficiency, reliability and safety of industrial systems, is motivating an increasing interest in fault diagnosis. Indeed, to maintain a high level of performance and reliability, system errors, component faults and abnormal operation must be diagnosed promptly. Faults in process equipment can result in off-specification production, increased operating costs, the chance of line shutdown and the possibility of detrimental environmental impact. The source and severity of each malfunction must be diagnosed so that corrective actions can be taken quickly. Therefore, prompt detection and diagnosis of process malfunctions are strategically important for companies to remain competitive in world markets. Furthermore, large-scale systems such as nuclear power plants (NPPs) are required to provide safe and reliable operation for long periods of time; nevertheless, system components are subject to manufacturing defects, interactions with the environment, wear and tear, and other causes of performance degradation (Razavi-Far et al., 2009). For these safety-critical systems, the problem of detecting faults occurrence is of a high importance due to their disastrous consequences.

In this respect, a large number of diagnostic methods for nuclear power plant sensors and components have been proposed in the past decade (Reifman, 1997), based on the advances of soft computing techniques such as artificial neural networks (Kim and Bartlett, 1996; Roverso 2000; Embrechts and Benedek, 2004), fuzzy logic (Zio and Baraldi 2005; Zhao and Upadhyaya 2005) and neuro-fuzzy based techniques (Evsukoff and Gentil 2005; Zio and Gola 2006; Razavi-Far et al., 2009). These methods differ for their diagnostic scope and for the type of computing approach. With regards to the former, the main tasks that these diagnostic systems tackle are fault detection (i.e., acknowledging the existence of a plant anomaly, based on observed trends of measured signals) and identification (i.e., establishing the cause of the detected anomaly). With regards to the latter, classification techniques are widely used.

Real applications are usually characterized by complex decision boundaries between the different classes which are difficult to be learned by linear or circular classifiers or lie outside the space of functions that can be implemented by the chosen classifiers. In this respect, an evolutionary optimized fuzzy clustering algorithm (Zio and Baraldi, 2005) and local classification approaches such as Kernel methods (Müller and Mika, 2001), Radial Basis Networks (Leonard and Kramer, 1991), Fuzzy rules (Wang and Mendel, 1992) and Locally Linear Neuro-Fuzzy system (Razavi-Far et al., 2009) have been proved to provide good performances due to their inherent ability of splitting the input space in local zones. However, given the difficulty in providing a test dataset sufficiently representative of the future field data, these classification algorithms may perform differently in the field. In order to overtake this limitation of single classifier systems, a solution may be of combining the outputs of several classifiers in order to reduce the risk of an unfortunate selection of a poorly performing classifier. The ensemble may or may not beat the performance of the best classifier in the ensemble, but it certainly reduces the overall risk of making a particularly poor selection.

Other two instances in which the performance of a single classifier in real applications may not be satisfactory are (Polikar 2006): (a) a large volume of data is available for training; (b) a small dataset, inadequate for satisfactory training, is available. In case of type (a), training a single classifier with such a huge amount of data can be not practical; on the contrary partitioning the data into smaller subsets, training different classifiers with different partitions of data, and combining their outputs using an intelligent combination rule often demonstrates to be a more efficient approach. In case of type (b), resampling techniques have been successfully used for drawing overlapping random subsets of the available data, each of which is used to train a different classifier, creating an ensemble.

Furthermore, a classification scheme which follows a divide-and-conquer approach by dividing the dataset into smaller and easier-to-learn partitions, where each classifier learns only one of the simpler partitions allows to approximate the original complex decision boundary by a suitable combination of different classifiers. Figure 1, shows an example in which a complex decision boundary cannot be learned by a single linear or circular classifier (curve-shaped boundary), whereas ensemble of classifiers can span the decision space (Polikar, 2006).

Figure 1. Ensemble of classifiers spanning the decision space (Polikar, 2006)

In an ensemble of classifiers, each base classifier is trained on different distributions of data. One of the most popular approaches for constructing different training set for the base classifiers of the ensemble is bagging (Breiman, 1996). Bagging presents the classifiers with a training set that consists of training patterns randomly selected with replacement from the original dataset. Such a training set is called a bootstrap replicate of the original dataset. Empirical studies have established that bagging is a simple and robust method that generally increases the accuracy of a single learner (Breiman, 1996; Bauer and Kohavi, 1999). The generalization performance of ensemble classifiers is dependent on the diversity and accuracy trade-off of the base classifiers. The combination of the single base classifiers outputs can be implemented using a variety of strategies, such as combination of multiple classifiers (Woods et al., 1997), classifier fusion (Keller et al., 1994), committees of neural networks (Breiman, 1996; Schwenk and Bengio, 2000), decision tree ensembles (Banfield et al., 2007; Bauer and Kohavi, 1999) and several combinations of these schemes (Xu et al., 1992).

Whereas neural network, decision tree and support vector machine ensembles have been largely investigated as base classifiers of ensemble systems, there is no comprehensive empirical research in fuzzy clustering ensembles. To fill this void, this paper considers a supervised, evolutionary-optimized Fuzzy C-Means (FCM) clustering procedure (Zio and Baraldi, 2005) for building the base classifiers of the ensemble and proposes a bagging algorithm to generate the ensemble. The final classification decision is produced by using the majority-voting scheme (Parhami, 1994), one of the simplest and most common aggregation methods. There are three versions of majority voting (Polikar, 2006), where the ensemble chooses the class (i) on which all classifiers agree (*unanimous voting*); (ii) predicted by at least one more than half the number of classifiers (*simple majority*); or (iii) that receives the highest number of votes, whether or not the sum of those votes exceeds 50% (*plurality voting* or just *majority voting*). This latter is here used to determine the aggregated outcome of the base classifier decisions in the ensemble.

A comparison of the proposed bagging ensemble with a single supervised, evolutionary-optimized FCM classifier (Zio and Baraldi, 2005) is firstly conducted on Gaussian datasets with different sizes and complexities. The results are then confirmed through an application to the classification of simulated transients in the feedwater system of a Boiling Water Reactor (BWR) (Puska and Norman, 2002).

In order to accurately estimate the performance of the considered classification methods on test data not previously seen, a cross-validation (CV) procedure has been adopted (Efron, 1983; Efron and Tibshirani, 1995; Kohavi, 1995, Polikar, 2007). In particular, the so called “*K*-fold” cross-validation error estimate is used to compare the performance of the ensemble system with that of a single optimized classifier. A special case of the *K*-fold CV, called *leave-one-out* (LOO), is used in those cases in which the size of the available dataset is small.

The remainder of the paper is organized as follows. Section 2 presents a brief overview of the ensemble techniques for classification, outlining the methods and their value to transient identification; further, an ensemble of classifiers based on the supervised, evolutionary-optimized FCM clustering is proposed. In Section 3, a comparison of the bagging ensemble and the single supervised, evolutionary-optimized FCM classifier is presented, using Gaussian datasets with different sizes and complexities. Section 4 describes how the ensemble-based model of the supervised, evolutionary-optimized FCM classifier is used for transient identification in the feedwater system of a BWR, the case studies and the experimental results are compared in this Section. Finally, conclusions are drawn in Section 5.

## 2. Ensemble-Based Classification for Transient Identification

The strategy underlying ensemble-based classification is to create many classifiers and combine their outputs so that the combination improves upon the performance of a single classifier. This requires that individual classifiers perform well in different regions of the feature space and make errors on different patterns, which are balanced out in the combination. The intuition is that if each classifier makes different errors, then a strategic combination of these classifiers can reduce the total error. The overarching principle in ensembles is therefore to make each classifier as unique as possible, particularly with respect to misclassified instances. Specifically, we need classifiers whose decision boundaries are sufficiently different from those of others. Various techniques have been suggested for obtaining diversity in the base models of an ensemble, e.g. using different training parameters (Hansen and Salamon, 1990), different training patterns (Breiman 1996), different feature subsets (Zio et al., 2008) and different learning methods for each classifier of the ensemble (Xu et al., 1992).

In this work, the possibility of using different datasets to train individual classifiers is considered; such datasets are obtained through the resampling technique of bagging (Breiman, 1996), where training data subsets are drawn randomly, usually with

replacement, from the entire training dataset. Bagging, short for bootstrap aggregating, is one of the earliest ensemble-based algorithms (Breiman, 1996; Breiman, 1999). It is also one of the most intuitive and simplest to implement, with a surprisingly good performance. In bagging, different training data subsets are used to train different classifiers of the same type. The individual classifiers are then combined by taking a majority vote of their decisions: for any given instance, the class chosen by most classifiers is the ensemble decision. To ensure that there are adequate training samples in each subset, relatively large portions of the samples (75% to 100%) are drawn into each subset. This causes individual training subsets to overlap significantly, with many of the same instances appearing in most subsets, and some instances appearing multiple times in a given subset. In order to ensure diversity under this scenario, a relatively unstable model is used so that sufficiently different decision boundaries can be obtained for small perturbations in different training datasets.

The main structure of our ensemble scheme is shown in Figure 2. In the next Section 2.1, the supervised, evolutionary-optimized FCM clustering algorithm used to build the base classifiers of the ensemble is briefly described (Zio and Baraldi, 2005). The procedure of ensemble construction and its algorithm is presented in Section 2.2.

Figure 2. The main structure of the FCM-based ensemble scheme

## 2.1. The Supervised, Evolutionary- Optimized FCM Classifier

The supervised, evolutionary-optimized FCM classifier is here briefly introduced with reference to a problem of pattern classification in which: a set of  $N$  available patterns  $\vec{x}_k$ , of a priori known class  $\Gamma_k^t$ , is assumed available. The subscript  $t$  indicates that  $\Gamma_k^t$  represents the true, a priori known physical class of  $\vec{x}_k$ . The total number of classes of the  $N$  patterns  $\vec{x}_k$  is  $c$  and thus  $\Gamma_k^t$  assumes a value in  $(1, \dots, c)$ . The information regarding this known, physical class  $\Gamma_k^t$  of the  $k$ -th pattern is used to supervise an evolutionary algorithm for finding  $c$  optimal Mahalanobis metrics which define  $c$  geometric clusters as close as possible to the a priori known physical classes (Yuan and Klir, 1997). The Mahalanobis metrics are defined by the matrices  $M_i$ ,  $i=1, \dots, c$  whose elements are identified by the supervised, evolutionary-optimized algorithm so as to minimize the distances  $s_{ik} = (\vec{x}_k^i - \vec{v}_i^*)^T \underline{\underline{M_i}} (\vec{x}_k^i - \vec{v}_i^*)$  between the patterns  $\vec{x}_k^i$  belonging to class  $i$  and the class prototype  $\vec{v}_i^*$ , i.e. the cluster center. The overall iterative training scheme can be summarized as follows (Zio and Baraldi, 2005):

1. At the first iteration ( $\tau=1$ ), initialize the metrics of all the  $c$  clusters to the Euclidean metrics.
2. At the generic iteration step  $\tau$ , run the FCM clustering algorithm to partition the  $N$  training data into  $c$  clusters, based on the current metric  $M_i(\tau)$  and on the “supervising” initial partition  $\Gamma^t = (\Gamma_1^t, \dots, \Gamma_N^t)$ . Let  $\Gamma_k(\tau)$  be the cluster to which the  $k$ -th pattern  $\vec{x}_k$  is assigned. Then  $\Gamma(\tau) = (\Gamma_1(\tau), \dots, \Gamma_N(\tau))$  is set equal to the obtained optimal partition  $\Gamma^*$ .
3. Compute the distance  $D(\Gamma^t, \Gamma(\tau))$  between the a priori known physical classes and the geometric clusters. At the first iteration ( $\tau=1$ ) initialize the best distance  $D^+$  to  $D(\Gamma^t, \Gamma(1))$ ,  $D_i^+$  to  $D(\Gamma_i^t, \Gamma_i(1))$  and the best metric  $\underline{\underline{M_i^+}}$  to  $\underline{\underline{M_i(1)}}$  and go to step 5.
4. If  $\Gamma(\tau)$  is close to  $\Gamma^t$  or if the number of iterations  $\tau$  is greater than the predefined maximum allowed number of iterations  $\tau_{max}$ , stop:  $\Gamma(\tau)$  is the optimal cluster partition  $\Gamma^*$ ; otherwise, upgrade  $D^+$  to  $D(\Gamma^t, \Gamma(\tau))$ ,  $\underline{\underline{M_i^+}}$  to  $\underline{\underline{M_i(\tau)}}$  and  $D_i^+ = D(\Gamma_i^t, \Gamma_i(\tau))$ .

5. Increment  $\tau$  by 1. Update each matrix  $\underline{M_i^+}$  by exploiting its unique decomposition into Cholesky factors.

6. Return to step 2.

Once the classifier is constructed, a new test pattern  $\vec{x}$  is classified, in fuzzy terms, by computing its value of membership to the  $c$  clusters, based on the Mahalanobis distances of matrices  $\underline{M_i^+}$ . Given the order correspondence between classes and clusters, the fuzzy membership information is finally used for the crisp assignment of the pattern  $\vec{x}$  to the class with the largest value of membership.

## 2.2. The Construction of the Bagging Ensemble

All ensemble models are built on two key components: first, a strategy is needed to build diverse classifiers; second, a strategy, is needed to combine the outputs of the individual classifiers that make up the ensemble in such a way that the correct decisions are amplified, and the incorrect ones are cancelled out. In this work, diversity is obtained by bagging and derives from using bootstrapped replicas of the training data: different training data subsets are randomly drawn, with replacement, from the entire training dataset; at each bagging iteration, a new training subset is created based on the original dataset. Each training data subset is used to train a different classifier of the FCM type. As for the combination of the outcomes of the individual base classifiers of the ensemble, this is, achieved by using a majority voting technique.

Given the training dataset  $S$  of  $N$  patterns with known physical class labels  $\Gamma_k^t = 1, \dots, c$ , the bagging pseudo-code is as follows (see Figure 2):

Weak learning algorithm *WeakLearn*.

Integer  $T$ , number of iterations.

Integer  $F$ , fraction of the total number  $N$  of training patterns that constitute each bootstrapped replica.

### **Train: Bagging**

**Do**  $\alpha = 1, \dots, T$

1- Construct a set of bootstrap samples. Take a bootstrapped replica  $S_\alpha$  by randomly drawing  $F \times N$  patterns of the original dataset  $S$ , with replacement (Jiawei and Kamber, 2001).

2- Train a classifier  $h_\alpha$  on  $S_\alpha$ .

3- Add  $h_\alpha$  to the ensemble  $E$ .

**End**

### **Aggregate: Majority Voting**

Majority voting is one of the simplest and most intuitive methods to combine classification decisions. The majority voting method consists in assigning to  $\vec{x}$  the class label which is supported by the majority of individual classifiers. Given an unlabeled test pattern  $\vec{x}$ :

**Do**  $\alpha = 1, \dots, T$

1- Call classifier  $h_\alpha$  from the ensemble  $E$  and classify the test pattern  $\vec{x}$ .

2- Let  $\Gamma^\alpha$  be the class assigned by classifier  $h_\alpha$  to  $\vec{x}$ .

3- Let  $v_{\alpha,j} = \begin{cases} 1, & \text{if } \Gamma^\alpha = j \\ 0, & \text{otherwise} \end{cases}$  be the vote given to class  $j$  by classifier  $h_\alpha$ .

**END**

4- Obtain the total vote received by each class:

$$V_j = \sum_{\alpha=1}^T v_{\alpha,j}, \quad j = 1, \dots, c$$

5- Choose as final assignment the class that receives the largest total vote. In case the number of votes to different classes is equal, assign randomly one among these classes with largest total vote.

In order to avoid over/under estimation of the ensemble performance, cross-validation of the results is performed according to the  $K$ -fold cross-validation scheme proposed in (Polikar, 2007). The original dataset is randomly partitioned into  $K$  blocks of equal size. One of these blocks is used as test data subset and the remaining  $K - 1$  blocks are combined together to constitute the training data subset. The cross-validation process is then repeated  $K$  times (the  $K$ -folds) using a different block as test set each time and consequently the bagging algorithm is repeated  $K$  times, each time with a different training set. The general structure of the bagging algorithm within the cross-validation scheme is shown in Figure 3. The  $K$ -fold cross-validation error estimate is then the average of the  $K$  individual error estimates.

In case of a small-size dataset, the so called “leave-one-out” (LOO) cross-validation method is used. As the name suggests, leave-one-out cross-validation (LOOCV) involves using a single pattern from the original dataset as the test data, and the remaining  $N - 1$  patterns as the training data (Polikar, 2007). This is repeated  $N$  times so that each pattern of the dataset is used once as the test data. This is a special case of a  $K$ -fold cross-validation, with  $K$  being equal to  $N$ , the number of patterns in the original data. LOOCV is often computationally expensive because of the large number of times the training process is repeated, but it permits to obtain an unbiased estimation of the error (Polikar, 2007).

Figure 3. The general structure of the ensemble scheme, with cross-validation; E-fold = Ensemble-fold

### 3. Experimental Results on Artificial Datasets

In this Section, the bagging ensemble is compared to a supervised, evolutionary-optimized FCM classifier with respect to the classification of Gaussian and artificial datasets of different size and complexity, including challenging artificial (“2-Spirals”) and real-world datasets. A summary of the datasets characteristics is reported in Table 1.

Table 1: Description of the datasets used in the experiments.

The following datasets have been considered:

**3.1. Small Gaussian dataset:** A Gaussian dataset, formed by 4 classes and 30 patterns is used to train, test and compare the performance of the bagging ensemble and the supervised, evolutionary-optimized FCM classifier.

**3.2. Medium Gaussian datasets:** Two datasets constituted by 60 and 120 patterns, respectively, randomly sampled from the same Gaussian distributions of the Small Gaussian Dataset.

**3.3. Large Gaussian dataset:** A dataset constituted by 1500 patterns randomly sampled from the same Gaussian distributions of the Small Gaussian Dataset. This dataset combines the difficulty of dealing with a large number of patterns with the advantages of having only two features.

**3.4. Laryngeal Dataset:** It is a real medical dataset which contains 692 patterns of two classes representing normal/pathological voices. The Laryngeal dataset comes from the Bulgarian Academy of Sciences and is available at (Kuncheva, 2005). The set was originally used for a computer decision support system, in order to aid diagnosing laryngeal pathology and especially in detecting its early stages. The main difficulty in dealing with the dataset is given by the large number of features of the patterns (16).

**3.5. 2-Spirals Dataset:** it contains 200 bi-dimensional patterns of 2 classes as shown in Figure 4, exhibiting complex cluster shapes (Minaei-Bidgoli et al., 2004). This dataset has been considered due to its complex decision boundaries.

Figure 4: “2-Spirals” Dataset with two classes, 200 patterns (100-100 per class)

Table 2 reports the basic parameters used to build the ensemble of classifiers. The supervised, evolutionary-optimized FCM classifier used for the comparison has been trained with a larger number of iterations,  $\tau_{\max} = 10000$ , than the individual FCM classifiers of the ensemble, to have a fair comparison in terms of computational time.

Table 2: basic parameters used to build the ensemble of classifiers

Table 3 reports the mean and the standard deviation of the performances, defined as the ratio between the number of test patterns correctly classified and the total number of test patterns, of the bagging ensemble and optimized supervised FCM classifier, with a 10-fold cross-validation scheme.

Table 3: Mean and standard deviation of the performances i.e., the ratio between the number of test patterns correctly classified and the total number of test patterns in a 10-fold cross-validation

The bagging ensemble performs better than the optimized supervised FCM classifier in the case of the small Gaussian dataset considered. This is due to the fact that in absence of an adequate number of training patterns, the resampling method allows drawing overlapping random subsets, each of which is used to train a different classifier, creating a robust ensemble with more satisfactory performance. For the other datasets, the performances are similar although the bagging ensemble seems to perform slightly better than the optimized supervised FCM classifier, especially when the dimension and complexity of the dataset increase. In particular, in the cases of the Laryngeal and Spiral datasets, the bagging ensemble seems able to learn better the complexity of the decision boundaries that separate the data of the different classes.

#### 4. Application to Nuclear Transient Identification

In this Section, the ensemble-based fuzzy FCM and the supervised, evolutionary-optimized FCM classifiers are applied to the classification of transients in the feedwater system of a BWR. The diagnostic problem of interest regards the early classification of a predefined set of transients in a BWR. The corresponding transients have been simulated by the HAMBO simulator of the Forsmark 3 BWR plant in Sweden (Puska and Noemann, 2002).

#### 4.1. Process Description and Transient Simulation

The considered faults occur in the section of the feedwater system where the feedwater is preheated from 169°C to 214°C in two parallel lines of high-pressure preheaters while going from the feedwater tank to the reactor. Figure 5 shows a sketch of the system. A set of faults, that are generally hard to detect for an operator and that produce efficiency losses if undetected are chosen for this application (see (Roverso 2004) for their description). The proposed faults are as follows:

1. F1-F5 regard line 1 of the feedwater system.
2. F6, F7, F8 regard both lines.

Among the 363 measured signals, only the 20 reported in Table 4 have been used for the fault classification in the two case studies here considered. These signals have been chosen considering the results of the application of a feature selection algorithm (Zio et al., 2006) and some benchmark tests.

Figure 5: A sketch of the feedwater system of the BWR

Table 4: Selected signals for the fault classification.

Five transients were simulated for each of the 8 faults, considering different degrees of leakage and valve closures and with step and ramp changes at different leak sizes. The data relative to the selected 20 signals were recorded with a sampling frequency of 1 Hz. All transients start after 60 seconds of steady state operation. Given that the ramp changes cause variations of the parameters later than the step changes, only the three step changes for each fault were considered.

#### 4.2 Transient identification in the case of a small dataset

In this case study, the two faults (F6 and F8) have been considered. The dataset is obtained by taking for each fault a transient caused by a step change of large magnitude. A pattern every 10 seconds has been considered in the interval [100,150] seconds. Thus, the dataset contains 12 patterns and 20 signals. The dataset is classified by the bagging ensemble and the supervised, evolutionary-optimized FCM classifiers in a LOOCV scheme. Figure 6 shows the results of the classification of each single test pattern when the classifier is trained with the remaining 11 patterns, i.e. a single cross-validation in the LOOCV scheme. The bagging ensemble correctly classifies all the patterns (Performance = 1) whereas the supervised, evolutionary-optimized FCM classifier misclassifies pattern 5 (Performance = 0.917).

Figure 6: Performance comparison on each single test pattern – Small transient dataset (OFCM = Optimized FCM classifier)

Figure 7 reports the values of membership to classes F6 and F8 assigned to each of the 12 patterns by the ensemble base classifiers and by the supervised, evolutionary-optimized FCM classifier. Notice that all the classifiers tend to assign to pattern 5 membership values close to 0.5, indicating uncertainty in the assignment; however, the supervised, evolutionary-optimized FCM classifier assigns a slightly higher membership to class F8 and thus the pattern is assigned to class F8, whereas its true class is F6.

Figure 7: Values of membership to each class assigned to the 12 test patterns by the individual base classifiers of the bagging ensemble (●, ◆) and the supervised, evolutionary-optimized FCM classifier (OFCM —, ····).

On the other hand, 6 over 10 classifiers of the bagging ensemble assign pattern 5 with a higher membership degree to class F6, leading to a correct classification of the pattern. In this case, the diversity of the individual base classifiers obtained by using different training sets and aggregated by the majority voting rule allows the ensemble to correctly classify the test pattern.

#### 4.3 Transient identification in the case of a large datasets

The objective of this case study is a prompt and correct classification of the seven faults  $F_1, \dots, F_7$ . For each fault, three transients corresponding to step changes have been considered since they lead to an early variation of the measured variables. Given that the goal is early fault diagnosis, only the data of the first 220 seconds after the beginning of the transients have been considered. The obtained dataset contains 4641 patterns of 7 different classes with 20 features. Figure 8 shows the values of the 20 features in the patterns of the 7 classes.

Figure 8: Behavior of the signals in the different fault classes ( $F_1 - F_7$ ).

Table 5 reports the mean performance obtained by the bagging ensemble and by the supervised, evolutionary-optimized FCM classifier in a 10-fold cross-validation scheme. The results indicate higher performance and robustness of the bagging ensemble in the classification of this high-dimensional dataset.

Table 5: Mean and standard deviation of the performance

As seen, the performance of the bagging classifier is slightly higher than that of the supervised, evolutionary-optimized FCM classifier when the large dataset, with high dimensionality, is used.

In order to investigate the performance of the bagging classifier and its diversity, the results achieved testing a single fold during the cross-validation process are analyzed. Figure 9 shows the number of test patterns which have been correctly classified by 0, 1, 2, ..., 10 individual base classifiers of the ensemble. Notice that all the ten individual base classifiers of the ensemble correctly classify 389 out of the 464 test patterns, indicating an overall satisfactory performance of the individual base classifiers. The remaining 75 patterns are assigned to the class which is supported by the majority of individual classifiers. Thus, if more than 5 of the 10 classifiers correctly classify a pattern, it is correctly classified by the ensemble. The remaining cases are analyzed in Figures 10 and 11 which show the distribution of the number of test patterns correctly classified and misclassified, respectively, by the ensemble. For example, Figure 10 shows that there are 3 patterns which are correctly classified by the ensemble although only 4 out of the 10 individual classifiers correctly classify them. This is due to the fact that the 6 classifiers misclassify the patterns assigning them to different classes, i.e, the diversity of the incorrect votes lead to a correct overall classification of the ensemble. Reversely, notice in Figure 11 that there are 2 cases in which all the 6 classifiers misclassifying the pattern agree in assigning them to the same incorrect class and thus the resulting classification of the pattern is incorrect.

**These analyses show the importance of obtaining an ensemble formed by single classifiers which make different errors. The diversity of the classification errors permits to correctly classify a pattern, even in those cases of Figure 10 in which the correct class is identified by a minority of the ensemble classifiers.**

Figure 9: Number of classifier responses coherent with true classes

Figure 10. Analysis of the cases in which the ensemble final decision is correct although only a minority of classifiers correctly classify the patterns. The numbers [X,Y1,Y2,Y3] next to the column bars indicate the number of individual base classifiers with correct vote [X] and with incorrect votes to different classes [Y1,Y2,Y3].

Figure 11. Analysis of the cases in which the ensemble final decision is wrong. The numbers [X,Y1,Y2,Y3] next to the column bars indicate the number of individual base classifiers with correct vote [X] and with incorrect votes to different classes [Y1,Y2,Y3].

*In order to investigate the time needed by the diagnostic algorithm to perform a correct diagnosis of the transient, the developed model has been tested by considering patterns taken every second from  $t=0s$  to  $t=XXX$ . ...*

*Figures and tables*

## 5. Conclusions

The accurate and timely identification of transients in safety critical nuclear systems is important because it can decrease the probability of catastrophic failures, increase the life of the plant and reduce maintenance costs. In this paper, an ensemble-based scheme has been presented for nuclear transient identification. An ensemble is a group of learners which are combined together in order to obtain better generalization performance than that obtained by a single learner. The errors of the individual ensemble members cancel out to some degree when their predictions are combined. One of the most effective techniques for ensemble construction is the bagging method. This approach promises to give superior classification results with improved accuracy. The expectation is that each classifier will make a different error, and strategically combining these classifiers can reduce the total error. This is because the sets of data examples misclassified by the different individual classifiers would not necessarily overlap. Thus, different classifier designs potentially offer complementary information about the data examples to be classified which could improve the performance of the selected classifiers.

In this work, bagging is used to construct the ensemble of classifiers; the novelty is in the use of a supervised Fuzzy C-Means (FCM) classifier as individual base classifier of the ensemble.

**A comparative experimental evaluation of the bagging ensemble with a supervised, evolutionary-optimized FCM classifier has been performed with reference to the classification of artificial Gaussian datasets and simulated transients in the feedwater system of a Boiling Water Reactor.**

**The experimental results show that the bagging ensemble performs as well or slightly better than the optimized FCM classifier on medium and large datasets and outperforms the optimized FCM classifier in the classification of small datasets. This is due to the fact that the single classifiers tend to overfit the data whereas the diversity of the ensemble base classifiers reduces the risk of overfitting.**

**Notice that some issues have still to be addressed for the real application of automatic diagnostics models such as the one here proposed to operating NPPs. First of all, diagnostic systems based on empirical models have performance which may vary in different zones of the training space, according to the density and information content of the example patterns available to train the model. Then, for their application in practice, it is important to a-priori verify whether the model is working in a zone where its performance is expected to be satisfactory. Another factor that has limited the practical application of automatic diagnostic algorithms is the difficulty of obtaining the recognition of transients and faults at different plant operational conditions. Finally, diagnostic approaches able to learn novel information becoming available during the plant life without forgetting the previously acquired knowledge should be investigated in order to reduce the maintenance efforts and costs of the diagnostic system.**

**In this respect, future developments of the present work include the investigation of ensemble techniques for estimating the degree of confidence in the fault diagnosis, for automatically taking into account changes in the plant operating state and updating the diagnostic model with new information.**

**Acknowledgments:** The authors wish to thank Drs. Paolo Fantoni and Davide Roverso of the IFE, Halden Reactor Project for providing the transient simulation data.

#### **References:**

- Banfield, R.E., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P., 2007. A comparison of decision tree ensemble creation techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 (1), pp. 173–180.
- Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting and variants. *Machine Learning*, 36, pp. 105–139.
- Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24 (2), pp. 123–140.
- Breiman, L., 1999. Combining predictors, in *Combining Artificial Neural Nets*. A. J. C. Sharkey (ed.), Springer.
- Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78 (382), pp. 316-331.
- Efron, B., Tibshirani, R.J., 1995. Improvements on cross-validation: The .632+ bootstrap method. *Journal of the American Statistical Association*, 92 (43), pp. 548-560.
- Embrechts, M.J., Benedek, S., 2004. Hybrid identification of nuclear power plant transients with artificial neural networks. *IEEE Transactions on Industrial Electronics*, 51, pp. 686–693.
- Evsukoff, A., Gentil, S., 2005. Recurrent neuro-fuzzy system for fault detection and isolation in nuclear reactors, *Advanced Engineering Informatics*, 19 (1), pp. 55-66.
- Hansen, L.K., Salamon, P., 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (10), pp. 993–1001.
- Jiawei, H., Kamber, M., 2001. *Data Mining-Concepts and Techniques*, Morgan Kaufmann Publishers.
- Keller, J.M., Gader, P., Tahani, H., Chiang, J.H., Mohamed, M., 1994. Advances in fuzzy integration for pattern recognition. *Fuzzy Sets and Systems*, 65, pp. 273–283.
- Kim, K., Bartlett, E.B., 1996. Nuclear power plant fault diagnosis using neural networks with error estimation. *IEEE Transactions on Nuclear Science*, 43 (4), pp. 2373-2388.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. San Mateo, CA: Morgan Kaufman. pp. 1137–1143.

Kuncheva, L.I., 2005. Real medical datasets. Technical report, School of Informatics, University of Wales, Bangor, UK. [http://www.bangor.ac.uk/~mas00a/activities/real\\_data.htm](http://www.bangor.ac.uk/~mas00a/activities/real_data.htm).

Leonard, J.A., Kramer, M.A., 1991. Radial basis function networks for classifying process faults. *IEEE Control Systems Magazine*, 11 (3), pp. 31-38.

Minaei-Bidgoli, B., Topchy, A., Punch, W.F., 2004. Ensembles of Partitions via Data Resampling. International conference on information technology. Las Vega, Nevada, USA, Vol 2, pp. 188-192.

Müller, K.R., Mika, S., 2001. An introduction to kernel based learning algorithms. *IEEE Transactions on Neural Networks*, 12 (2), pp. 181-201.

Parhami, B., 1994. Voting algorithms. *IEEE Transactions on Reliability*, 43 (4), pp. 617–629.

Polikar, R., 2006. Ensemble Based Systems in Decision Making. *IEEE Circuits and Systems Magazine*, 6 (3), pp. 21-45.

Polikar, R., 2007. Bootstrap-Inspired Techniques in Computational Intelligence. *IEEE signal processing magazine*, 59, pp. 59-72.

Puska, E., Normann, S., 2002. 3-d core studies for hambo simulator. In proceedings of presentations on man-machine system research, Enlarged Halden programme group meeting, vol. 2.

Razavi-Far, R., Davilu, H., Palade, V., Lucas, C., 2009. Model based fault detection and isolation of a steam generator using neuro-fuzzy networks. *Neurocomputing Journal*, 72, pp. 2939–2951.

Rovero, D., 2000. Soft computing tools for transient classification. *Information Sciences*, 127, pp. 137-156.

Rovero, D., 2004. On-line early fault detection and diagnosis with the alladin transient classifier. Proceedings of PNPIC and HMIT-2004, the 4<sup>th</sup> American Nuclear Society, International Topical Meeting on Nuclear Plant Instrumentation Control and Human-Machine Interface Technologies, Columbus, Ohio, September 19-22.

Reifman, J., 1997. Survey of artificial intelligence methods for detection and identification of component faults in nuclear power plants. *Nuclear Technology* 119, pp. 76–97.

Schwenk, H., Bengio, Y., 2000. Boosting neural network. *Neural Computation*, 12, pp. 1869–1887.

Wang, L.X., Mendel, M., 1992. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems Man Cybernet.* 22 (6), 1414–1427.

Woods, K., Kegelmeyer, W.P., Bowyer, K., 1997. Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19, pp. 405–410.

Xu, L., Krzyzak, A., Suen, C.Y., 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Syst., Man, and Cybern.*, 22, pp. 418–435.

Yuan, B., Klir, G., 1997. *Intelligent Hybrid Systems Fuzzy Logic, Neural Network, and Genetic Algorithms*, Kluwer Academic Publishers.

Zhao, K., Upadhyaya, B.R., 2005. Adaptive fuzzy inference causal graph approach to fault detection and isolation of field devices in nuclear power plants. *Progress in Nuclear Energy*, 46 (3-4), pp. 226-240.

Zio, E., Baraldi, P., Pedroni, N., 2006. Selecting Features for Nuclear Transients Classification by Means of Genetic Algorithms. *IEEE transactions on nuclear science*, 53 (3), pp. 1479-1493.

Zio, E., Baraldi, P., Gola, G., 2008. Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery. *Applied Soft Computing*, 8, pp. 1365–1380.

Zio, E., Baraldi, P., 2005. Identification of nuclear transients via optimized fuzzy clustering. *Annals of Nuclear Energy*, 32, pp. 1068-1080.

Zio, E., Gola, G., Neuro-fuzzy pattern classification for fault diagnosis in nuclear components. *Annals of Nuclear Energy*, 33, pp. 415–426.