HAL
open science

A METHOD FOR ESTIMATING THE CONFIDENCE IN THE IDENTIFICATION OF NUCLEAR TRANSIENTS BY A BAGGED ENSEMBLE OF FCM CLASSIFIERS

Piero Baraldi, Roozbeh Razavi-Far, Enrico Zio

To cite this version:

Piero Baraldi, Roozbeh Razavi-Far, Enrico Zio. A METHOD FOR ESTIMATING THE CONFIDENCE IN THE IDENTIFICATION OF NUCLEAR TRANSIENTS BY A BAGGED ENSEMBLE OF FCM CLASSIFIERS. Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies NPIC&HMIT, 2010, pp.283 - 293. hal-00721035

HAL Id: hal-00721035

https://hal.science/hal-00721035v1

Submitted on 26 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.

A METHOD FOR ESTIMATING THE CONFIDENCE IN THE IDENTIFICATION OF NUCLEAR TRANSIENTS BY A BAGGED ENSEMBLE OF FCM CLASSIFIERS

Piero Baraldi, Roozbeh Razavi-Far
Dipartimento di Energia - Sezione Ingegneria Nucleare
Politecnico di Milano
Via Ponzio 34/3, 20133 Milano, Italy
piero.baraldi@polimi.it, razavi@nuclearmi.cesnef.polimi.it

Enrico Zio
Dipartimento di Energia - Sezione Ingegneria Nucleare
Politecnico di Milano
Via Ponzio 34/3, 20133 Milano, Italy
and
Ecole Centrale Paris-Supelec, Paris, France
enrico.zio@polimi.it

ABSTRACT

The performance of diagnostic systems based on empirical models may vary in different zones of the training space. It is, thus, important to a-priori verify whether the model is working in a zone where the performance is expected to be satisfactory. In this respect, the objective of this work is to estimate the degree of confidence in the identification of nuclear transients by a diagnostic system based on a bagged ensemble of Supervised Fuzzy C-Means (FCM) classifiers. The method has been applied for classifying simulated transients in the feedwater system of a nuclear Boiling Water Reactor (BWR). The obtained results indicate that the bagging ensemble permits to achieve satisfactory performance, with a reliable estimation of the degree of confidence in the classification.

Key Words: Confidence Estimation; Bagging Ensemble; Fuzzy C-Means Classifiers; Transient Identification; Nuclear Power Plant.

1 INTRODUCTION:

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 [1]. For these safety-critical systems, fault detection and identification are of paramount importance due to their disastrous consequences. In this respect, a large number of diagnostic methods have been proposed in the past decade for application to NPP sensors and components, based on the advances of soft computing techniques such as artificial neural networks [2, 3, 4], fuzzy logic [5, 6] and neuro-fuzzy based techniques [7, 8, 1]. Furthermore, in recent years, there has been an increasing interest in ensemble-based classifiers for fault

diagnosis due to their superior performance. Indeed, ensemble-based classifiers have been shown to successfully generate strong classifiers from weak classifiers [9].

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 [10].

The objective of the present work is to provide a method for estimating the confidence in the classification of any pattern. The diagnostic system considered is an ensemble of classifiers for nuclear transient identification, which has been proposed by the authors in a previous work [11]. Bagging [12] is used to construct an ensemble of supervised Fuzzy C-Means (FCM) classifiers [5]. The final classification decision is obtained by majority-voting of the outcomes of the individual classifiers [13].

The confidence in the final classification is obtained by exploiting the nature of the ensemble itself, to compute the Bayesian posterior probability of the class by a softmax approach [14]. The ensemble-based classification approach with confidence estimation has been applied for the identification of simulated transients in the feedwater system of a Boiling Water Reactor (BWR) [15]. The remainder of the paper is organized as follows. Section (2) presents a brief overview of the ensemble-based classification approach. In Section (3), the technique for confidence estimation is illustrated. Section (4) describes the results obtained by applying the classification approach with confidence estimation to the transients in the feedwater system of a BWR. Finally, conclusions are drawn in Section (5).

2 ENSEMBLE-BASED CLASSIFICATION FOR TRANSIENT IDENTIFICATION

Ensemble-based classification techniques base the class assignment on the combination of the outputs of a set of individual classifier models. The individual classifiers perform well in different regions of the feature space and thus make errors on different patterns; these are balanced out in the combination so that the performance is superior than that of a single classifier. In this work, different datasets are used to train the individual classifiers; such datasets are obtained through the resampling technique of bagging. Bagging, short for bootstrap aggregating, is one of the earliest ensemble-based algorithms [12]. Training data subsets are drawn randomly, usually with replacement, from the entire training dataset. The output classes of the individual classifiers are combined by majority voting: the class chosen by most classifiers is the ensemble assignment. 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 the ensemble scheme is shown in Fig. 1. In the next Section 2.1, the supervised, evolutionary-optimized FCM algorithm [5] used to build the base classifiers of the ensemble is

briefly described. The procedure of ensemble construction and its algorithm is presented in Section 2.2.

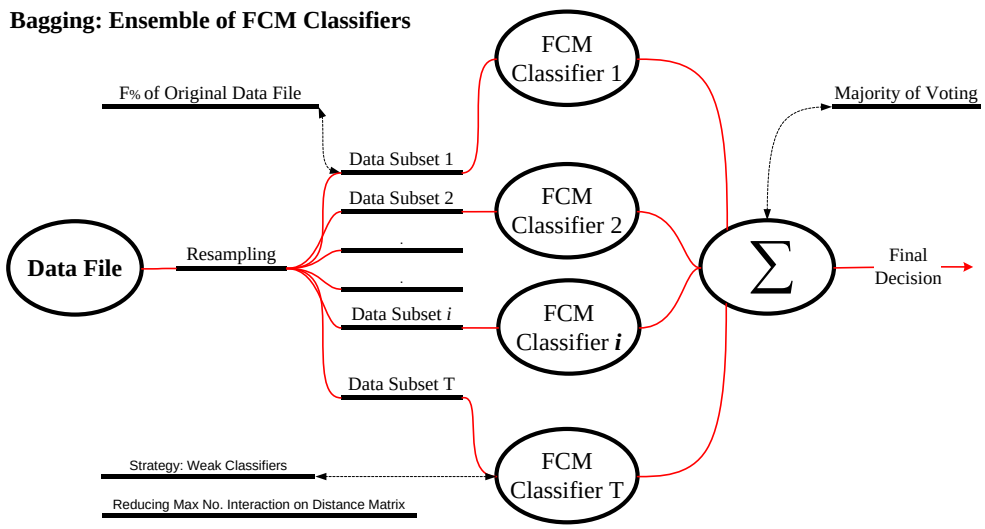

Figure 1. The main structure of the FCM ensemble scheme

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 $\vec{x}_k$, whose corresponding class Γ_k^t is a priori known, is assumed available. The subscript t indicates that Γ_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 Γ_k^t assumes a value in $(1,\ldots,c)$. The information regarding the known, physical class Γ_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 [16]. The Mahalanobis metrics are defined by the matrices $M_i, i=1,\ldots,c$ whose elements are identified by the supervised evolutionary algorithm so as to minimize the distances $s_{ik}=\left(\vec{x}_k^i-\vec{v}_i^*\right)^T \underline{\underline{M_i}}\left(\vec{x}_k^i-\vec{v}_i^*\right)$ between the patterns $\vec{x}_k^i$ belonging to class i and the class prototype, i.e. the cluster center $\vec{v}_i^*$.

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. 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

Ensemble models are built on two key components: the strategy for building classifiers that are as diverse as possible; the strategy for combining 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 data; 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 FCM classifier. As for the combination of the outcomes of the individual base classifiers of the ensemble, this is, achieved by majority voting.

To evaluate the ensemble error, cross-validation is performed according to the *K-fold* cross-validation scheme proposed in [17]. 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 *K-fold* cross-validation error estimate is then the average of the *K* individual error estimates. The general structure of the bagging algorithm within the cross-validation scheme is shown in Fig. 2.

Basic parameters used to build the ensemble of classifiers are: T "Number of base classifiers in the ensemble" = 10 in the application; F "fraction of the total number N of training patterns which constitute each bootstrapped replica" = 0.75 in the application; $\tau_{\max}$ "Number of iterations of the supervised algorithm used to train the single base classifiers of the ensemble" = 500 in the application.

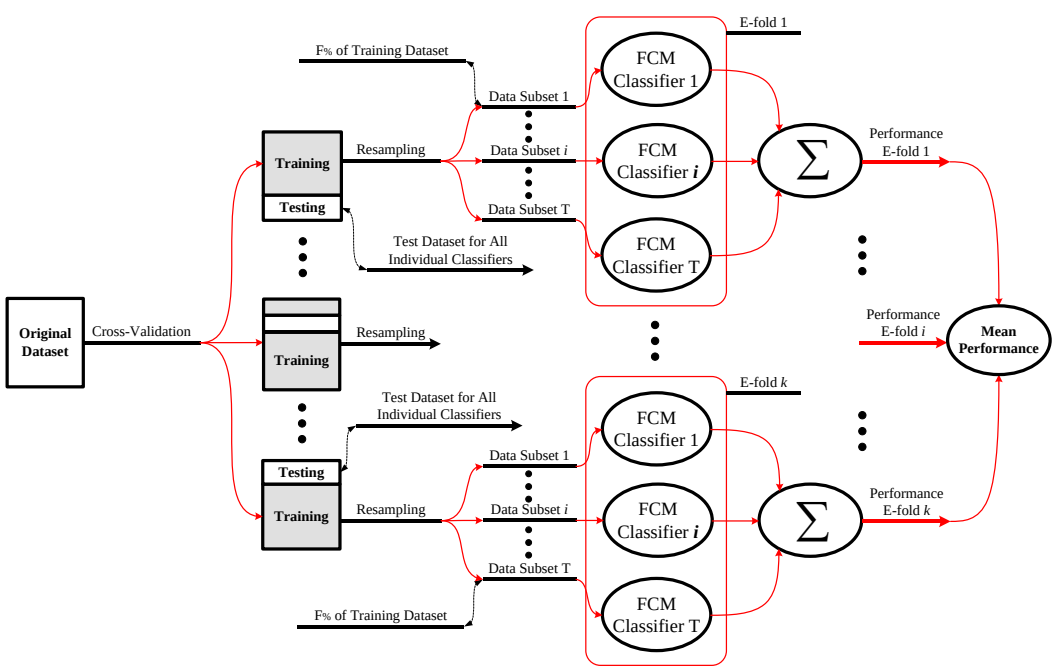

Figure 2. The general structure of the ensemble scheme, with cross-validation; E-fold = Ensemble-fold

3 ESTIMATION OF THE CLASSIFICATION CONFIDENCE

The test performance of a classifier, i.e. the ratio between the number of test patterns correctly classified and the total number of test patterns, is usually computed on a set of test patterns different from those used to train the model. Thus, if we consider a new pattern $\vec{x}$, the test performance can be interpreted as the probability that $\vec{x}$ is correctly classified, and it represents a measure of the expected confidence in its classification. However, since the performance may vary in different zones of the input space, this confidence measure is not satisfactory given that it does not take into account the location of the test pattern.

In the present work, as measure of the confidence in the assignment of the test pattern $\vec{x}$ to class Γ_i^t, we consider the posterior probability $P\left(\Gamma_i^t \middle| \vec{x}\right)$, i.e. the probability that the class Γ_i^t is the correct class given that the test pattern is $\vec{x}$. Using Bayes rule:

$$P\left(\Gamma_i^t \middle| \vec{x}\right) = \frac{P\left(\vec{x} \middle| \Gamma_i^t\right) P\left(\Gamma_i^t\right)}{\sum_{j=1}^{c} P\left(\vec{x} \middle| \Gamma_j^t\right) P\left(\Gamma_j^t\right)} \tag{1}$$

where $P\left(\vec{x} \middle| \Gamma_i^t\right) d\vec{x}$ is the probability that a generic pattern of class Γ_i^t be in $d\vec{x}$ around $\vec{x}$ and $P\left(\Gamma_i^t\right)$ is the a priori probability that a pattern is of class Γ_i^t, i.e. the probability that a pattern belongs to class Γ_i^t prior to observing any information about the pattern. Since the probabilities $P\left(\vec{x} \middle| \Gamma_i^t\right),\ i = 1, \ldots, c$, are not known in practical classification problems, it is necessary to estimate $P\left(\Gamma_i^t\right)$ from the available experimental information [18]. Several approaches have been proposed to this purpose, both based on the estimation of $P\left(\vec{x} \middle| \Gamma_i^t\right)$ (sampling paradigm) or on the direct estimation of $P\left(\Gamma_i^t \middle| \vec{x}\right)$ (diagnostic paradigm). Since the former approaches are usually difficult to apply to large dimension problems, a direct estimator of $P\left(\Gamma_i^t \middle| \vec{x}\right)$ for ensemble classification systems, based on the softmax function which ensures all $P\left(\Gamma_i^t \middle| \vec{x}\right)$ are between 0 and 1 and have summation equal to 1 [19], is here applied:

$$P\left(\Gamma_i^t \middle| \vec{x}\right) = \frac{\exp^{A_i\left(\vec{x}\right)}}{\sum_{j=1}^{c} \exp^{A_j\left(\vec{x}\right)}} \tag{2}$$

where $A_i(x)$ is the sum of the classifier weights that select class Γ_i^t, given by:

$$A_i\left(\vec{x}\right) = \sum_{\alpha=1}^{T} v_{\alpha,i} \quad with \quad v_{\alpha,i} = \begin{cases} w_\alpha & if\ the\ \alpha-th\ classifier\ assigns\ pattern\ \vec{x}\ to\ class\ \Gamma_i^t \\ 0 & otherwise \end{cases} \tag{3}$$

being T the number of classifiers of the ensemble and w_α the weight assigned to the $\alpha-th$ classifier in the combination of the individual classifier outputs according to the weighted

majority voting scheme. This estimator has been shown to be unbiased, i.e. on the average, a fraction p of all patterns classified with confidence $P\left(\Gamma_i^t \middle| \vec{x}\right) = p$ is correctly classified [19].

In the present work, equation 3 is applied to a bagging ensemble algorithm. Since the individual base classifier outcomes are aggregated according to the majority voting rule, the same weight is assigned to all the classifiers. Thus, in equation 3, $v_{\alpha,i}$ is given by:

$$v_{\alpha,i} = \begin{cases} 1 & \textit{if the } \alpha_{th} \textit{ classifier assigns pattern } \vec{x} \textit{ to class } \Gamma_i^t \\ 0 & \textit{otherwise} \end{cases} \tag{4}$$

Notice that the confidence in the classification to a given class Γ_i^t of a test pattern increases if the classifiers that do not identify Γ_i^t assign different classes among themselves, instead. For example, let us consider the two following cases in the classification of a pattern $\vec{x}$ by an ensemble of ten classifiers:

a) Six classifiers assign $\vec{x}$ to class 1, four classifiers to class 2

b) Six classifiers assign $\vec{x}$ to class 1, one classifier to class 2, one classifier to class 3, one classifier to class 4, one classifier to class 5

In case a), the confidence associated to class 1 is 0.875 and to class 2 is 0.125; in case b), the confidence associated to class 1 is 0.974 and to the other four classes is 0.006.

4 APPLICATION TO NUCLEAR TRANSIENT IDENTIFICATION

In this Section, the ensemble-based fuzzy FCM is applied to the classification of transients in the feedwater system of a BWR. Fig. 3 shows a sketch of the system. Seven fault classes F1-F7 are considered. A number of transients corresponding to these faults have been simulated by the HAMBO simulator of the Forsmark 3 BWR plant in Sweden [15]. The proposed faults are as follows:

1. F1-F4 regard line 1 of the feedwater system.

2. F5-F7 regard both lines.

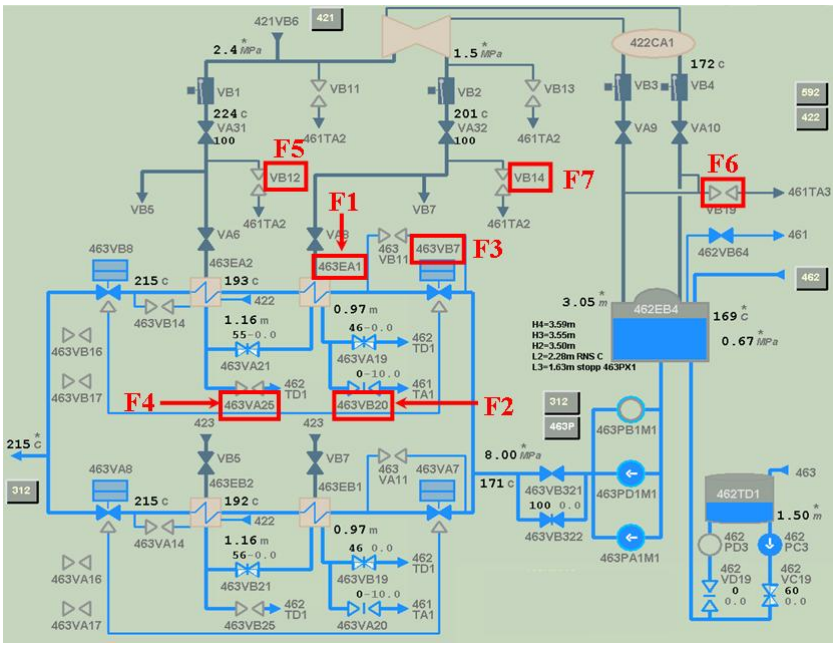

Figure 3. A sketch of the feedwater system of the BWR [11]

Twenty measured signal features have been used for the fault classification: Temperature of the high-pressure drain, Temperature after EA1 (high pressure preheater A1), Temperature of drain 4 before VB3, Temperature of condensate after EB2 train B, Temperature of condensate after EB2 train A, Temperature of condensate after EB3 train B, Water level tank TD1, Water level in EA2 train A (preheater A2), Water level in EB2 train B (preheater B2), Temperature of the feedwater before EA2 train A, Temperature of the feedwater after EA2 train A, Temperature of the feedwater before EB2 train B, Temperature feedwater after EB2 train B, Position level of the control valve for EA1, Position level of the control valve for EB1, Position level of the control valve before EA2, Position of the emergency drain valve EB2, Position of the valve for level I EB4, Position of the steam stop control valve.

Five transients were simulated for each of the eight faults, considering different degrees of leakage and valve closures and with step and ramp changes. All transients start after 60 seconds of steady state operation. For each fault, three transients corresponding to step changes have been considered since they lead to an early variation of the measured variables. The data relative to the selected twenty signals were recorded with a sampling frequency of 1 Hz. With the goal of 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. Ten data subsets each one made of 4177 patterns have been bagged from this dataset and used to train 10 FCM classifiers, as explained in Section 2. The performance of classification obtained by the bagging ensemble in a 10-fold cross-validation scheme turns out to be 0.973 ± 0.013.

In order to investigate the confidence of the algorithm in the classification of the test patterns, the results achieved testing a single fold during the cross-validation process are analyzed. In the considered case, 455 patterns out of the total 464 test patterns have been correctly classified by the algorithm. Equation 2 provides the degrees of confidence in the

assignment of each test pattern to each of the seven classes. Notice, however, that the most interesting information regards the confidence in the assignment of the test pattern to the class with the highest number of votes, i.e. the class actually assigned by the ensemble system according to the majority voting rule adopted. In this respect, Fig. 4 reports the distribution of the confidence values (equation 2) assigned to the class to which each of the 464 test patterns has been assigned.

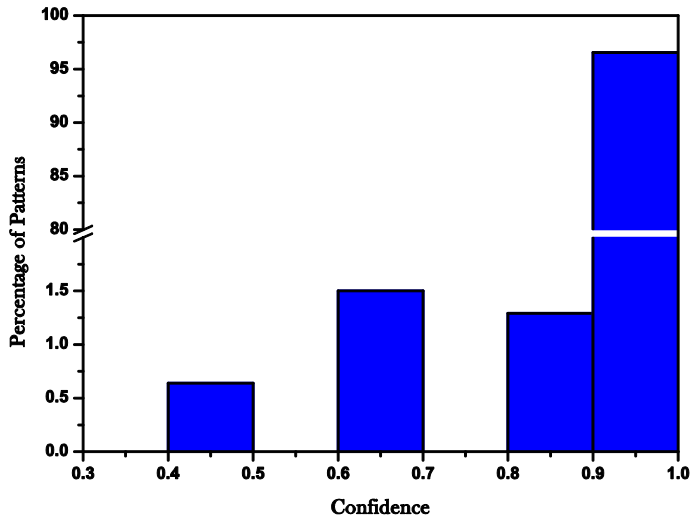

Figure 4. Distribution of the confidence assigned to the class with the highest number of votes. All the patterns in the last bin [0.9-1] are actually assigned with a confidence greater than 0.97.

Notice that 96.55% of all the test patterns are assigned to a class with a confidence greater than 0.97, indicating a very high degree of confidence in the classification of a great fraction of the test patterns. Table I shows that these high degrees of confidence are justified since the performance in the classification of these 448 patterns is very high (0.987).

Table I: Number of patterns classified with a confidence value in the range indicated in the first row, and performance in the classification of the corresponding patterns.

Confidence range	(0.4, 0.5]	(0.5, 0.6]	(0.6,0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1]
Number of patterns	3	0	7	0	6	448
Performance	0.667		1.000		0.667	0.987

Furthermore, notice that the three patterns assigned to a class with a confidence lower than 0.5 are all characterized by a tie between two classes in the votes assigned by the ten classifiers of the ensemble (Table II). Although the majority voting rule assigns the three patterns to a class randomly chosen between the two classes with the highest number of votes, each of these patterns is actually assigned to two classes with the same confidence of 0.492. It is interesting to observe that there are globally six classes with associated confidence close to 0.5 (class 2 and 7 for patterns 25 and 159, class 6 and 7 for pattern 144) and that out of these six assignments, three

are correct (patterns 25 and 159 are of class 2 and pattern 144 is of class 7). Thus, a 50% of all class assignments with confidence close to 0.5 are correct.

Table II. Number of votes assigned by the 10 classifiers of the ensemble to the three patterns classified with a confidence of 0.49. The patterns misclassified by the ensemble are reported in red.

Pattern	Number of votes (confidence in the assignment)							Assigned Class	True Class
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7		
25	0 (0.003)	5 (0.492)	0 (0.003)	0 (0.003)	0 (0.003)	0 (0.003)	5 (0.492)	2	2
144	0 (0.003)	0 (0.003)	0 (0.003)	0 (0.003)	0 (0.003)	5 (0.492)	5 (0.492)	6	7
159	0 (0.003)	5 (0.492)	0 (0.003)	0 (0.003)	0 (0.003)	0 (0.003)	5 (0.492)	2	2

Finally, Table III reports the number of votes obtained by the classes for those patterns classified with a confidence between 0.8 and 0.9. In this case there are two errors.

Table III. Number of votes assigned by the 10 classifiers of the ensemble to the patterns associated to a confidence between 0.8 and 0.9. The patterns misclassified by the ensemble are reported in red.

Pattern	Number of votes (confidence in the assignment)							Assigned class	True Class
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7		
42	0 (0.002)	4 (0.118)	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	6 (0.871)	7	2
231	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	4 (0.118)	6 (0.871)	7	7
274	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	6 (0.871)	4 (0.118)	6	6
326	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	6 (0.871)	4 (0.118)	6	6
443	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	0 (0.002)	6 (0.871)	4 (0.118)	6	7
271	0 (0.006)	1 (0.016)	5 (0.879)	0 (0.006)	0 (0.006)	2 (0.044)	2 (0.044)	3	3

5 CONCLUSIONS

This work proposes a method for the estimation of the confidence in the identification of nuclear transients by a bagged ensemble of FCM classifiers. The confidence in the final classification is obtained by estimating the Bayesian posterior probability of the class by a softmax approach applied to the outcomes of the ensemble classifiers.

The application of the proposed method to the classification of nuclear transients in the feedwater system of a BWR nuclear power plant proves effective in achieving a reliable estimation of the confidence in the classification. In particular, very few patterns classified with high confidence values are misclassified, whereas, as expected, low classification performances are obtained when considering the few patterns classified with low confidence values.

6 REFERENCES

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

[2] K. Kim, E.B. Bartlett, "Nuclear power plant fault diagnosis using neural networks with error estimation," *IEEE Transactions on Nuclear Science*, **43 (4)**, pp. 2373-2388 (1996).

[3] D. Roverso, "Soft computing tools for transient classification," *Information Sciences*, **127**, pp. 137-156 (2000).

[4] M.J. Embrechts, S. Benedek, "Hybrid identification of nuclear power plant transients with artificial neural networks," *IEEE Transactions on Industrial Electronics*, **51**, pp. 686–693 (2004).

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

[6] K. Zhao, B.R. Upadhyaya, "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 (2005).

[7] A. Evsukoff, S. Gentil, "Recurrent neuro-fuzzy system for fault detection and isolation in nuclear reactors," *Advanced Engineering Informatics*, **19 (1)**, pp. 55-66 (2005).

[8] E. Zio, G. Gola, "Neuro-fuzzy pattern classification for fault diagnosis in nuclear components," *Annals of Nuclear Energy*, **33**, pp. 415–426 (2006).

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

[10] J. Reifman, "Survey of artificial intelligence methods for detection and identification of component faults in nuclear power plants," *Nuclear Technology*, **119**, pp. 76-97 (1997).

[11] P. Baraldi, R. Razavi-Far, E. Zio, "Bagged Ensemble of FCM Classifier for Nuclear Transient Identification," *Submitted to Annals of Nuclear Energy*, (2010).

[12] L. Breiman, "Bagging predictors," *Machine Learning*, **24 (2)**, pp. 123–140 (1996).

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

[14] R.P. Duin, M. Tax, "Classifier conditional posterior probabilities," *Lecture Notes in Computer Science*, **LNCS 1451**, pp. 611-619, (1998).

[15] E. Puska, S. Normann, "3-d core studies for hambo simulator," *Proceedings of presentations on man-machine system research,* Enlarged Halden programme group meeting, Vol. 2, (2002).

[16] B. Yuan, G. Klir, Intelligent Hybrid Systems Fuzzy Logic, Neural Network, and Genetic Algorithms, Kluwer Academic Publishers, (1997).

[17] R. Polikar, "Bootstrap-inspired techniques in computational intelligence," *IEEE signal processing magazine*, **59**, pp. 59-72 (2007).

[18] D.J. Hand, K. Yu, "Idiots Bayes: not so stupid after all," *International Statistical Review*, **69**, pp. 385-398 (2001).

[19] M. Muhlbaier, A. Topalis, R. Polikar, "Ensemble Confidence Estimates Posterior Probability," *Lecture Notes in Computer Science*, *Multiple Classifier Systems*, **LNCS 3541**, pp. 326–335 (2005).