Power Plant Condition Monitoring by Means of Coal Powder Granulometry Classification

Damiano Rossetti[a,b,*], Stefano Squartini[a], Stefano Collura[b], Yu Zhang[c]

[a]*Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy*
[b]*Research for Innovation, Loccioni Group, Angeli di Rosora, Ancona, Italy*
[c]*School of Engineering, University of Lincoln, Lincoln, United Kingdom*

Abstract

In this work, a condition monitoring approach suitable for coal fired power plant is proposed. This approach is based on classification techniques and it is applied for the monitoring of the Particle Size Distribution (PSD) of coal powder. For coal fired power plant, the PSD of coal can affect the combustion performance, therefore it is a meaningful parameter of the operating condition of the plant. Three tests have been carried out aimed to study the effect of the class numbers, the dataset size, and the reduction of the number of false positives on the effectiveness of the approach. For each designed test, three standard classification algorithms, i.e. Artificial Neural Network, Extreme Learning Machine and Support Vector Machine, have been employed and compared. Experimental data taken from 13 measuring point on 13 burners of two different industrial power plants have been used. Obtained results showed that, using two classes give the most accurate results, using only the 90% of the available data can still provide comparable classification results, and the level of false positive can be effectively reduced.

Abbreviations: PSD, Particle Size Distribution; AE, Acoustic Emission; SVM, Support Vector Machine; ANN, Artificial Neural Network; ELM, Extreme Learning Machine; FP, False Positive; WP, Wavelet Packet; CV, Cross Validation; G-PSD, Good PSD, P-PSD, Poor PSD; DTT, Decision Threshold Technique; ROC, Receive Operating Curve; DET, Detection Error Trade-off; TPR, True Positive Rate; FPR, False Positive Rate; FNR, False Negative Rate; ACC, Accuracy; SENS, Sensitivity; TP, True Positive; TN, True Negative; FN, False Negative; STD, Original Classifier.

*Corresponding author

Email address: d.rossetti@pm.univpm.it (Damiano Rossetti)

1. Introduction

In many industrial processes, the particle size of powder represents an important parameter as it affects both physical and chemical behaviours of the powder. In powder analysis it is often interesting to focus the attention on quantities that represent average characteristics of the whole flux, instead of a single particle. In particular, it is commonly agreed that particle size can be well described by means of cumulative parameters known as PSD. The PSD is a list of values representing the discretization of the size distribution curve. Each value is usually expressed in terms of percentage, typically by mass, of particles with a certain size. Each industrial application has its reference range of particle size that can be related to optimal process conditions. Within an industrial plant, the most used approach to measure PSD is the sampling and sieving method: a certain amount of powder is sampled inside the process by introducing a probe into the duct conveying the powder; in a second phase, the sample is sent to the laboratory where it is sieved and classified through a nested column of sieves of decreasing screen openings. This method produces an accurate estimation of the PSD for a given time instant, but it is time consuming and it is not suitable for continuous monitoring. In order to cope with this latter issue, it is necessary to use a system that is able to perform the PSD estimation on line and in line with a non-intrusive equipment.

1.1. Background

For the purpose of this work, two thermal power plants are considered as case study. A thermal power plant produces energy from the coal combustion and it must keep the condition of highest efficiency in order to reduce fuel consumption and emissions. Inside a power plant, fine ground coal powder is used as fuel and it is conveyed by air within ducts in order to feed boiler burners. One key

aspect that affected the combustion efficiency is the size of the coal powder, for this reason it is important that PSD of coal feeding the burners remains within specific ranges to avoid an efficiency dropping. It is possible to correlate PSD to operational failure or wrong setting of coal grinding mills that lead to poor efficiency, thus the possibility to on line monitor of coal particle size within the process can be an useful tool in order to set the plant working parameters to keep acceptable combustion efficiency. One possible way to gain information about the flux of coal powder is by monitoring the Acoustic Emission (AE) signals produced by the impact of powder on the inner surface of these ducts [15].

The relationship between particle size and AE was demonstrated by Leach *et al.* [1, 2] that were the first to use AE signals produced by the impact of a single particle regularly shaped on a metallic surface as meaningful quantity to measure the size of the particle itself. During the years, many applications have appeared in the literature [3, 4, 5], confirming the suitability of AEs for PSD measurement in engineering problems.

Moving from these premises, the authors have used diverse machine learning techniques to train models for the estimation of PSD of coal powder by exploiting AE based information [6]. Data collected from different burner feeding ducts of the same power plant were used for training and testing suitable supervised learning algorithms for regression.

For many practical applications, where it is not necessary to have a punctual estimation of the PSD, a classification approach can be a valuable alternative to relate the AE and PSD. Once it has been defined a threshold PSD beyond which the plant performances decrease, it is possible to distinguish at least 2 classes of PSD, the one associated with good working conditions and the one associated with bad working conditions. Therefore, it is possible identifying the working conditions that ensure an useful monitoring of the plant, and dividing these conditions according the defined classes.

The problem of powder, or particles, detection and classification is common in many applicative contexts and during the past years several solutions were

3

presented to solve specific tasks that involve pulverized material for both civil [7, 8] and industrial [9, 10] environments. In a preliminary study [11], the authors have already investigated the efficacy of an algorithm based on Support Vector Machine (SVM) for the classification of two types of food powder characterized by different particle size. In that case AE based information were exploited to characterize the powder and used to train a classifier based on SVM. The study demonstrated that a machine learning approach can be used to solve powder classification problem by exploiting AE signals.

1.2. Contribution

All experiences with earlier work have been used by the authors to develop a new approach for condition monitoring of coal powder burned as fuel in power plants. The method proposed in this paper uses the PSD, measured by means of AEs, as an indication of the operating state of a monitored boiler burner where the combustion of the powdered coal takes place.

Applying a regression on the data provides with an estimation of the actual PSD value, but the training of a monitoring system via supervised regression algorithms implies the need of reference PSD data obtained by collecting and sieving samples of powder during specific system set-ups. This procedure must be repeated several times to collect enough examples for the training, leading to longer time for the set-up of the whole monitoring system.

For a condition monitoring purpose the evaluation of the general state of the system is needing, so it is not necessary to know the punctual PSD value. In this situation, the usage of a classification approach can allow to reduce the effort for labelling data and to use qualitative feedbacks get from the final phase of the monitored process for clustering the AEs signals associated with different PSDs and system set-ups. For this case study, coal powder is used to feed a boiler and the combustion efficiency can be evaluated by measuring the coal specific consumption, the ashes and the exhausted gas composition.

Three different machine learning algorithms have been implemented and compared in our analysis, i.e. SVM, Artificial Neural Network (ANN) and

4

Extreme Learning Machine (ELM), to correlate AE signals acquired on the feeding ducts and PSD. The proposed method was tested on data coming from two different industrial power plants.

As first step some tests have been carried out to evaluate the performance variation due to the number of classes. Three divisions of the output space with 2, 4 or 6 classes are proposed. The binary classification is the simplest classification that allows to divide the output space into two regions associated respectively with the negative and the positive elements. Adding more classes provides a better precision and the possibility to monitor multiple operating conditions to monitor.

A second set of tests was carried out in order to find the minimum number of examples required to train the classifier. Indeed, a crucial issue for a supervised approach is the amount of examples that are necessary to train properly the classification models. For augmenting the number of training examples, the simplest approach is to use the standard sample and sieve method, but this procedure is time consuming and costly. In a previous work, the authors proposed an alternative approach [12], to use data collected from multiple sources in order to increase the amount of training examples. The authors explored supervised strategies that are able to exploit such extended availability of heterogeneous data to improve the PSD estimation performance. The present work modifies the learning paradigm and designs a learning algorithm based on classification. The classification provides a target evaluation less precise than the regression, being able to identify only the group to which the example belongs and not the real value associated with it. The loss of precision may indicate that a classification model is less affected by the examples number than a regression model, and the same classification accuracy is achievable with fewer examples and without the use of more complicated techniques. For this reason, this paper presents a series of tests carried out to asses the performance variation due to reduction of the training samples and identifying the minimum number needing to obtain the desired classification accuracy.

The third set of tests was aimed to reduce the amount of False Positives (FP).

The proposed method exploits a variation of the Receiver Operating Curve to find the optimal decision threshold that provides the lowest level of false positive and the highest accuracy.

The effectiveness of the proposed monitoring scheme has been demonstrated on experimental data by performed computer tests, as detailed in the following. Although this method is applied on a specific task, as the monitoring of boiler burners in power plants, it can be suitable for all those scenarios where it is requested a monitoring system to discriminate between a discrete number powder types, different by material or dimension.

This paper is organized as follows. Section 2 describes the acquisition system and data processing, with details about the algorithms used for computer simulations. The methodological approach is presented in Section 3. Section 4 describes the used dataset and the results obtained in the diverse experiments addressed. Finally, conclusions are drawn in Section 5.

2. Data Acquisition and Processing

2.1. Acquisition System and Methodology

In this work, experimental data have been collected by means of an industrial AE monitoring system named POWdER [13], installed on the burners feeding ducts of two different power plants. This system continuously monitors AE signals and converts them in terms of PSD of coal powder conveyed inside the ducts. The POWdER sensors are installed near a duct curve because in this point there is the highest probability that the particles hit the inner surface of the duct and generate the AE. The curve is the final part of a feeding duct that carries the coal powder from mill to the burners in the boiler. In Figure 1 the typical installation of AE sensors on a plant is sketched; the mill that grinds the coal, the feeding ducts that carry the coal powder and the boiler where the coal combustion occurs are shown.

The measure of interest is the PSD of coal powder flowing inside the feeding ducts of the burners. This quantity can be characterized by several meshes,
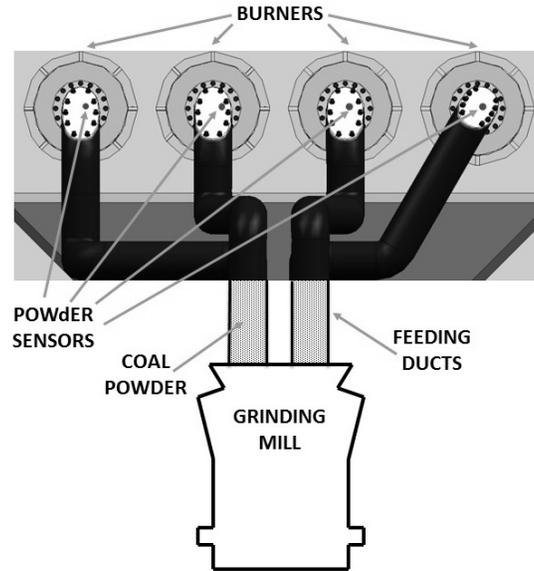
Figure 1: Sketch of the power plant structure

in the case study proposed in this work, two sieves, corresponding to 50MESH (i.e. 300μm), and 200MESH (i.e. 75μm), were taken into account. Each PSD is therefore represented by two numerical values, corresponding to the percentage of coal particles in the initial sample whose dimensions are, respectively, lower than 300μm and 75μm.

Signal acquisitions have been carried out in several ducts and different operating conditions, acquiring 200 consecutive acquisitions sampled on a 50ms time interval and a sample-rate of 2MHz, for a total of 10 seconds. The recorded signals are used in the model training procedure along with a label associated with the corresponding PSD measure.

2.2. Data Processing

Let $x_i(t)$ be the generic voltage signal acquired trough an AE transducer, where t is discrete time index and i denotes the specific acquisition, as the one shown in Figure 2. The signal energy distribution in frequency domain holds the information about the PSD profile, therefore, it is important to identify a suitable set of features able to characterize that distribution so that the AE
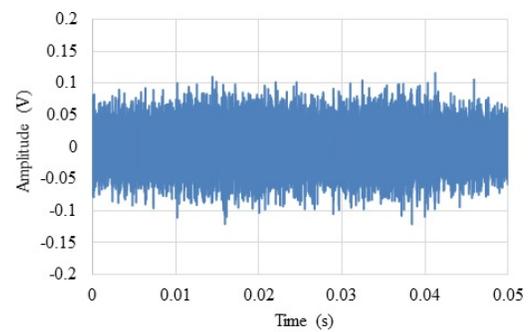
7

Figure 2: Typical time-waveform of acquired AE signal. On the y-axis, the amplitude represents the sampled voltage signal at the sensor output.

signals corresponding to similar PSD lead to similar feature vectors, while the difference among feature vectors increases as PSD changes.

The proposed feature extraction procedure is based on complete Wavelet Packet (WP) decomposition [14] as introduced by Bastari and al. [15], who demonstrated the effectiveness of WP for the estimation of PSD of coal powder via AEs on a similar application.

A WP decomposition over 6 levels was used to decompose the signal $x_i(t)$ into $2^6 = 64$ sub-band sets of coefficients. The feature vector X_i in then obtained by calculating the energy content of each of these sub-bands. These 64 values are the parameters used to characterize the AE signal and, trough this, the PSD. Figure 3 shows an example of how the features look like, 4 vectors X_i extracted from data of one duct, in two operating conditions: 2 vectors associated to good working condition (dashed line) and 2 associated to bad condition (solid line). Vectors belonging to the same condition are almost identical, while it is possible to distinguish vectors associated with good condition from those associated with bad condition.

Feature vectors have been averaged on a proper time interval in order to obtain a single vector $\overline{X}_i$ for minimizing the effects of signal fluctuations. Thus, averages on vectors X_i have been performed on 200 time contiguous acquisitions for 10 seconds of time signal, obtaining an WP-based average vector $\overline{X}_i$.
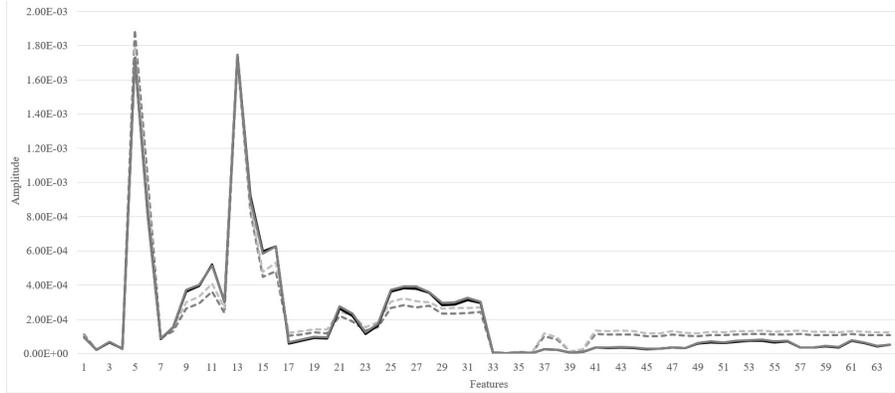
8

Figure 3: Wavelet packet based features extracted from data of a single duct in positive condition (dashed line) and in negative condition (solid line).

2.3. Machine Learning Algorithms

In this work, three machine learning algorithms have been employed for classification purposes: Artificial Neural Network, Support Vectors Machine and Extreme Learning Machine. All tests have been performed by using Cross Validation (CV) K-fold procedure with 6 not-overlapping folds. At each CV iteration, a different combination of Training and Testing sets is selected from the dataset under test. At the end of the CV procedure, all the outcomes for each combination Training-Testing set were gathered together in an array and used for the performance evaluation. A second CV procedure is performed to find the optimal parameters needed for the training of the classifiers. In this case, the Validation set is extracted from the Training set and the parameters that provide the best result on this dataset are selected.

2.3.1. Artificial Neural Networks [16]

The ANN is a well documented approach, used for a classification purpose in many studies. This kind of network is composed by layers characterized by a fixed number of parallel neurons. Assuming a network with M layers and N_l neurons in the l-th layer, with $l = 1, \ldots, M$, the description of the output state

9

for the k-th neuron in the l-th layer, $y_k^{(l)}$, is obtained with the equation:

$$y_k^{(l)} = f\left(\sum_{j=0}^{N_{l-1}} \mathbf{W}_{kj}^{(l)} \cdot \mathbf{X}_j^{(l)}\right), \tag{1}$$

where $\mathbf{W}$ and $\mathbf{X}$ respectively are the weights and the inputs matrices, and f is the activation function. In particular, $\mathbf{W}_{kj}^{(l)}$ represents the weight adopted by k-th neuron into the l-th layer and applied at the j-th input. $\mathbf{X}_j^{(l)}$ represents the j-th input value for the neurons in the l-th layer. Consequently, for the first and the last layer it is assumed that:

$$\mathbf{x} = \left[\mathbf{X}_0^{(1)}, \ldots, \mathbf{X}_{N_1}^{(1)}\right], \qquad \mathbf{y} = \left[y_0^{(M)}, \ldots, y_{N_M}^{(M)}\right], \tag{2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the input and output data vectors, respectively. For this study, the Unipolar Sigmoid has been chosen. During the training process, the weights and the bias are update with the Backpropagation [17] algorithm.

The standard structure selected for the tests uses one input layer, two hidden layers and one output layer. The number of hidden layers nodes have been varied, from 50 to 100 for the first layer and from 30 to 80, in order to identify the configuration that minimizes the estimation error. A wide range of experimental tests with the available datasets have shown that the classification performance typically decreases when more than 100 neurons and 80 neurons for the first and the second layer respectively.

2.3.2. Support Vector Machine [18]

SVMs are binary classifiers that discriminate whether an input vector x belongs to class $+1$ or to class -1 based on the following discriminant function:

$$f(x) = \sum_{i=1}^{N} \alpha_i d_i K(x, x_i) + b, \tag{3}$$

where $d_i \in \{-1, +1\}$, $0 < \alpha_i < C$ and $\sum_{i=1}^{N} \alpha_i d_i = 0$. The terms x_i are the support vectors and b is a bias term that together with the α_i are determined during the training process of the SVM. The input vector x is classified as $+1$ if $f(x) \geq 0$ and -1 if $f(x) < 0$. The input vector x is classified as $+1$ if $f(x) \geq 0$

and -1 if $f(x) < 0$, given by the separation hyperplane defined by the support vectors.

The kernel function $K(\cdot, \cdot)$ can assume different forms [19]. The Radial Basis Function (RBF) kernel, which is defined in (4), has been selected for the tests.

$$K(x_i, x_j) = \exp(-\gamma \|x_i + x_j\|^2), \tag{4}$$

γ plays a major role in the performance of the kernel and must be optimized.

The C and γ parameters are determined by performing the grid search approach to find the optimal values, within the following ranges:

$$\mathbf{C} = \left[2^0, 2^1, \ldots, 2^8\right], \qquad \gamma = \left[2^{-5}, 2^{-4}, \ldots, 2^3\right], \tag{5}$$

All the experiments have been performed using the tools provided by LibSVM [20], a library for Support Vector Machines.

2.3.3. Extreme Learning Machine [21, 22]

The ELM has been presented as a fast learning algorithm for Single Layer Feedforward Networks. Differently from standard ANN approaches, the input weights are randomly generated and the output ones are tuned by a least-square method. In this paper a kernel-based version of the ELM has been implemented. Given a set of pairs (x_i, t_i), $i = 1, \cdots, N$, where $x_i \in \mathbb{R}^L$ is the training data, and $t_i \in \{-1, 1\}$ denotes the corresponding label, the output function of the ELM is computed as:

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \mathbf{\Omega}\right)^{-1} \mathbf{T}, \tag{6}$$

where $K(\cdot, \cdot)$ denotes the chosen kernel function, C is the regularization coefficient and $\mathbf{\Omega}$ defines the kernel matrix, so that $\Omega_{i,j} = h(x_i) \cdot h(x_j) = K(x_i, x_j)$. In this paper the kernel-based approach is used to implement a classification algorithm and the the RBF kernel has been evaluated, as for SVM. Also in this case the optimal γ parameters for the kernel are calculated by performing the grid search approach, within the same range of SVM in (5).

3. The Proposed Condition Monitoring Approach

The proposed condition monitoring approach uses the AE signals for the classification of pulverized coal PSD in order to monitor the combustion performances of the boilers in a power plant. This approach has been designed and applied on data coming from two industrial coal fired power plants, in order to prove its effectiveness in an actual industrial context and to demonstrate its generality and independence from the a specific plant type. This Section reports in details the tests designed and carried out to show the response of the proposed approach to specific issues.

As mentioned in Section 2.1, the PSD for the coal powder is expressed by mean two quantities, identified as 50 mesh and 200 mesh. Therefore, two classifiers have been trained on the data of these two meshes.

3.1. Binary Classification

The standard approach to the problem of PSD classification involves the use of only two classes to identify two size ranges associated with different operating conditions of the system. With the aim of developing a control system, the classes association was designed to distinguish those values of PSD that are good for efficient combustion from the ones that can lead to non-optimal combustion. For this reason the first class, that can be identified as the corrected class or good PSD (G-PSD), is associated with the PSD values which are required for the correct operation of the system and for maximizing the combustion efficiency. Under these conditions, the monitoring system provides with a positive feedback the operator who can verify that the monitored system is operating within the parameters established for proper functioning. The negative class is associated with those PSD values which indicate the non-optimal operating conditions or poor PSD (P-PSD), in this case the monitoring system reports a negative feedback that indicates the necessity to modify the system parameters in order to return to the proper functioning.

The association of the examples with the proper labels is performed by mean a threshold-based labelling that divides the range of boiler operating conditions.

12

Control room operator may set different action for the mill, the burners and the boiler, and each set-up leads to a different combustion efficiency. The threshold is identified for each mesh size and each power plant by using the information from real coal combustion process in terms of cycle efficiency, ash composition and exhaust gas composition.

3.2. Multi-class Classification

Adding more classes has an impact on the accuracy of the classification. This kind of modification to the original classification can be useful in order to obtain early warnings as the plant approaches to the critical condition. Two series of tests with 4 and 6 classes were carried out to show the variation of performance depending on the number of classes. The performance obtained is compared with that obtained with the algorithms with 2 classes.

In order to increase the number of classes, it is necessary to set up new thresholds for a further division of the range of the operating conditions. These thresholds are chosen in the range of the critical threshold, previously defined for the binary classification, for a double purpose: identifying how much the powder size is getting close to the critical threshold, and quantifying how much the PSD is over this threshold. Figure 4 shows a comparison between the cases with 2 and 6 classes. Figure 4a explains how it is possible to detect the occurrence of a problem only when the problem is already occurred if the output space divided by two classes. In contrast, the presence of additional thresholds provides a form of early alert that can allow to prevent the problem occurrence, as shown in Figure 4b.

3.3. Dataset Reduction

Algorithms based on machine learning techniques are powerful tools that allow to create models able to associate the measured AEs with the PSD of the powder that generated them. These techniques are strongly influenced by the data available for training, therefore it is essential to gather a sufficient amount of data from the monitored system, in order to be able to properly
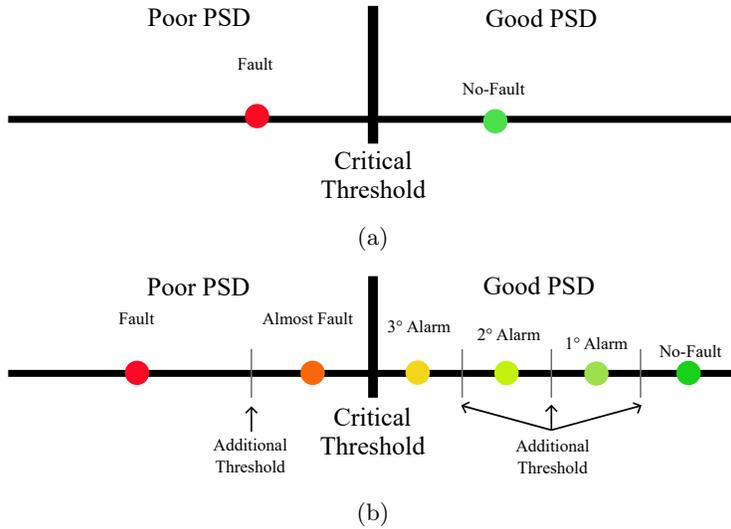
13

Figure 4: Comparison between the monitoring capabilities with (a) 1 threshold and 2 classes, and (b) 5 thresholds and 6 classes.

characterize it. As detailed in the Section 1, it is expensive, both in terms of time and resources, to get examples with associated targets needed for a supervised training. Thus, it is important to be able to design accurate models with as less need of examples as possible. Due of the centrality of this aspect, a study was carried out to understand the evolution of the classification accuracy for a binary classifier by varying the number of examples used for training the models, with the goal of identifying the minimum amount of examples necessary to achieve an accuracy that can be an acceptable compromise between classification accuracy and development cost.

To cope with this aim, in addition to the models that were already trained with the whole dataset, other three sets of models were trained using a sub-set of the whole dataset. To select the sub-sets, the last acquired examples in terms of time were removed from the original dataset. This selection aims to simulate a shorter period for the acquisition of the examples that populate the training set, delivering insight on how much these examples should be for a proper training. The three sub-set contain respectively 90%, 75% and 50% of the dataset. In

order to exploit all available data, the removed portion of the dataset is added to the selected Testing set during the K-folds CV procedure. In this way, the dataset used for the test is composed by a fixed part, the removed portion from the original dataset, and a variable part, chosen among the K folds. The Training set is extracted at each CV iteration from the reduced dataset, so each of the three tests with a different sub-set has a different amount of examples for the training.

3.4. False Positive Reduction

A crucial issue for a monitoring system is the possible presence of P-PSD that are classified as G-PSD. The presence of FP can lead to unwanted operating situations and compromising the system operation.

Many works in literature face this issue and provide with many solutions [23, 24, 25]. The complete elimination is not achievable because in the real case it is always possible dealing with a peculiar scenario. On the other hand, reducing the FP it is possible by improving the classification algorithm with other techniques, as it is detailed in this section where the procedure will be presented.

The proposed procedure, named Decision Threshold Technique (DTT), employs a variant of the Receive Operating Curve (ROC), the Detection Error Trade-off (DET). The ROC is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings, the DET is slightly different and it is a graphical plot of False Negative Rate (FNR) against the FPR. Through this representation the decision threshold that provides zero FPR and minimum FNR on the Training set is selected and used with the Testing data.

4. Computer Tests and Results

All tests have been performed with Matlab 2013a$^{\circledR}$ running on a PC with a processor i5 dual-core 2.3 GHz, 4 GB of RAM and Windows 7$^{\circledR}$ OS.

4.1. Datasets

Two distinct datasets acquired from two different coal-fired power plants are involved in this work. The two plants are identified as Plant A and Plant B. The proposed method has been developed by using the data from the Plant A, while the data from the Plant B have been used for testing the generality of the method.

The datasets gather the data collected from sensors installed on 13 ducts of the power plants, specifically 5 for Plant A and 8 for Plant B. The vector $\overline{X}_i = [\overline{x}_1, \overline{x}_2, ..., \overline{x}_{64}]$ represents the vector of the features extracted for the i-th AE signal measured on a duct. The vector of labels corresponding to the single observation is $L_i = [l_1, l_2]$, with l_1 and l_2 that are respectively the classes associated with 50 mesh and 200 mesh. Each dataset is split into 2 sub-dataset, for the 50 mesh and for the 200 mesh, that share the same features with different labels.

4.2. Metrics

Different metrics can be employed to evaluate the performance of the proposed techniques, so for each technique discussed in Section 3 the most appropriate metrics are reported. All the proposed metrics are calculated by using the results obtained with the Testing set at the end of the 6 iterations of the CV procedure, thus the reported values are average value of those obtained for each fold.

To evaluate the performance of the proposed techniques and to asses the uncertainty of the trained models for binary and multi-class classification, also with the dataset reduction, the Accuracy (ACC) and the Sensitivity (SENS) have been chosen. Let's define the True Positives (TP) and the True Negatives (TN) as actual positive and negative examples of the Testing set that were correctly classified as positive and negative. The FP are negative examples that were incorrectly labelled as positive and the False Negative (FN) are the positive examples marked as negative. The two metrics for binary classification

are defined as follow:

$$ACC_{bin} = \frac{TP + TN}{TP + TN + FP + TN} \qquad (7)$$

$$SENS_{bin} = \frac{TP}{TP + FN} \qquad (8)$$

To generalize these metrics to multi-class classification, it is possible to use the principle of One-Vs-All in order to obtain the accuracy and the sensitivity for each class, and then to compute a macro-averaging for the final metrics:

$$ACC_{multi} = \frac{ACC_1 + ... + ACC_k}{k} \qquad (9)$$

$$SENS_{multi} = \frac{SENS_1 + ... + SENS_k}{k} \qquad (10)$$

where k is the number of classes, and $[ACC_k; SENS_k]$ are the metrics of the kth class.

The metric used to evaluate the performance of the techniques proposed to reduce the FP, is the FPR defined as:

$$FPR = \frac{FP}{FP + TN} \qquad (11)$$

4.3. Binary Classification Results

The first set of results is related to the tests carried out with the three implemented algorithms for binary classification. The binary classification is the simplest classification used in this study, for this reason these outcomes will be used from now on as reference for assessing the performance of all the results shown in this Section.

17

In order to make easier the comparison between the results of the various tests, Tables 1-2 show the average values of the accuracy and sensitivity of the signals collected on all the ducts for each plant and algorithm.

Looking at the differences between the performance obtained for 50 mesh and 200 mesh, it is possible to observe that for Plant A, all the algorithms give similar results for both meshes. For Plant B, all the algorithms provide lower performance on 200 mesh than on 50 mesh. This disparity may be addressed to the higher degree of dispersion of PSD values for 200 mesh than for 50 mesh. A greater data diversity leads to more uncertainty during the classification process and thus a higher number of miss-classifications.

By comparing the results of different algorithms, none of them emerges as the best overall, but for all cases the difference between the best and the worst result is less than 2%. In detail, the ANN-based algorithm provides the best performance for Plant A 50 mesh with an average accuracy of 97.77% and a sensitivity of 97.34%, if compared with ACC=96.59%\SENS=96.06% of SVM and A=96,50%\S=96.09% of ELM; for Plant B the performance is very similar for all the algorithms, with ACC≃97% and SENS≃98%. The algorithm which uses the SVM achieves for both plants for the 200 mesh the best performance, providing an average accuracy of 97.13% and sensitivity of 98.74% for Plant A, while ANN and ELM obtain respectively the ACC=96.59%\SENS=98.46% and ACC=96.75%\SENS=98.29%, and an average accuracy of 94.02% and sensitivity of 98.86% for Plant B, higher than ACC=92.86%\SENS=95.73% and ACC=93.28%\SENS=96.54% respectively of ANN and ELM.

4.4. Multi-class Classification Results

In Tables 1-2 the results obtained with binary classification are compared with those obtained with 4 and 6-classes classification.

By analysing the averages for each plant, it is possible to appreciate a decreasing trend correlated with the increasing of the classes number, for both the considered meshes. As it is natural to expect, the introduction of more classes leads a greater uncertainty for the classification model and the accuracy get

Table 1: Binary Classification VS 4 and 6 classes classification for 50 mesh

%Dataset	SVM		ANN		ELM	
	ACC	SENS	ACC	SENS	ACC	SENS
	Plant A					
2 classes	96.59%	96.06%	**97.77%**	**97.34%**	96.50%	96.09%
4 classes	**94.47%**	**93.29%**	94.26%	93.23%	92.92%	90.54%
6 classes	**91.57%**	**89.10%**	92.26%	88.95%	90.54%	88.06%
	Plant B					
2 classes	97.51%	98.56%	97.34%	98.25%	**97.55%**	**98.57%**
4 classes	93.08%	**85.94%**	**93.37%**	84.45%	92.53%	84.18%
6 classes	**90.21%**	64.06%	88.72%	**70.29%**	88.58%	60.91%

worse. Despite this loss of performance, the average accuracy remain above the 90% with 4 classes for both plants in the case of 50 mesh. The sensitivity is above the 90% only for the Plant A, for the Plant B the values are around the 85%. For the 200 mesh only for Plant A all the algorithms provide an accuracy higher than 90% and sensitivity around the 90%, with Plant B data all the algorithms return lower accuracies and sensitivities.

By increasing the number of classes to 6, there is a further overall reduction of accuracy and sensitivity and in this case only the 50 mesh for Plant A, with all the algorithms, and for Plant B, with SVM, there is a mean accuracy higher than 90%. ELM and ANN for Plant B and 50 mesh and all the algorithms for both of the plants and 200 mesh return average accuracies between 80% and 90%. The sensitivities are around the 89% for Plant A and both meshes. For the Plant B the sensitivities are lower than Plant A, with values in the range

Table 2: Binary Classification VS 4 and 6 classes classification for 200 mesh

%Dataset	SVM		ANN		ELM	
	ACC	SENS	ACC	SENS	ACC	SENS
	Plant A					
2 classes	**97.13%**	**98.74%**	96.59%	98.46%	96.75%	98.29%
4 classes	**91.37%**	**90.86%**	91.21%	89.82%	90.95%	90.12%
6 classes	**89.07%**	**88.38%**	88.91%	86.00%	88.44%	87.77%
	Plant B					
2 classes	**94.02%**	**96.86%**	92.86%	95.73%	93.28%	96.54%
4 classes	**87.98%**	**74.54%**	85.30%	72.88%	86.03%	73.26%
6 classes	**82.70%**	**69.92%**	80.02%	65.40%	80.50%	67.84%

60%-70%.

Comparing the SVM, ANN and ELM models, the SVM-based algorithm achieve the best results for all the datasets with 200 mesh, and some of the dataset of 50 mesh. With 50 mesh for Plant A 2 classes the ANN provides the best results, for Plant B with 2 classes it is the ELM that get the best performance and with 4 classes the ANN is again the best one.

An important parameter that must be considered in this analysis, together with the mean accuracy, is the accuracy variation due to the classes increasing. For this reason, Table 3 reports the accuracy changes between the 2 and 6 class classifiers. By taking into consideration the tests for 50 mesh, it can be seen for both plants that the SVM based algorithm is the less affected by the increase of the classes, presenting an average accuracy reduction of 5.02% for Plant A and of 7.30% for Plant B, while ANN and ELM have a greater reduction in

Table 3: Accuracy variation between classification with 2 and 6 classes.

		PLANT A	PLANT B
50 mesh	SVM	**-5.02**%	**-7.30%**
	ANN	-5.48%	-8.62%
	ELM	-5.96%	-8.97%
200 mesh	SVM	-8.06%	**-11.32%**
	ANN	**-7.68%**	-12.84%
	ELM	-8.31%	-12.78%

both cases. For 200 mesh, ANN provides the best results for Plant A, with a variation of 7.68%, while SVM returns the minimum variation for Plant B with a difference of 11.32%. Comparing the performance variation for the two meshes, 50 mesh is less affected by the classes variation then 200 mesh, confirming as said in the Section 4.3 about the 50 mesh data and 200 mesh data.

4.5. Dataset Reduction Results

The reduction of the number of samples needed for the models training is a crucial aspect to consider during the design of a monitoring system based on machine learning techniques. Since reducing the number of training samples leads to the reduction of costs and development time. The tests carried out have assessed the variation of model accuracy and sensitivity on the Testing set in relation with the progressive reduction of the observations number used for training the models with the developed algorithms. The sensitivity provides an additional indication of the proportion of positives samples that are correctly identified as such. The studies were performed by using binary classification and three sub-sets of data, respectively with 90%, 75% and 50% of the whole dataset. The following tables report the values of mean accuracy and mean

Table 4: Dataset reduction for 50 mesh

%Dataset	SVM		ANN		ELM	
	ACC	SENS	ACC	SENS	ACC	SENS
	Plant A					
100%	96.59%	96.05%	**97.77%**	**97.34%**	96.50%	96.09%
90%	97.40%	96.71%	97.39%	96.61%	**97.50%**	**96.97%**
75%	84.61%	75.99%	79.83%	**87.37%**	**93.73%**	84.85%
50%	76.12%	80.57%	74.07%	81.86%	**81.63%**	**94.82%**
	Plant B					
100%	97.15%	98.23%	97.34%	98.25%	**97.55%**	**98.57%**
90%	**97.51%**	98.45%	97.02%	97.99%	97.38%	**98.51%**
75%	84.98%	**99.30%**	**85.46%**	98.00%	83.82%	98.68%
50%	79.68%	**89.94%**	**80.92%**	83.91%	80.42%	82.73%

sensitivity for both plants.

According to Table 4 referring to the 50 mesh data, it can be seen that with 90% of the dataset, for all cases, the accuracies do not decrease significantly and the performance are comparable with those obtained with 100% of the dataset. The performance starts to get worse by reducing to 75% and 50% of the samples. For both the target datasets, the accuracies go down under the 90%-threshold with Training sets extracted from 75% of the datasets, the only exception is the ELM-based algorithm with Plant A data. With 50% of the datasets, the decreasing continues and the accuracies reach values under 80% with SVM and ANN algorithms. The sensitivity roughly follows the accuracy

Table 5: Dataset reduction for 200 mesh

%Dataset	SVM		ANN		ELM	
	ACC	SENS	ACC	SENS	ACC	SENS
	Plant A					
100%	**97.13%**	**98.74%**	96.59%	98.46%	96.75%	98.29%
90%	**96.57%**	**98.28%**	96.13%	97.74%	96.37%	97.87%
75%	91.19%	**99.07%**	**92.44%**	98.03%	90.47%	98.42%
50%	81.72%	85.58%	81.55%	**92.76%**	**87.57%**	87.57%
	Plant B					
100%	**94.02%**	**96.86%**	92.86%	95.73%	93.28%	96.54%
90%	**94.86%**	**96.66%**	94.39%	95.89%	94.27%	96.14%
75%	**95.15%**	**97.12%**	94.98%	96.81%	94.85%	96.88%
50%	89.14%	88.45%	90.60%	**90.49%**	**91.40%**	87.68%

Table 6: Accuracy variation between classification with 100% and 50% of the dataset.

		PLANT A	PLANT B
50 mesh	SVM	-20.47%	-17.83%
	ANN	-23.70%	**-16.42%**
	ELM	**-14.87%**	-17.13%
200 mesh	SVM	-15.41%	-4.88%
	ANN	-15.04%	-2.26%
	ELM	**-9.18%**	**-1.88%**

trend, with the exception of SVM and ELM Plant A where it is possible to appreciate a sensitivity increment between the 75% and the 50% of the dataset.

Comparing the two plants, better performance is achieved with Plant B data, since the accuracy follows an uniform decreasing and reaches minimum value in the range of 80%, and the sensitivity maintains similar values for 100%, 90% and 75% of dataset, whereas it notably decreases with the 50%. With data from Plant A data the decreasing trend is less uniform and with the SVM and ANN, the accuracies reach values of 76.12% and 74.07% with 50% of the dataset. Also the sensitivity has this irregular behaviour, with a notably decrement since the 75% of the dataset.

Following the results show in Table 5, it is possible to see a similar trend for the two plants, both for accuracy and sensitivity. There is a uniform decreasing trend from the 100% to the 50% of the dataset. In terms of values, the accuracy does not go down the 80%-threshold, while the sensitivity has values greater than 85%.

Table 6 summarizes the variations of accuracy going from 100% to 50% of the dataset. Observing the case of 50 mesh, it can be noted as for Plant A, the ELM-based algorithm returns the smallest variation in the accuracy on the

Testing set, with a difference of 14.87%. For Plant B, instead the ANN returns the smallest difference of 16.42%. The 200 mesh data are subjected to a lower reduction than the 50 mesh data, and, for both systems, the ELM returns the least significant reductions, which are -9.18% for Plant A and -1.88% for Plant B. Comparing the performance among the three algorithms, the ELM outperforms the others for 3 out of 4 cases, and therefore it demonstrates to be less affected by the dataset variation.

4.6. False Positive Reduction Results

In the Figures 5, 6, 8 and 7 the FPRs are reported for each duct (5 ducts for Plant A and 8 ducts for Plant B) and three graph compared the values of FPR calculated with Eq.(11) for the original classifier (STD), and the proposed DTT. For sake of simplicity of exposition, the average FPR for each plant is used as metric to evaluate the performance in the comments.

The Figure 5 shows that the proposed technique succeeds to reduce the false positives. The SVM-based algorithm obtained an average FPR of 3.27% with STD and 0.63% with DTT. The original ANN classifier achieves the average FPR of 2.26%, while DTT the 0.97%. Also with ELM-based algorithm, DTT outperforms STD, 1.20% against the 3.40%.

As it is represented in the Figure 6, for the data of 50 mesh, Plant B, SVM classifier achieves an average FPR of 6.74%, with the addition of DTT a FPR of 2.29%. ANN classifier allows to obtain an average FPR of 5.96%, this value decrease at 2.02% with DTT. With ELM the behaviour of the proposed technique is the same, it provides better results then STD, an average FPR of 1.96% for DTT and 6.40% for the STD.

Also with the 200 mesh data the employment DTT improves the performance of STD. In the Figure 7 the results of Plant A are reported. The average FPR with original SVM classifier is 8.79%, the DTT provide the best result with the 3.10%. ANN results are 9.46% for STD, and 3.91% for the DTT. With ELM, DTT has again the best result of 6.32%, against the 11.18% without any technique.

The Figure 8 shows the results for Plant B and 200 mesh. The best average FPR with the SVM is obtained with the DTT (4.83%), whereas the STD achieves worst FPR of 19.82%. The ANN-based algorithm achieves the lowest FPR of 10.03% with DTT, and a value of 21.37% with STD.
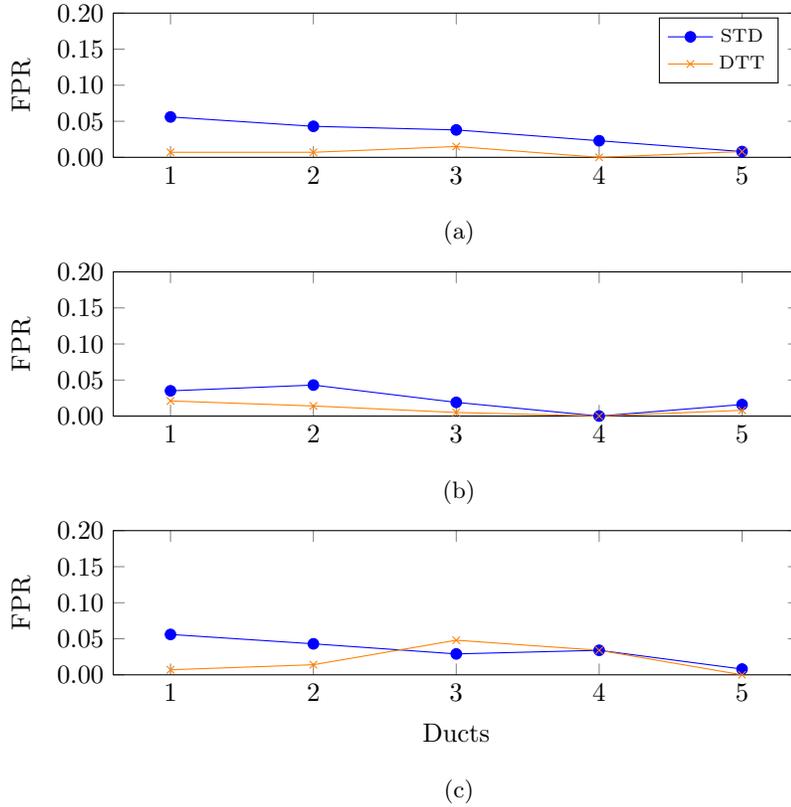

Figure 5: FPRs comparison for data 50 mesh, Plant A: (a) SVM, (b) ANN, (c) ELM.

Comparing the averages for the two meshes, for the 50 mesh the proposed technique succeeds to reduce the level of FP and the resulting FPR is under the 3%, with the data of 200 mesh the technique still reduces FPR with values under the 20%. This performance difference is due to the same reason highlighted in the Section 4.3, the 200 mesh data have more variability in the output space and this do not allow to find an optimal threshold to obtain a greater FP reduction.
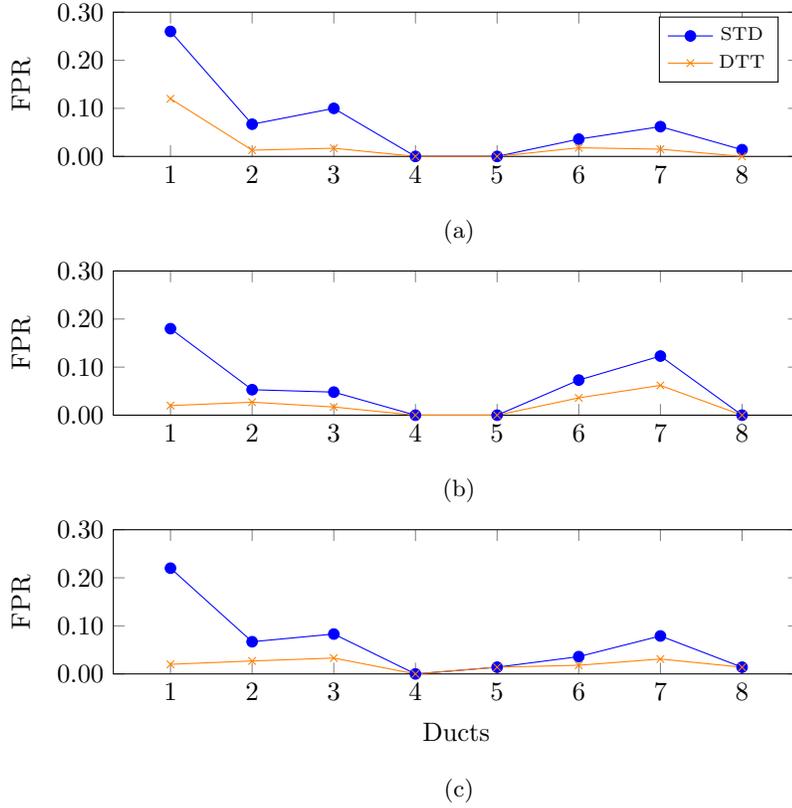
Figure 6: FPRs comparison for data 50 mesh, Plant B: (a) SVM, (b) ANN, (c) ELM.

5. Conclusion

This paper presents a new scheme for a non-invasive condition monitoring approach based on classification for burners devoted to coal powder combustion in an industrial power plant. The proposed scheme used PSD to monitor the boiler burner operating condition with the purpose of holding the optimal powder particle size that ensures the maximum combustion efficiency. The information about the PSD was extracted by means of the AE signals generated by the impact of the powder on the inner surface of the duct. The data used for the study were collected on two different power plants and describe the granulometry of pulverized coal. Two sizes for the PSD have been taken into account: 50 mesh and 200 mesh. The data have been labelled with appropriate threshold
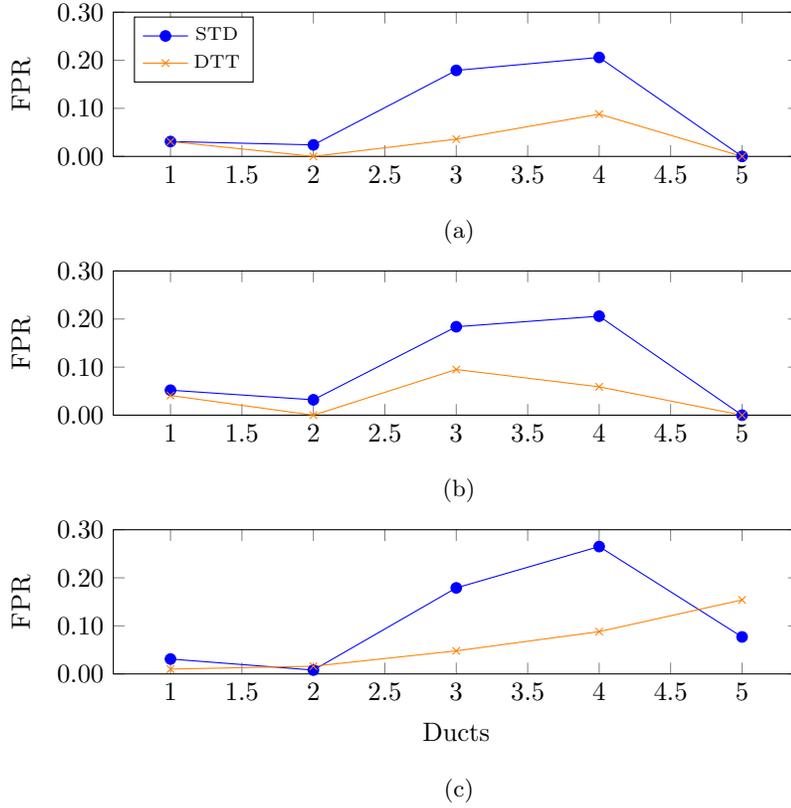
Figure 7: FPRs comparison for data 200 mesh, Plant A: (a) SVM, (b) ANN, (c) ELM.

system in order to be separated into different classes distinguished by the particle size. Three classification algorithms have been implemented and compared: SVM, ANN and ELM. Three key issues have been addressed in this study: the number of classes used to divide the output space, the number of samples used for training and the reduction of FP.

A series of tests was performed with three sets of classes, containing respectively 2, 4 and 6 classes, to evaluate the variation of accuracy depending on the number of classes. As it is easy to predict, the results demonstrated that the binary classification is the most accurate, but at the same time such results point out the capability of using multiple classes to improve the condition monitoring capabilities by accepting the accuracy reduction as a compromise.
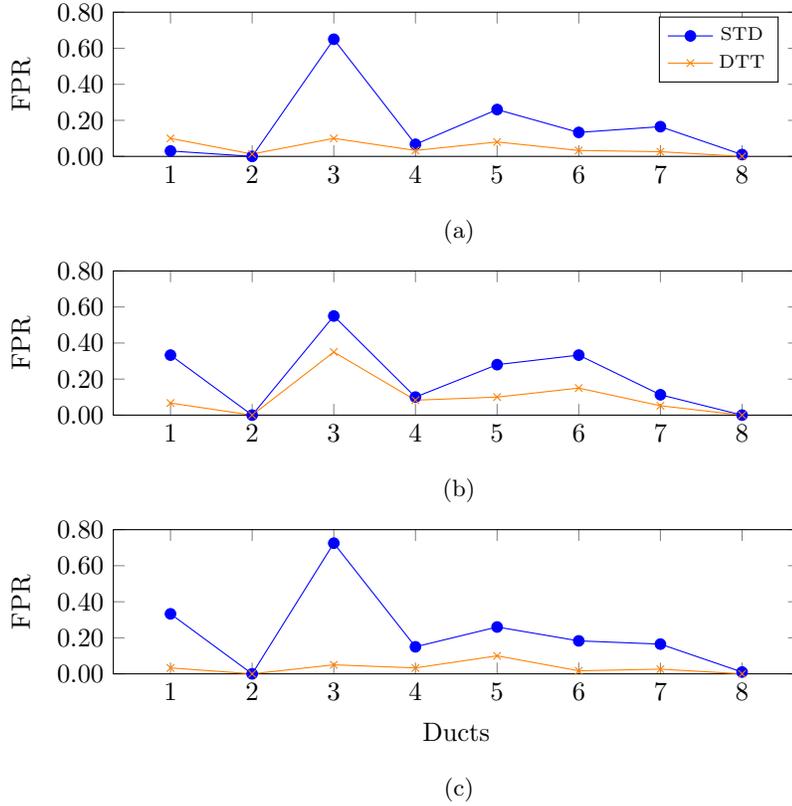
Figure 8: FPRs comparison for data 200 mesh, Plant B: (a) SVM, (b) ANN, (c) ELM.

It was shown that a decrease of the number of samples used for the training causes a decrease of performance of algorithms. Therefore a 10% data reduction is reasonable and in that case the accuracies don't decrease significantly. A major reduction of 25% is possible but it must be accepted that the accuracies are lower than 90%. The results obtained using a technique for the false positive reduction showed that for the problem of granulometry classification it is not possible eliminate all the occurrences of false positives, but it is possible to reduce their level.

In conclusion, the proposed work confirms that classification can be used for on-line monitoring of the pulverized coal particle size. Indeed, the use of classification algorithms allowed a reduction of the number of samples required the

training by 10%. Moreover, a control over the false positive and their reduction can be achieved. There are no apparent priorities of the three classification algorithms, although in average, SVM performs better for more classes, and ELM performs better with significant reduced datasets in this case study.

Future works will be targeted to the application of the proposed strategies to other type of industrial plants, in which the estimation of the powder size is used for monitoring purposes, and AE data exploited to train the expert systems. Moreover, different machine learning techniques will be also employed in substitution of those addressed in this paper, either with supervised and unsupervised learning, in order to increase the accuracy of classification. There is also the intention of investigating the use of fuzzy logic to improve the discrimination between two contiguous classes.

References

[1] M. Leach, G. Rubin, J. Williams, Particle size determination from acoustic emissions, Powder Technology 16 (2) (1977) 153 – 158.

[2] M. Leach, G. Rubin, J. Williams, Particle size distribution characterization from acoustic emissions, Powder Technology 19 (2) (1978) 157 – 167.

[3] A. Boschetto, F. Quadrini, Powder size measurement by acoustic emission, Measurement 44 (1) (2011) 290 – 297.

[4] M. Uher, P. Benes, Measurement of particle size distribution by the use of acoustic emission method, in: Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International, 2012, pp. 1194–1198. `doi:10.1109/I2MTC.2012.6229375`.

[5] Y. Hu, X. Huang, X. Qian, L. Gao, Y. Yan, Online particle size measurement through acoustic emission detection and signal analysis, in: Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2014 IEEE International, 2014, pp. 949–953. `doi:10.1109/I2MTC.2014.6860883`.

[6] D. Rossetti, S. Squartini, S. Collura, Machine learning techniques for the estimation of particle size distribution in industrial plants, in: 25th Italian Workshop on Neural Networks, Vietri Sul Mare, Italy, 2015.

[7] K. Nurzynska, M. Kubo, K. i. Muramoto, Snow particle automatic classification with texture operators, in: 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011, pp. 2892–2895. `doi:10.1109/IGARSS.2011.6049819`.

[8] K. K. Pahalawatta, R. Green, Particle detection and classification in photoelectric smoke detectors using image histogram features, in: 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2013, pp. 1–8. `doi:10.1109/DICTA.2013.6691509`.

[9] F. Previdi, S. M. Savaresi, P. Guazzoni, L. Zetta, Light charged particle classification using subspace identification methods and neural networks, in: 2006 American Control Conference, 2006, pp. 6 pp.–. `doi:10.1109/ACC.2006.1657190`.

[10] C. Yibo, X. Xiaopeng, Wear particles classification based on wavelet transform and back-propagation neural network, in: The Proceedings of the Multiconference on "Computational Engineering in Systems Applications", Vol. 2, 2006, pp. 2010–2014. `doi:10.1109/CESA.2006.4281969`.

[11] M. Dumont, S. Collura, T. Urbank, D. Rossetti, Optimized atex acoustic emission measuring chains for particle size monitoring in industrial plants, in: 19th World Conference on Non-Destructive Testing, 2016.

[12] D. Rossetti1, S. Squartini, S. Collura, Y. Zhang, Exploiting heterogeneous data for the estimation of particles size distribution in industrial plant, in: The 17th International Conference on Mechatronics Mechatronika, 2016.

[13] S. Collura, D. Possanzini, M. Gualerci, L. Bonelli, D. Pestonesi, Coal mill performances optimization through non-invasive online coal fineness monitoring, in: Powergen, Wien, 2013.

[14] M. Stphane, Chapter 8 - wavelet packet and local cosine bases, in: M. Stphane (Ed.), A Wavelet Tour of Signal Processing, third edition Edition, Academic Press, Boston, 2009, pp. 377 – 434. `doi:http://dx.doi.org/10.1016/B978-0-12-374370-1.00012-4`.

[15] A. Bastari, C. Cristalli, R. Morlacchi, E. Pomponi, Acoustic emissions for particle sizing of powders through signal processing techniques, Mechanical Systems and Signal Processing 25 (3) (2011) 901 – 916. `doi:http://dx.doi.org/10.1016/j.ymssp.2010.08.019`.

[16] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd Edition, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.

[17] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Neurocomputing: Foundations of research, MIT Press, Cambridge, MA, USA, 1988, Ch. Learning Representations by Back-propagating Errors, pp. 696–699.

[18] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[19] C. Bishop, Pattern Recognition and Machine Learning, 1st Edition, Springer-Verlag New York, 2006.

[20] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27.

[21] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, Vol. 2, 2004, pp. 985–990 vol.2. `doi:10.1109/IJCNN.2004.1380068`.

[22] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (2) (2012) 513–529. `doi:10.1109/TSMCB.2011.2168604`.

[23] D. Breitgand, M. Goldstein, E. Henis, O. Shehory, Performance manage-ment via adaptive thresholds with separate control of false positive and false negative errors, in: 2009 IFIP/IEEE International Symposium on Integrated Network Management, 2009, pp. 195–202. `doi:10.1109/INM.2009.5188810`.

[24] S. Yusuf, D. J. Brown, A. Mackinnon, R. Papanicolaou, Fault classification improvement in industrial condition monitoring via hidden markov models and naive bayesian modeling, in: 2013 IEEE Symposium on Industrial Elec-tronics Applications, 2013, pp. 75–80. `doi:10.1109/ISIEA.2013.6738971`.

[25] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Snchez, B. van Ginneken, Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks, IEEE Transactions on Medical Imaging 35 (5) (2016) 1160–1169. `doi:10.1109/TMI.2016.2536809`.