Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Biomedica, Elettronica e delle Telecomunicazioni

Advanced Machine Learning Techniques for Condition Monitoring in Industrial Engineering Applications

Ph.D. Dissertation of:
Damiano Rossetti

Advisor:
Prof. Stefano Squartini

Coadvisor:
Dott. Stefano Collura

Curriculum Supervisor:
Prof. Franco Chiaraluce

XVI edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Biomedica, Elettronica e delle Telecomunicazioni

Advanced Machine Learning Techniques for Condition Monitoring in Industrial Engineering Applications

Ph.D. Dissertation of:
Damiano Rossetti

Advisor:
Prof. Stefano Squartini

Coadvisor:
Dott. Stefano Collura

Curriculum Supervisor:
Prof. Franco Chiaraluce

XVI edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Facoltà di Ingegneria
Via Brecce Bianche – 60131 Ancona (AN), Italy

Abstract

Over recent years, the improvement of Computational Intelligence tools has led to the development of industrial techniques for monitoring conditions both for processes and single devices. Each monitored system has its peculiar characteristics, so each monitoring solution must be adapted to the specifics of the case. The present thesis forms part of this scenario with the aim of proposing innovative solutions for condition monitoring of industrial systems by exploiting Machine Learning techniques. The flexibility and versatility of these techniques make it possible to develop customized solutions for each problem. In this work, different case studies coming from industrial contexts have been addressed and for each one the most appropriate Machine Learning tools have been selected, with the purpose of developing effective solutions for the specific condition monitoring problem. The proposed solutions want to represent an improvement from those currently employed in the industrial procedure by applying Machine Learning techniques in new industrial contexts.

Firstly, a new system for monitoring the operating conditions of pulverized coal burners employed in a coal fired power plant is developed with the aim to determine if the whole combustion system is working properly under the most energy efficient conditions. This part of the work begins with the investigation of Machine Learning techniques to provide a solution for non-invasive estimation of the coal powder particle size. The particle size estimation forms the foundation of the proposed method. Further research exploited heterogeneous data from similar but not same application, thus delivering more accurate modelling. With the aim of improving the implementability of the method by reducing the number of observations needed, the problem of particle size estimation has been reviewed in the form of a classification problem, therefore the previously implemented Machine Learning algorithms have been appropriately modified and adapted to solve a classification problem. Finally, the classification-based algorithms have been used to develop a complete monitoring system.

Fault detection for roller bearings is another industrial relevant problem that has been tried to solve using a machine learning approach similar to the previous case. In particular a classifier based on Support Vector Machine has been developed to detect and classify different types of faults. Real experimental data has allowed development of a fault detection system that exploits the

time-frequency analysis and image analysis, and the obtained results proved its suitability for both stationary and non-stationary operating conditions of roller bearings.

The last case study proposed is a novel unsupervised method to detect faults in DC electric motors during quality control at the end of the production line. The proposed scheme uses a novelty detection approach consisting in using of a Denoising Autoencoder network for modelling the normality condition of the motors without faults, with the purpose of distinguish from this background the motors with faults that do not respect the quality requirements.

The results obtained with tests on experimental data demonstrate the validity of the solutions developed for each addressed case study, applicable for the implementation of monitoring systems for industrial processes and devices devoted to quality check and equipment maintenance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is intended to study new applications for Machine Learning techniques developed over the years in order to design for specific industrial problems effective solutions that can represent an improvement on the existing standard procedures.

In the past decades, the industrial processes have been subjected to a rapid evolution, with new standard requirements to meet in order to ensure the highest quality level to the final user. At the beginning of the industrial era, during the so called first industrial revolution, the goal for every system designer was to develop systems able to make products at the lowest cost, however over the years, new fundamental aspects have emerged. These include workers safety, quality of the products and energy consumed. Nowadays in the fourth industrial revolution, with the increasing demand of production quality, system performance, economic and energetic requirement, the industrial processes are becoming more complicated in both structure and degrees of automation. The reliability and safety issues on complex industrial processes have became more critical for system design. In response to these new tasks, the attention of both industrial and academic research has started to focus on two important aspects for providing valid solutions to the raised issues: the maintenance process and the quality control.

The maintenance process aims to preserve the original features of a system for its all life cycle. Many strategies can be adopted for maintenance purpose, and among them the Preventive Maintenance (PM) [2, 3] strategies are the most efficient to reduce the failure rate of the equipment because they involve the performance of maintenance activities prior to the failure. In the last years a new strategy of PM based on Condition-based Maintenance (CBM) [4, 5] is emerging aimed to maximise the effectiveness of PM decision making by recommending maintenance operations based on the information collected through condition monitoring process. In CBM, the equipment status is monitored through its operating condition, which can be measured based on various monitoring parameters, such as vibration, temperature, pressure, and noise levels. The CBM goal is to detect the signs that precede device failures or

breakdowns and to alert for the incoming fault. The foundation of CBM is the Condition Monitoring (CM) process, that analyses signals from sensors installed on the working component. In this way, CBM performs a real-time assessment of equipment conditions making smart decisions about maintenance activity, consequently reducing unnecessary maintenance and related costs.

If, from one side, maintenance process aspire to conserve the system characteristics during its working life, from the other side quality control has the purpose of assuring that since the production these characteristics are within a strict specification range. Quality control is a process by which entities review the quality of all factors involved in production, and it is a part of quality management focused on fulfilling quality requirements. In other words, quality control is a continuous check, along or at the end of production line, aimed to monitor the conditions of produced products

The CM is then a cross aspect for maintenance and quality assurance, and a complete CM system can provide a useful solution to both issues. Many works presented in literature have proposed the CM as adequate method for process monitoring in order to improve reliability, quality, and productivity in industry [6, 7, 8]. CM techniques are continually evolving by embedding latest technological advances and computer technology is supporting the development of both hardware and software. Moreover, under the push of Industry 4.0 more and more industrial systems have been embedding complex monitoring and metering solutions that provide a great amount of experimental data. Recent developed methods have been exploiting the new data availability to train expert system based on Machine Learning (ML) [9] techniques and use them as tools for improving the performance of monitoring systems. These techniques develop an intelligent monitoring system, which aims to provide a conditions and quality control similar to human decision making.

1.1 Literature Survey

ML techniques have been widely exploited by many contributions to deal with condition monitoring problems. For this reason, an overview on the recent researches proposed in literature is fundamental to understand the context in which the industrial solutions proposed in this work have been developed. In the following, a survey on the most popular ML technique employed for the condition monitoring is reported.

Among the solution for industrial CM systems, the Artificial Neural Networks (ANN) is undoubtedly one of those techniques that have raised a strong interest.

Ghate and Dudul [10] developed a simple and economical fault detection scheme for small and medium sizes three-phase induction motors. The proposed method uses a cascade connection of multilayer perceptrons ANN with

Radial Basis Function as activation function. In this scheme, stator current is captured, and simple statistical parameters of current waveform are used as network inputs to detect the conditions of motor. Also other works dealt with the fault detection problem for induction motors and provided valid solutions with ANN-based algorithms [11, 12].

An evolution of classic ANN was used by Jia et al. [13] to design a new method for intelligent diagnosis of rotating machinery. Through deep learning, deep neural networks (DNNs) with deep architectures, could be established to mine the useful information from raw data and approximate complex non-linear functions. Based on DNNs, a novel intelligent method is proposed and its effectiveness is validated using datasets from rolling element bearings and planetary gearboxes.

In a completely different field of application, Fast and Palmé [14] created an online system for condition monitoring and diagnosis of a combined heat and power plant. The system in question consisted of ANN models, representing each main component of the combined heat and power plant, connected to a graphical user interface. The ANN models were integrated on a power generation information manager server in the computer system of the combined heat and power plant. The proposed method obtain accurate predictions from the ANN models on experimental data from a working power plant.

An ANN based method is developed by Tian [15] for achieving more accurate remaining useful life prediction of equipment subject to condition monitoring. The ANN model takes the age and multiple condition monitoring measurement values at the present and previous inspection points as the inputs, and the life percentage as the output. A validation mechanism is introduced in the ANN training process to improve the prediction performance of the ANN model. The proposed ANN method is validated using real-world vibration monitoring data collected from pump bearings in the field.

The Support Vector Machine (SVM) is another CI technique that gained its relevance among the solution for CM [16]. Particularly, the One Class SVM (OC-SVM) has been used by many works to provide an unsupervised method for CM and fault diagnosis.

Yang et al. [17] performed condition classification of small reciprocating compressor for refrigerator using SVM. In this paper, wavelet transform and statistical method were used to extract salient features from row noise and vibration signal. Moreover, iteration method was employed to select the proper RBF kernel parameters in SVM.

Another research was conducted by Han et al. [18] for hot spot detection in power plant boiler air pre-heater based on Least Squares SVM. In this system, discriminate models of 3 pairs of fire status have been built based on Least Squares SVM using RBF kernel and polynomial kernel. The hyper parameters

of classifiers were tuned by leave-one-out cross validation. Receiver operating characteristic curve showed that Least Squares SVM has good classification and generalization ability.

In another research concerning CM for rotating machines, Jack and Nandi [19] performed fault detection of roller bearing using SVM and ANN. They used vibration data taken from small test rig and simulate the bearing condition, which has four faults: inner race fault, outer race fault, rolling element fault and cage fault. They defined and calculated statistical features based on moments and cumulants and selected the optimal features using genetic algorithm. In the classification process, they employed SVM using RBF kernel with constant kernel parameter.

Above the well-know ANN and SVM, others techniques have been successfully employed for the monitoring purpose.

The Canonical Variant Analysis has been used by Ruiz-Cárcel et al. [8] to design a method that merges process data and vibration features to improve the detectability of mechanical faults in systems working under varying operational conditions. Their results show how CVA is able to detect changes in selected vibration, current and pressure features during the test, distinguishing deviations from normal operation even under operational conditions that were not considered in the training period.

The article proposed by Kruger et al. [20] presents an extension to the standard Partial Least Squares (PLS) algorithm, referred to as Extended-PLS, which leads to the definition of new PLS scores, denoted as generalized scores. This approach is capable of detecting abnormal process conditions that manifest themselves in process response variables as well as predictor variables, even when no process feedback is present.

Allen et al. [21] presented a novel method of monitoring the health of heating ventilating and air conditioning equipment. Fuzzy Logic techniques were demonstrated to monitor the health of a variable air volume unit and to create a fault signature for each individual fault type. A health monitor fault classification neural network was also demonstrated to recognize and classify the fault signature.

Two computational intelligence algorithms, i.e. Particle Swarm Optimization and Bacterial Foraging Optimization, have been used to detect a developing induction motor stator winding fault by Ethni et al. [22]. The condition monitoring method is based on the comparison of measured machine stator currents with those obtained from a machine mathematical model, and then using the algorithms to minimise the resulting error function. The two algorithms have been shown to be effective in determining the winding fault type and location.

A different method that applies many ML techniques for condition monitoring and fault detection is the Novelty Detection approach. During the past

years, many research works have proposed application for this approach to address different industrial monitoring problems.

Carino et al. [23] presented a condition-based monitoring methodology based on novelty detection applied to industrial machinery. The proposed approach includes both the classical classification of multiple a priori known scenarios, and the detection capability of new operating modes not previously available. The posterior combination of a feed-forward ANN and OC-SVM machine allows the proper interpretation of known and unknown operating conditions.

The Gaussian Mixture Model (GMM) was used by Filev et al. [24] to present a practical framework for autonomous monitoring of industrial equipment based on novelty detection. The approach utilizes unsupervised learning techniques to continuously learn the necessary parameters needed to identify a significant change in the pattern of monitored features. The input to the evolving novelty detection model is a feature vector that characterizes the status of monitored equipment. The model of the feature space is essentially a set of evolving GMM which are dynamically updated as new feature vectors become available through continuous monitoring. This approach proved its efficacy on experimental data of two accelerated bearing failure.

Wang and Jiao [25] a non-linear quality-related fault detection approach is proposed based on Kernel Least Squares (KLS) model. The proposed method uses a non-linear projection function to map original process variables into feature space in which the correlation between feature and output matrices is realized by means of KLS. Then, the feature matrix is decomposed into two orthogonal parts by singular value decomposition and the statistics for each part are determined appropriately for the purpose of quality-related fault detection.

The SVM is used also by Clifton and al. [26] in a method that calibrates the novelty scores output by the one-class SVM into estimated posterior class probabilities, where special care was required due to the one-class formulation, utilises the probabilistic nature of the result to define a novelty threshold without the need for the conventional validation set, and proposes a procedure for determining other SVM parameters.

In addition to the present survey, in each chapter that expose one of the industrial issues addressed in this thesis, a more detailed literature is reported in a dedicated paragraph.

1.2 Thesis Outline and Contribution

Although there are common issues for the various scenarios, each monitoring solution needs to be customized for the specific analysed system by taking into account the available resources for the monitoring. Each case brings with it

different problems due to the working environment of the system, the operating conditions, and the time under work. In other words, each condition monitoring solution must be able to use the analysis tools made available by the technology and to make them fit the particular problem dealt with.

As aforementioned, the works presented in this thesis aim to propose new applications of ML tools to solve specific industrial issues. The starting point of each proposed study has been a problem raised in a particular industrial field, for which there was the necessity to find a more effective solution than those already available. The solution development explained in this dissertation is based on the available experimental data for the process under examination and uses signal processing tools and ML techniques to implement new data-driven monitoring methods that could provide the information needed for the control of these processes. The outline of the thesis is the following.

An overview of the Machine Learning techniques adopted for the condition monitoring is presented in Chapter 2 concerning the following methods: Artificial Neural Networks, Autoencoders (AE), Support Vector Machines, Support Vector Regression (SVR), One-Class Support Vector Machines, and Extreme Learning Machines (ELM).

One of the proposed monitoring solution is based on the case study of a monitoring system, named POWdER, which performs the estimation of the coal powder size in a non-invasive way. The data collected by this system have been used to lead on the research activities concerning this solution, and in Chapter 3 an overview of this system is proposed together with the description of the datasets employed for the studies.

The development of an effective method for condition monitoring based on particle size analysis has started from the study of the most suitable ML techniques for this task. In Chapter 4 SVR, ANN and ELM will be used and compared for the estimation of particle size. The available data for the study are limited in quantity, for this reason a solution for the use of data acquired from different sources will be presented.

The previously proposed algorithms have been modified and used to perform the particle size classification, in order to provide a solution that does not need reference targets for the model training. In Chapter 5 a condition monitoring approach suitable for coal fired power plant is proposed. This approach use the particle size classification as an indication of the operating state of a monitored boiler burner where the combustion of the powdered coal takes place. The results shows that the powder size classification is suitable to face the monitoring task for a boiler burner.

Some experimental tests have been conducted to explore a different context for the application of ML techniques for monitoring. A common problem for many industrial devices are connected with the faults related to basic device

components, therefore a fault detection method for rolling bearing is reported in Chapter 6. A SVM- based algorithm is proposed to provide a fault detection and classification for both stationary and non-stationary working conditions.

A novelty detection based approach for the identification of faulty DC motor during the en-line check is presented in Chapter 7, Autoencoder ANN and One-Class SVM, are exploited to characterize the normality condition, or background. The validity of the approach is confirmed by the achieved results, proving its suitability for the end-line quality control. The two algorithms are compared and the Autoencoder emerged as the best one.

The main findings of this research related to the case studies are presented in Chapter 8.

Chapter 2

Background

2.1 Neural Networks

In order to mimic the brain capabilities a first model of artificial neuron, namely Threshold Logic Unit (TLU), has been proposed by McCulloch and Pitts *et al.* [27]. After many years, inspired by the biological neuron, Figure 2.1, the original idea has been improved to accomplish the nowadays artificial neuron model illustrated in Figure 2.2.

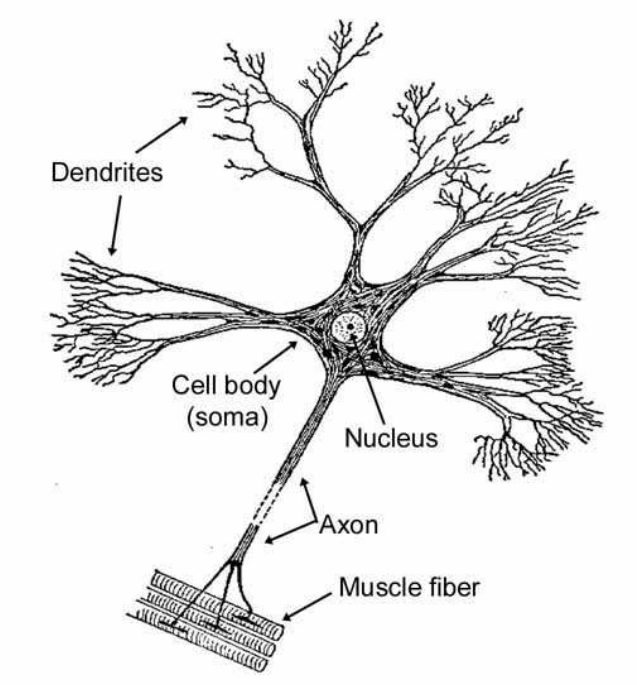


Figure 2.1: Biological multipolar neuron.

As the biological neuron, the artificial one presents the same fundamental parts:

- dendrites: propagate the electrochemical stimulation received from other neural cells to the cell body, or soma;
- cell body (soma): sum all the received electrochemical stimulations;
- axon: reached a certain potential conducts the electrical impulses away from the neuron's cell body.

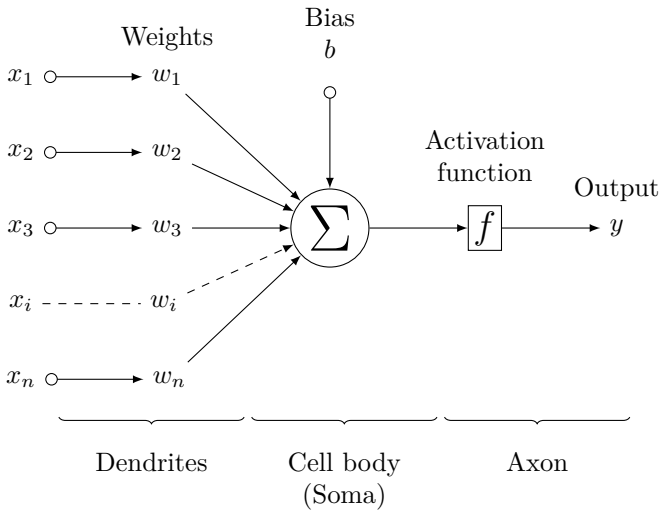


Figure 2.2: Artificial neuron.

Therefore, considering the general representation of an artificial neuron, Figure 2.2, with N_c input connection, its output state, y, is given as:

$$y = f\left(\sum_{j=1}^{N_c} \mathbf{W}_j \cdot \mathbf{X}_j - b\right), \tag{2.1}$$

assuming $W_0 = -b$ and $x_0 = 1$, is obtained the generic formulation below: N_c input connection, its output state, y, is given as:

$$y = f\left(\sum_{j=0}^{N_c} \mathbf{W}_j \cdot \mathbf{X}_j\right), \tag{2.2}$$

where $\mathbf{W}$ and $\mathbf{X}$ are the weight and input vectors, respectively, and f is the *activation function* (e.g., linear, sigmoid, etc.).

In a typical Neural Network (NN), without feedback connections (feedforward), the neurons are organized in layers, characterized by a fixed number of parallel neurons. Specifically, in each layer the neurons share the same inputs,

but not the weights, and these are connected with the outputs of the previous layer neurons. In the first layer, the inputs correspond to the system inputs. In the same way, for the last layer the number of outputs, thus the neurons number, is set by the system outputs. Therefore, 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 for the k-th neuron in the l-th layer, $y_k^{(l)}$, is obtained reformulating the equation (2.2) in:

$$y_k^{(l)} = f\left(\sum_{j=0}^{N_{l-1}} \mathbf{W}_{kj}^{(l)} \cdot \mathbf{X}_j^{(l)}\right), \tag{2.3}$$

where $\mathbf{W}$ and $\mathbf{X}$ respectively become the weights and the inputs matrices. 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:

$$\mathbf{x} = \left[\mathbf{X}_0^{(1)}, \ldots, \mathbf{X}_{N_1}^{(1)}\right], \qquad \mathbf{y} = \left[y_0^{(M)}, \ldots, y_{N_M}^{(M)}\right], \tag{2.4}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the input and output data vectors, respectively.

Different activation functions can be chosen for the neurons: hyperbolic tangent, unipolar and bipolar sigmoid, and a set of radial basis functions. In the experiments, the unipolar sigmoid function has been used:

$$\frac{1}{1 + e^{-x}}, \tag{2.5}$$

During the training process, the weights $\mathbf{W}$ are updated with the Backpropagation [28] algorithm. The standard structure selected for the tests uses one input layer, two hidden layers and one output layer. For each test, in order to select the best configuration, the number of hidden layers nodes have been varied, from 50 to 100 for the first layer and from 30 to 80.

The analysis with the ANN have been performed by using a proper implementation of the algorithm running on MATLAB®.

2.1.1 Autoencoders and Denoising Autoencoders

Generally speaking, an autoencoder [29] is any neural network trained to get target values equal to input ones: $\mathbf{x} = \mathbf{y}$. Specifically, given a generic input $\mathbf{x} \in \mathbb{R}^n$, the goal of the autoencoder is to first map the input with an encoder

to an hidden representation $\mathbf{z} \in \mathbb{R}^m$, as:

$$\mathbf{z} = f(\mathbf{W_1} \cdot \mathbf{x} + \mathbf{b_1}), \tag{2.6}$$

where f is a non-linear activation function (e.g. tanh or sigmoid), $\mathbf{W_1}$ the weight matrix, and $\mathbf{b_1}$ the bias vector. This level of representation, $\mathbf{y}$ is then mapped back into a reconstruction $\mathbf{y}$ of the same shape of $\mathbf{x}$, as:

$$\mathbf{y} = f(\mathbf{W_2} \cdot \mathbf{z} + \mathbf{b_2}), \tag{2.7}$$

where $\mathbf{W_2}$ and $\mathbf{b_2}$ denote the weight matrix and the bias vector of the reverse mapping, respectively. $\mathbf{y}$ is seen as a prediction of $\mathbf{x}$, give the code $\mathbf{z}$.

Given an input set of examples $\mathbf{X}$ autoencoder training consists in finding parameters $\theta = \{\mathbf{W_1}, \mathbf{W_2}, \mathbf{b_1}, \mathbf{b_2}\}$ that minimise the reconstruction error, which corresponds to minimising the following objective function:

$$O(\theta) = \sum_{\mathbf{x} \in \mathbf{X}} \|\mathbf{x} - \mathbf{y}\|^2, \tag{2.8}$$

The minimization is usually realised by stochastic gradient descent or Adam algorithm [30].

Different type of configurations can be obtained by varying the size and the dimension of the hidden representation, thus layers and neurons number. Among the state-of-the-art applications, the Denoising Autoencoder (DAE) [31] and the Compression Autoencoder (CAE) are ones of the most adopted configurations.

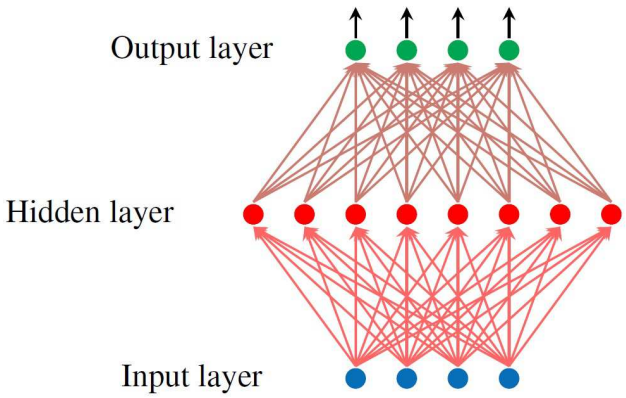


Figure 2.3: Structure of the Denoising Autoencoder [1].

Denoising autoencoders are characterized by the use of a corrupted training sequence, in order to force the hidden layer to learn robust features and prevent it from learning the identity. Typically, a DAE is composed of multiple hidden

layers that have a number of neurons greater than the input one, as the Figure 2.3 shows. The initial input $\mathbf{x}$ is corrupted by means of additive isotropic Gaussian noise in order to obtain: $\mathbf{x}'|\mathbf{x} \sim N(\mathbf{x}, \sigma^2 I)$. The corrupted input $\mathbf{x}'$ is then mapped, as with the basic autoencoder, to a hidden representation:

$$\mathbf{z}' = f(\mathbf{W_1}' \cdot \mathbf{x}' + \mathbf{b_1}'), \tag{2.9}$$

from which we reconstruct a the original signal as follows:

$$\mathbf{y}' = f(\mathbf{W_2}' \cdot \mathbf{z}' + \mathbf{b_2}'), \tag{2.10}$$

The parameters θ' are trained to minimise the average reconstruction error over the training set, to have $\mathbf{y}'$ as close as possible to the uncorrupted input $\mathbf{x}$ which corresponds to minimising the objective function in (2.8).

The tests with DAE have been carried out by using the Keras API [32], written in Python and running on Theano.

2.2 Support Vector Machines

Support Vector Machines [33] are binary classifiers that discriminate whether an input vector $\mathbf{x}$ belongs to class $+1$ or to class -1 based on the following discriminant function:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i t_i K(\mathbf{x}_i, \mathbf{x}) + d, \tag{2.11}$$

where $t_i \in \{+1, -1\}$, $\alpha_i > 0$ and $\sum_{i=1}^{N} \alpha_i t_i = 0$. The terms $\mathbf{x}_i$ are the "support vectors" and d is a bias term that together with the α_i are determined during the training process of the SVM. The input vector $\mathbf{x}$ is classified as $+1$ if $f(\mathbf{x}) \geq 0$ and -1 if $f(\mathbf{x}) < 0$, given by the separation hyperplane, as depicted in Figure 2.4, defined by the "support vectors".

The kernel function $K(\cdot, \cdot)$ can assume different forms [34], common ones are the linear, the polynomial, and the Gaussian. Specifically, the linear one is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i{}^T \mathbf{x}_j + c \ , \tag{2.12}$$

that gives the inner product between $\mathbf{x}_i$ and $\mathbf{x}_j$ plus an optional constant c. This is the simplest kernel function that provide a separation hyperplane only for linearly separable vectors. By considering the case of non-separable data, a non-linear kernel is needing to provide an optimal hyperplane. The Gaussian kernel is an example of Radial Basis Function (RBF) kernel, which is defined

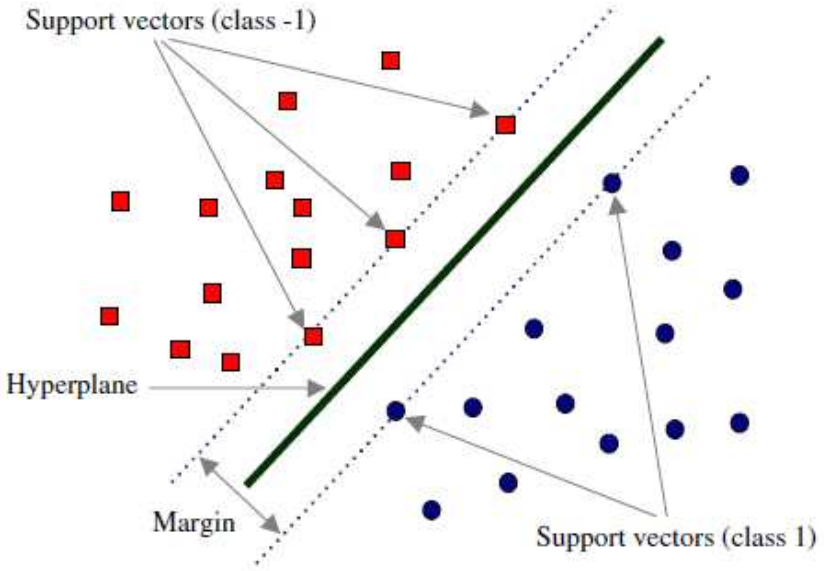


Figure 2.4: Example of SVM separation hyperplane with linear kernel.

as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i + \mathbf{x}_j\|^2\right), \tag{2.13}$$

where γ plays a major role in the performance of the kernel. If underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data. On the other hand, if overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. The γ parameter is determined by performing the grid search approach to find the optimal value, within the following range: $\gamma = \left[2^{-5}, 2^{-4}, \ldots, 2^{3}\right]$.

All the experiments have been performed using the tools provided by LibSVM [35], a library for Support Vector Machines running on MATLAB®.

2.2.1 Support Vector Regression

The Support Vector Regression (SVR) [36] approach, or SVM regression, is derived from the SVM [33] technique. Its goal is to find the function $f(x)$ that represents the target outputs with a maximum deviation ϵ, and, at the same time, is as flat as possible. Therefore, deviations lower than ϵ are not considered as errors, however, higher deviations are not allowed.

Differently from SVM, the solution of a linear model (in the feature space)

for the SVR is obtained starting from a dual optimization problem, given as:

$$\begin{aligned} \text{maximize} \quad & \begin{cases} \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)\left(\alpha_j + \alpha_j^*\right) K(\mathbf{x}_i, \mathbf{x}_j) \ , \\ -\epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) \ , \end{cases} \\ \text{subject to} \quad & \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) & = 0 \ , \\ \alpha_i, \alpha_i^* & \in [0, C] \ , \end{cases} \end{aligned} \tag{2.14}$$

to obtain:

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) \cdot K(\mathbf{x}_i, \mathbf{x}) + d, \tag{2.15}$$

where α_i, α_i^* are determined during the training process, and $C > 0$ represents the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated.

Also for the SVR, the experiments have been performed using LIBSVM library. The experiments have been carried out by using the C and γ parameters chosen by performing the grid search approach within the ranges:

$$\mathbf{C} = \left[2^0, 2^1, \ldots, 2^8\right], \qquad \gamma = \left[2^{-5}, 2^{-4}, \ldots, 2^3\right], \tag{2.16}$$

2.2.2 One-Class Support Vector Machine

In the One-Class Support Vector Machine (OC-SVM) theory [37], in order to separate the data set form the origin, the following quadratic problem has to be solved:

$$\begin{aligned} \min_{w \in \mathbf{F}, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_i \xi_i - \rho \ , \\ \text{subject to} \quad & (w \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0 \ , \end{aligned} \tag{2.17}$$

to obtain the decision function:

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_i \alpha_i \cdot k(\mathbf{x}_i, \mathbf{x}) - \rho\right) \ , \tag{2.18}$$

where $\nu \in (0, 1)$, $\mathbf{x}_i$ denotes the i-th Support Vector (SV) and $k(\cdot, \cdot)$ represents one of the kernel presented above.

The function f returns $+1$ in the "small" region where most of the data points fall and -1 everywhere else. The strategy is to map the data into a features space corresponding to the kernel and to separate the data from the origin with the maximum margin. For a new point $\mathbf{x}$, the value of $f(\mathbf{x})$ is determined by evaluating on which side of the hyperplane of the features space $\mathbf{x}$ falls, as depicted in Figure 2.4.

2.3 Extreme Learning Machine

Extreme Learning Machine (ELM) has been presented by Huang *et al.* [38] as a fast learning algorithm for single hidden layer feedforward neural networks (SLFNs). Differently from standard ANNs approaches, the input weights are randomly generated and the output ones are tuned by a least-square method. In later work [39], an unified solution for regression, binary, and multi-class classification has been presented. Specifically, given a set of pairs $(\mathbf{x}_i, t_i)$, $i = 1, \cdots, N$, where $\mathbf{x}_i \in \mathbb{R}^L$ is the training data, and $t_i \in \{-1, 1\}$ denotes the corresponding label, the output function of the ELM for regression is computed as:

$$f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathbf{T}} \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^{\mathbf{T}} \right)^{-1} \mathbf{T}, \tag{2.19}$$

where $\mathbf{h}(\mathbf{x})$, namely the feature mapping, denotes the hidden layer output for the corresponding input $\mathbf{x}$, and C is the regularization coefficient, whereas the matrices $\mathbf{H}$ and $\mathbf{T}$ denote the hidden layer output matrix and the input labels matrix, respectively.

Moreover, a kernel-based approach has been also proposed [38] if a feature mapping $\mathbf{h}(\mathbf{x})$ is unknown. Differently from the standard ELM, the number of hidden neurons must not be known in advance, and the output relation (2.19) becomes:

$$f(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \mathbf{\Omega} \right)^{-1} \mathbf{T}, \tag{2.20}$$

where $K(\cdot, \cdot)$ denotes the chosen kernel function, and $\mathbf{\Omega}$ defines the kernel matrix, so that $\Omega_{i,j} = h(\mathbf{x}_i) \cdot h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$. In this study the RBF kernel has been evaluated, as for SVM Section 2.2. 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 (2.16).

The MATLAB® code of ELM with kernels provided by the algorithm developers has been used to perform the algorithms evaluation.

Chapter 3

The POWdER System

The POWdER system[40] is a particle size monitoring system used as case study for this thesis. The experimental data collected by this system have been used to develop a solution for condition monitoring of boiler burner used in an industrial power plant, as reported later in Chapter5.

3.1 Acquisition System

The POWdER system has been installed on the burners feeding ducts of two different industrial power plants. This system continuously monitors Acoustic Emission (AE) produced by coal powder impacting on the inner surface of the duct where its is conveyed. The AE signals are processed and used to provide an estimation of the size of coal powder. The POWdER sensors are installed near a duct curve because in this point there is the highest probability that the particles hit the 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 3.1 the scheme of a 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. Instead in Figure 3.2 an example of a real POWdER installation is shown.

The size of coal powder is expressed with Particle Size Distribution, a list of values that defines the relative amount, typically by mass percentage, of particles present in a sample according to size. It is given in terms of mesh, a measurement of particle size often used in determining the PSD of a granular material. Within an industrial plant, the most used approach to measure PSD is the sampling and sieving method: in a first phase, 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 to obtain the values that characterized the PSD. In the case study proposed in this work, three sieves, corresponding to 50MESH (i.e. 300μm), 100MESH (i.e. 150μm), and 200MESH (i.e. 75μm), were taken into account.

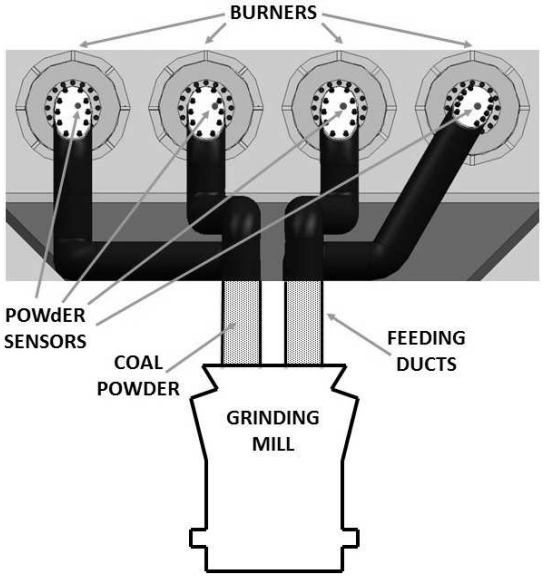


Figure 3.1: Schematic of the power plant structure and the POWdER sensors

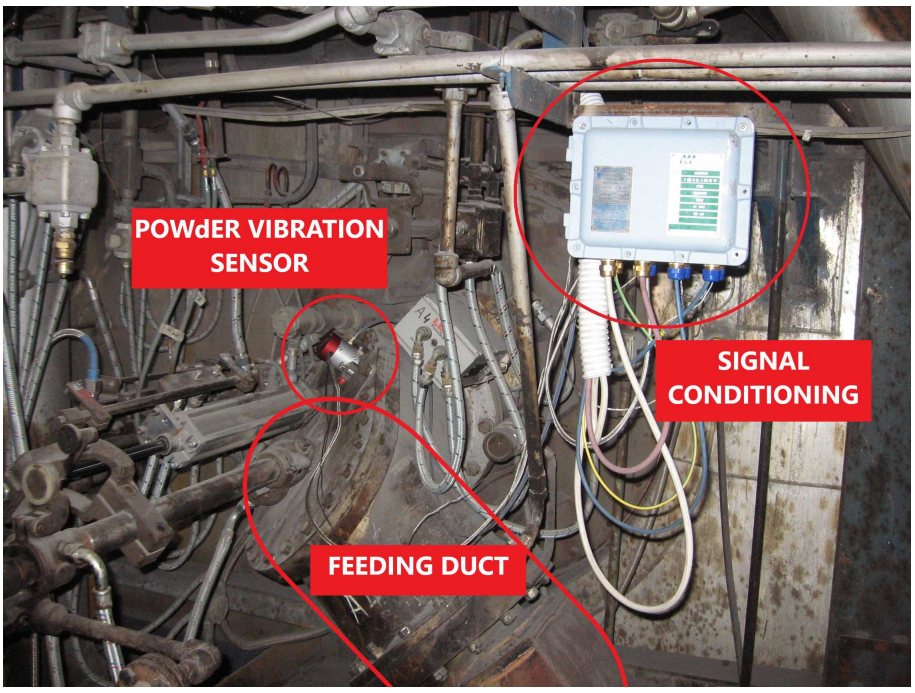
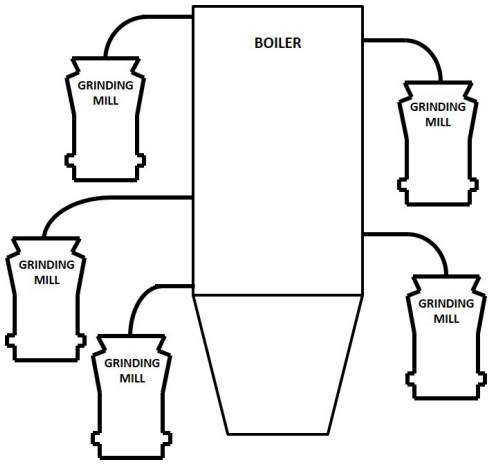
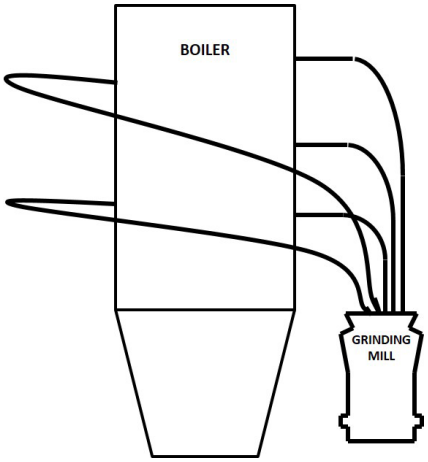


Figure 3.2: Example of real POWdER installation in a power plant

Each PSD is therefore represented by three numerical values, corresponding



(a) Plant A



(b) Plant B

Figure 3.3: Different pipeline structure of two plants

to the percentage of coal particles in the initial sample whose dimensions are, respectively, lower than 300μm, 150μm, and 75μm.

The acquisition of AE signals has 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 the targets vector or the label associated with the corresponding PSD measure.

The data related to two different industrial plants are involved in this work. Hereinafter, the two plants are referred as Plant A and Plant B. The plants differ in terms of pipeline layout, type of coal employed, and structure of mills,

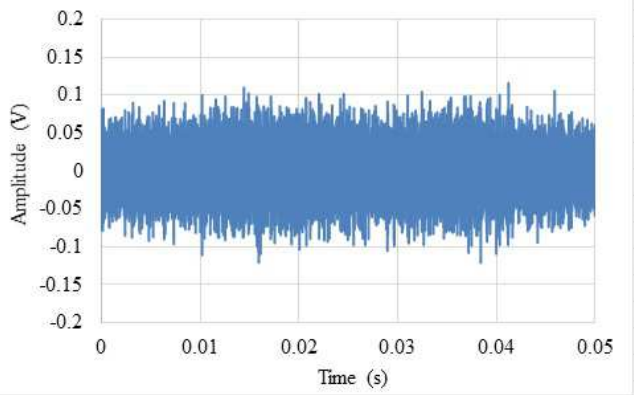


Figure 3.4: Typical time-waveform of acquired AE signal. On the y-axis, the amplitude represents the sampled voltage signal at the sensor output.

which use different grinding elements. The coal mass flow rates are 5 t/h and 4.5 t/h for the Plant A and the Plant B respectively. Furthermore, the plant B has a 660 MWh boiler, double compared to the 330 MWh of Plant A. In both plants five ducts were monitored and all AE acquisitions were performed under different working conditions of the plant. Another important difference between these datasets regards which ducts were monitored inside the plant: in Plant A the five monitored ducts are connected with five different mills (Figure 3.3a), whereas for Plant B the five ducts are connected with the same mill (Figure 3.3b).

3.2 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 3.4. 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 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 [41] as introduced by Bastari and al. [42], 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

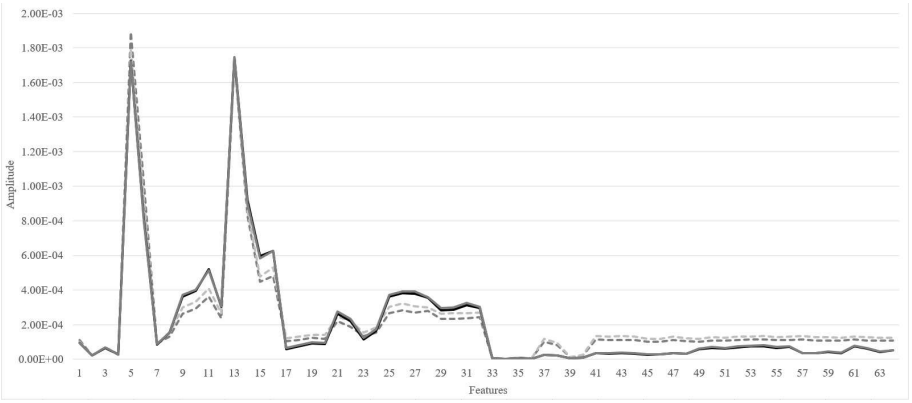


Figure 3.5: Wavelet packet based features extracted from data of a single duct in Condition 1 (dashed line) and in Condition 2 (solid line).

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 3.5 shows an example of how the features look like, 4 vectors X_i extracted from data of one duct, in two operating conditions. Vectors belonging to the same condition are almost identical, while it is possible to distinguish vectors associated with Condition 1 from those associated with Condition 2.

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 $\overline{X}_i$.

Chapter 4

Estimation of Particle Size Distribution for Industrial Plants

The particle size of the powder is an important parameter in many industrial processes, as it affects to the physical and chemical properties of materials. In most cases the powder particles have irregular shapes and speed, and travelling within structures that change its characteristics over time due to usury. Furthermore, it is not interesting to describe the size of single particle but the size of an ensemble of particles, so it is necessary to use cumulative parameters to describe it, such as the PSD. As it is mentioned in the Section 3.1, in industrial plants the evaluation of the PSD is typically performed with the sample and sieve method. This method produces an accurate estimation of the PSD for a given time instant, but it is time consuming and difficult to use for continuous monitoring. To have a continuous monitoring of powder size, it is necessary to employ a system that carries out the estimation in a non invasive way, for the all time horizon of interest.

The main challenge is to find physical model able to describe a very complex system, with unknown and uncontrolled variables, and then to select a set of physical quantities related to the particles size. In this study, AE signals produced by a powder impacting on a metallic surface have been identified as meaningful quantities in order to obtain PSD measure. Leach *et al.* [43, 44] were the first to use AE signals for particle sizing. They collected AE spectra in the range 50-200 kHz, from the impact among particles. By measuring the beat frequencies from different resonance frequencies of particles with varying diameters, they could determine their average diameter and size range. The method gave satisfactory results just for regularly shaped particles, i.e. spheres and cylinders. Unfortunately, this explicit method is impractical for most of industrial applications where are involved fluxes of irregular shaped of particles. However it was demonstrated, at least from a theoretical point of view, that a particle impinging on a metallic surface generate an AE signal containing the information about its size. Many applications of this theoretical result have appeared in the literature [45, 46, 47], confirming the suitability of AEs for

PSD measurement in engineering problems.

The industrial environment provides many other complications for an AE analysis, due to the presence of several sources of AE, so it is also not easy to distinguish between the AE produced by powder and the other sources. To deal with such a complex scenario, suitable techniques have to be developed in order to create a mathematical model without knowing all the variables and all noise sources that influence the process.

The Computational Intelligence methodologies, such as Machine Learning (ML) techniques, can provide useful tools with which develop an appropriate approach that allows to overcome all the raised issues.

4.1 Machine Learning Techniques for Particle Size Estimation

There are many different machine learning algorithms in the literature, in this study the most common algorithms used for regression problems have been considered: ANN, SVR and ELM. Each algorithm has been used with a procedure of Cross Validation (CV), in the form K-Folds with folds containing 15% of the entire dataset available. At each cycle of the CV is selected 15% of the total dataset that is then used as Testing set, while the remaining 85% is used as the Training set. In total 6 different pairs of Training and Testing are involved, making sure that each Testing set had different observations than the others. In order to perform the optimization of the parameters required for the training of the models, a further K-Folds CV has been performed in which, at each cycle, 15% of the Training set is used as a Validation Set. Also in this case 6 combinations of the Training and Validation set have been used. The parameters yielding the lower estimation error in the Validation set has been selected as the optimal parameters. In all tests, features and targets have been normalized in the range [0:1].

4.1.1 Computer Simulation and Results

The tests have been performed with MATLAB® running on a PC with a processor i5 dual-core 2.3 GHz, 4 GB of RAM and Windows 7® OS.

To evaluate the accuracy of the models created by means of the Machine Learning algorithms, the results obtained with the POWdER system3 on the same task, have been taken as useful reference. In this case, only the data related to three ducts from Plant A have been used. In Table 4.1 are reported the partition in Training, Validation and Testing for the employed datasets. The performance evaluation of various models is performed by taking into

Table 4.2: Algorithms Performance on DUCT 1 dataset.

Rmse	**Duct 1**			
	Mesh 50			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.0750	0.0780	0.0763	0.0813
Var Train		1.83e-4	2.20e-5	5.06e-5
Avg Test	0.1390	0.1386	0.1094	**0.1043**
Var Test		7.34e-4	2.94e-4	3.29e-4
	Mesh 100			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.7600	0.8533	0.8874	0.8461
Var Train		3.30e-3	1.00e-3	9.41e-2
Avg Test	1.4000	1.6845	**1.3494**	1.4007
Var Test		1.90e-2	1.62e-2	2.01e-1
	Mesh 200			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	1.4200	1.2191	1.7546	0.6798
Var Train		1.25e-2	3.70e-3	2.02e-2
Avg Test	3.2200	2.8612	2.4950	**2.3023**
Var Test		1.27e-1	2.45e-2	6.83e-2

account the values of the mean and variance of the Root Mean Square Error (RMSE) on the Testing set obtained with the CV.

Table 4.1: Datasets Partition.

DUCTS	**Observations Number**			
	TOTAL	*TRAIN*	*VALID*	*TEST*
Duct 1	272	188	38	46
Duct 2	251	174	35	42
Duct 3	286	198	40	48

In Table 4.2, Table 4.3 and Table 4.4 the error values obtained from tests on selected datasets have been reported. Each table refers to one different duct and reports the result for three Mesh values. The first column shows the values of the system reference POWdER and subsequent columns the error results obtained with the three algorithms of machine learning studied. In the tables, the better results for the Testing set are reported in bold.

By analysing the average values of the error on the Testing data for this firsts two Ducts, it can be observed that the SVR and ELM provide the lower estimation error, while the ANN has the worse results. This trend is also confirmed for the Training data, where the SVR and ELM outperforms the ANN. The SVR and ELM perform also better than POWdER in all cases.

Table 4.3: Algorithms Performance on DUCT 2 dataset.

Rmse	**Duct 2**			
	Mesh 50			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.0600	0.0968	0.0842	0.0408
Var Train		3.44e-4	8.63e-6	7.55e-6
Avg Test	0.1400	0.1453	0.1359	**0.1134**
Var Test		8.43e-4	1.07e-4	4.08e-4
	Mesh 100			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.5700	1.3245	0.6519	0.4421
Var Train		4.08e-2	2.20e-3	3.80e-3
Avg Test	1.1100	2.5003	1.0583	**1.0383**
Var Test		2.32e-1	1.20e-2	5.50e-2
	Mesh 200			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.9200	1.3331	0.7825	0.6591
Var Train		2.50e-2	7.50e-3	2.33e-2
Avg Test	1.9600	2.7403	1.6679	**1.6618**
Var Test		2.24e-1	5.41e-2	8.72e-2

Looking at the variance values, it can be seen that the SVR and the ELM to have a very similar variability, lower than ANN.

For the third Duct, only the algorithm SVR perform better than the POWdER system for he first two meshes, instead, for the Mesh 200, the POWdER system provides better result than all machine learning algorithms.

In the Figure 4.1 and Figure 4.2 the estimation results obtained with the various algorithms are plot in comparison with the corresponding real targets and the targets estimated by POWdER. Both training and testing data related to Mesh 50 of Duct 1 have been used on purpose. In x axis are shown the number of observation and in y axis are shown the values of PSD in terms of percentage, the real and estimated values.

4.2 Heterogeneous Data to Improve the Particle Size Estimation

Several tests conducted with the experimental setup reported in the previous section showed that the number of patterns used for training plays a crucial role and the availability of further data can significantly boost the regression performance. In order to augment the number of training examples, diverse strategies can be considered. The most immediate method involves collect-

Table 4.4: Algorithms Performance on DUCT 3 dataset.

Rmse	**Duct 3**			
	Mesh 50			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.0710	0.0935	0.1006	0.0443
Var Train		2.56e-4	1.10e-4	3.15e-5
Avg Test	0.1400	0.1776	0.1463	**0.1091**
Var Test		1.16e-4	2.49e-4	1.41e-4
	Mesh 100			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	0.6200	0.8489	0.8095	0.4370
Var Train		2.42e-2	1.60e-3	1.50e-3
Avg Test	1.0300	1.5623	1.1413	**0.9898**
Var Test		1.33e-2	2.03e-2	6.40e-3
	Mesh 200			
	POWdER	*ANN*	*ELM*	*SVR*
Avg Train	1.0900	1.2111	1.5067	0.5963
Var Train		3.88e-2	1.10e-3	1.50e-3
Avg Test	1.6400	2.5308	2.1146	**1.7313**
Var Test		6.17e-1	4.15e-2	2.31e-2

ing further samples of powder in the target industrial plant, then measuring the PSD in laboratory. However, this procedure can be time consuming and cost effective. During the acquisition phase, the normal production cycle of the plant has to be interrupted to allow labelling of the target data. A second possibility is represented by the adoption of an active learning strategy; a technique aimed at automatically labelling the unlabelled data gathered during standard operational conditions. The active learning paradigm has been used in many works [48, 49, 50] to face regression problems in which unlabelled data are abundant but labelled examples are difficult to obtain. Another way followed by many Machine Learning researchers, with special focus on the Neural Network area, consists of initializing the network free parameters using suitable unsupervised learning algorithms, thus not requiring target data and related labels. This paradigm has encountered recent success within the Deep Learning community. It has been investigated for different neural architectures, such as Convolutional Neural Network (CNN) [51], Deep Belief Networks (DBN) [52], Recurrent Neural Networks (RNN) [53, 54], and the objective is always the same: inserting some kind of a-priori knowledge into the network by exploiting the available data in order to improve the fine-tuning performance.

This section proposes an approach, alternative with respect to the previous ones. The idea is to use data collected from multiple sources in order to increase the amount of training examples. This was already investigated in the

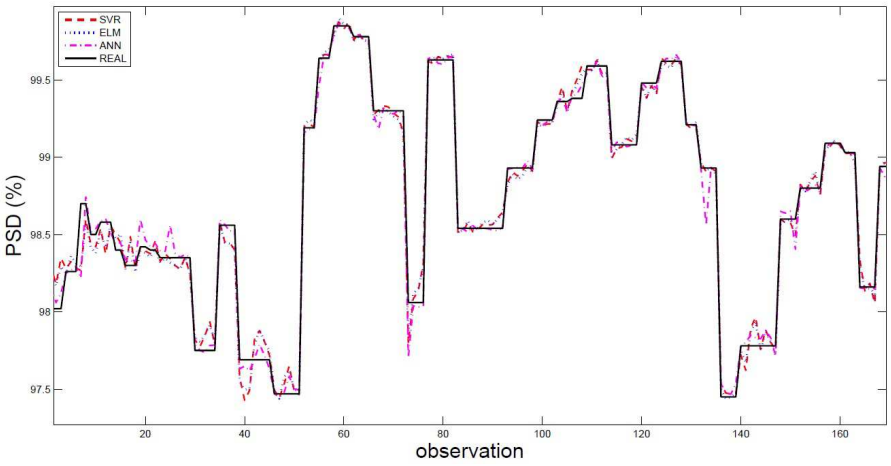


Figure 4.1: Comparison of estimated output values with real target ones on the Training set

field of image processing with CNN [55] and in a biological context with SVM [56]. In many industrial applications, it happens that data related to different plants present some important similarities. In our case study, we have the same type of sensors and process for acquisition, conditioning and processing of signals. Therefore the author wants to explore supervised strategies able to exploit extended availability of heterogeneous data, coming from the nature of the industrial environment and application under study, to improve the PSD estimation performance. Two distinct supervised techniques are proposed to show how such idea can be effectively implemented. They employ ANN as machine learning tool and use the heterogeneous data coming from two different plants to embed some form of a priori-knowledge into the expert system to enhance the regression performance. In particular, a first approach with supervised pre-training of the ANN free parameters by means of data related to the plant not addressed in testing, as preliminary step before the fine-tuning phase, has been investigated. A second approach uses multiple datasets from diverse plants for training ANN models, with the aim to create general mesh-based models. The effectiveness of the proposed techniques has been experimentally proven by computer simulations, as detailed later in the paper.

4.2.1 Datasets

Both the POWdER datasets, from Plant A and Plant B, have been involved in this study. For each duct, the set of data is divided into two distinct sets, namely Primary and Secondary. They differ as they were collected in distinct time periods and under diverse plant operating conditions (different coal flow

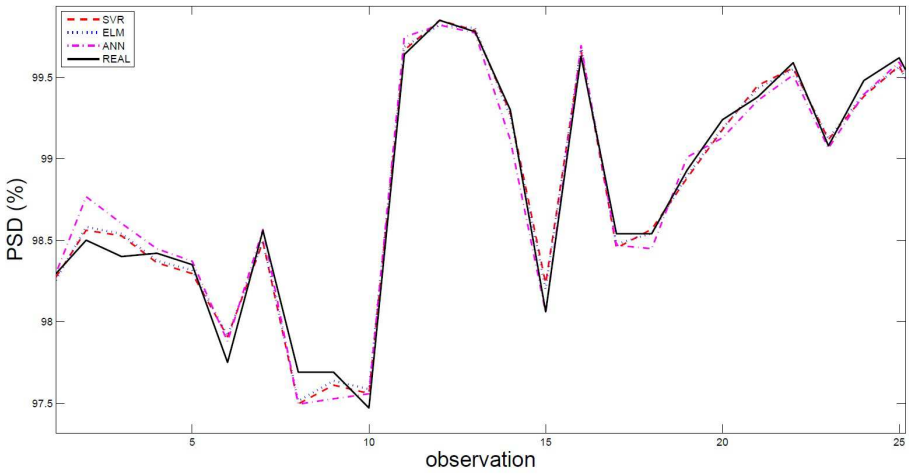


Figure 4.2: Comparison of estimated output values with real target ones on the Testing set

rates). The two datasets can be used to evaluate the generalization capability of the expert system for a certain duct. The number of observations in each Primary and Secondary dataset for both plants is shown in Table 4.5. As detailed later on, in some cases the Primary and Secondary sets correspond to the Training and Test ones, but in others cases the Training and Test sets are extracted solely from the Primary.

4.2.2 Pre-training

The first proposed method is based on a supervised pre-training of ANN by using a dataset related to a distinct plant with respect to the reference one considered for testing. In this way, the ANN free weights are initialized according to the a-priori knowledge learnt from data coming from the non-reference plant, and then fine-tuned to optimize the PSD estimation performance according to the characteristics of the reference plant. The available datasets of plants A and B are used for pre-training and fine-tuning, according to the following combinations:

- **Pre-train Case 1**: the pre-training phase is performed on the union of Primary and Secondary dataset of Plant B and Fine Tuning and Test on Primary dataset of Plant A;
- **Pre-train Case 2**: the pre-training phase is performed on the union of Primary and Secondary dataset of Plant A and Fine Tuning and Test on Primary dataset of Plant B.

Table 4.5: Number of observations for each dataset.

Plant A	*Primary*	*Secondary*
Duct 1	272	36
Duct 2	251	36
Duct 3	333	36
Duct 4	232	36
Duct 5	286	36
Plant B	***Primary***	***Secondary***
Duct 1	295	99
Duct 2	290	99
Duct 3	280	99
Duct 4	270	99
Duct 5	280	99

The overall algorithm is depicted in Figure 4.3, where training is indeed divided into two steps. First, the data for the pre-training are organized into three datasets, each one containing the data of all ducts referring to a specific mesh. Each dataset is used as a training set to obtain three models for each plant, identified as Model 50mesh, Model 100mesh and Model 200mesh. After that, the weights and biases obtained at the end of such training are used to initialize the network, before the completion of the fine-tuning phase. All tests have been performed by using CV with 6 not-overlapping folds. At each CV iteration, a different combination of Training and Test sets is selected from one Primary dataset. The Validation set is extracted from the selected Train set and used to identify the best number of neurons for each network layer.

4.2.3 Multiple Training Sets (TS)

The ANN is also involved in this second method simply based on the collection of both plant datasets in one single training set. The related block scheme is depicted in Figure 4.4. Three specific 50mesh, 100mesh and 200mesh models are trained by using all data from Primary datasets of both plants as training set. The three models created are then applied to the Secondary datasets of individual ducts, used as test set. It must be observed that, due to the nature of the plants and related datasets, there is no correspondence among the ducts of the two plants. This aspect motivated the choice to create the three mesh-based models, with the aim of using them for PSD estimation of any duct in any plant.

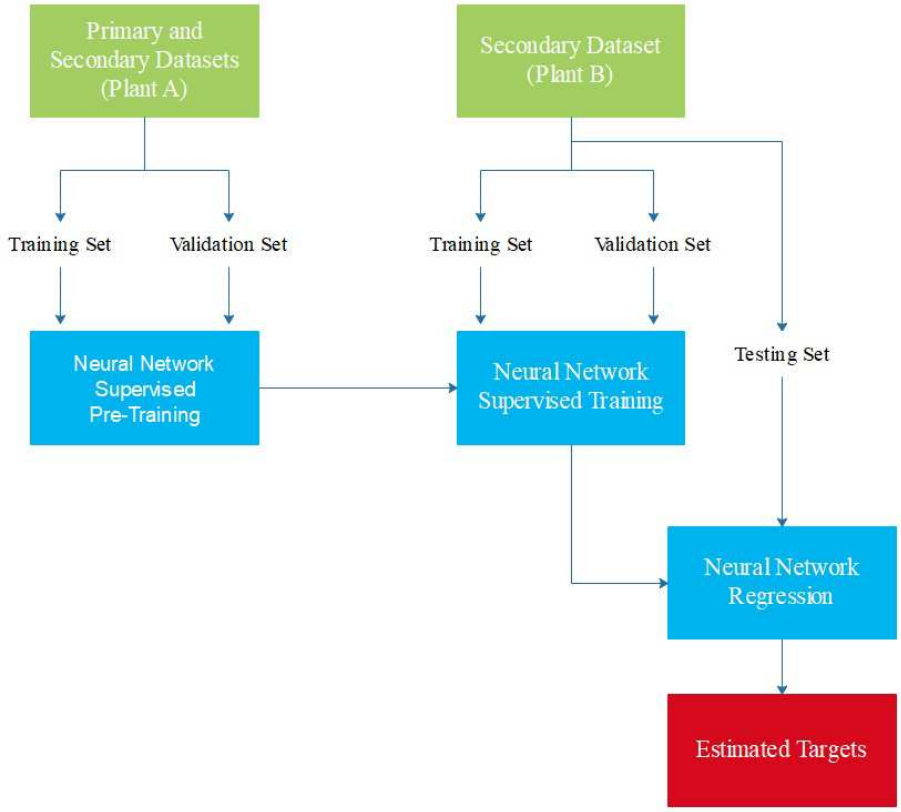


Figure 4.3: Supervised Pre-training Approach

The dataset combinations are the following:

- **Multiple TS Case 1**: the model is trained on the union of Primary datasets of Plant A and Primary datasets of Plant B; then this model is applied on Secondary of Plant A;

- **Multiple TS Case 2**: the model is trained on the union of Primary datasets of Plant A and Primary datasets of Plant B; then this model is applied on Secondary of Plant B.

Different to the previous procedure, in this case the model for the final regression is trained in one phase and not in two phases as previous. This is due to the fact that the heterogeneous data are merged together so just a single training phase is necessary for the PSD estimation.

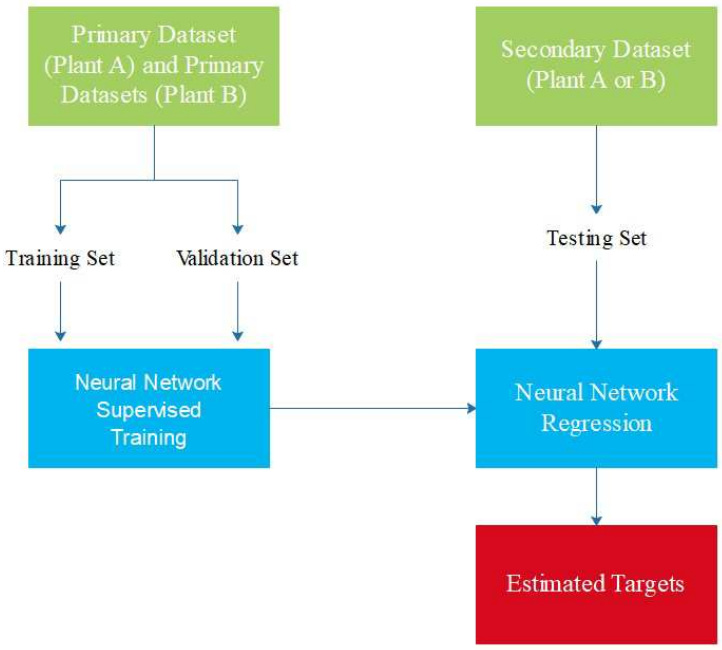


Figure 4.4: The Multiple Training Sets Approach

4.2.4 Computer Simulation and Results

Also in this case, the tests have been performed with MATLAB® running on a PC with a processor i5 dual-core 2.3 GHz, 4 GB of RAM and Windows 7® OS.

The performance evaluation of techniques discussed in Section 4.2 is performed in terms of Root Mean Square Error (RMSE) on the Testing set. For each case study, the standard approach (ANN directly trained and tested on a single duct and mesh dataset) and the proposed supervised method are compared for each duct and each mesh.

However, different RMSE ranges are experienced for the three meshes. This drove the authors to apply a normalization procedure to the RMSE values in order to better compare the regression performance of the three meshes, and likely provide an overall performance index for each experimental case study. According to this procedure, the RMSE performance of the standard and the proposed supervised approach are put in comparison for each duct and each mesh. The two RMSE values are divided by the higher between the two. Finally, due to the large number of duct/mesh combinations and of experiments accomplished, the Normalized RMSE values related to the five ducts of a plant are averaged, providing a total of three indicators (one per mesh) for each experimental case study. In Table 4.6 an example of normalization for the ANN Pre-training Case 1 Duct 1, is shown.

Table 4.6: Normalization Example. Reported values are related to the pre-trained ANN technique applied to Duct 1-Plant A data.

	50 MESH	*100 MESH*	*200 MESH*
RMSE without pre-train	0.1144	1.3555	2.3616
Normalized RMSE	0.8338	1.0000	1.0000
RMSE with pre-train	0.1372	1.2742	2.2329
Normalized RMSE	1.0000	0.9400	0.9455

The results obtained from these experiments will now be discussed. First, the performance achieved with the standard ANN-based approach and the proposed ANN pre-training technique will be compared. In this case study, training and testing are applied in CV and for each CV iteration, the test set is represented by one CV-fold taken from the Primary dataset of the reference plant. In Figure 4.5 the average Normalized RMSE values for the three meshes are reported, with and without pre-training. In this case, the pre-trained models outperform the models without pre-training for 50 mesh (0.9146 vs 0.9340) and 100 mesh (0.9305 vs 0.9368), only for 200 mesh the pre-training provides a mean error of 0.9456 that is higher than the average Normalized RMSE of 0.9436 obtained with the model without pre-training. Figure 4.6 shows the results of Case 2. With this combination of datasets, the standard ANN approach outperforms the new proposed approach for the 100 mesh, providing an average Normalize RMSE of 0.9789. For the others meshes, the new approach returns the best outcomes of 0.9174 for 50 mesh and 0.9610 for 200 mesh.

In the second set of experiments, the effectiveness of the second proposed supervised method described in Section 4.2.3 was tested. The models are trained by using all data contained in the Primary datasets of two plants. This means having three distinct models, one per mesh, which are finally tested against the Secondary dataset of the reference plant for each duct. In this case, the comparison is made with the standard approach involving ANN as an alternative. Figure 4.7 and Figure 4.8 show that the use of multiple training sets to train ANN models provides a lower estimation error than the model trained on a single training set.

In Table 4.7 all obtained results are summarized. A single indicator is reported for each technique denoting the mean of three average Normalized RMSE values associated with the three meshes. In the first comparison, the standard approach uses the ANN to model one single duct by means of a CV procedure. In contrast, in the second approach, the ANN is used to train Mesh-based models, by using the Secondary datasets as Test sets. In the light of this difference, we named the standard techniques in the first and second case study

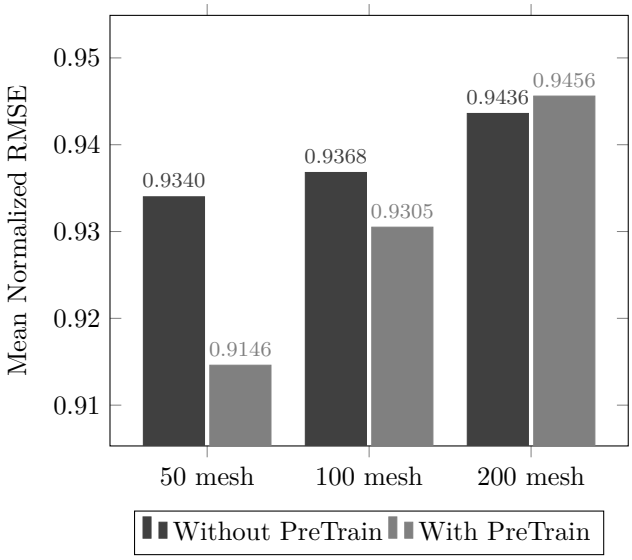


Figure 4.5: Pretrain Case 1

Table 4.7: Overall Mean Normalized RMSE for the two plants taken as reference for testing and for all machine learning techniques under study. Reported values are obtained by averaging the Normalized RMSEs of the three meshes.

	PLANT A	*PLANT B*
ANN-D	0.9594	0.9768
ANN with Pretrain	**0.9518**	**0.9537**
ANN-M	0.9091	0.9779
ANN Multiple TS	**0.7282**	**0.7938**

as ANN-D and ANN-M, respectively. The best results between the standard techniques and the proposed ones are reported in bold for each plant. It can be immediately observed that the new supervised approaches outperform the corresponding standard ones in simulations related to Plant A and Plant B.

4.3 Remarks

The results reported in this chapter demonstrate the possibility to apply the Machine Learning techniques for developing an effective solution for the PSD estimation in non-intrusive way. The new solution wants overcome the drawbacks of the sampling and sieving method usually employed for the PSD estimation.

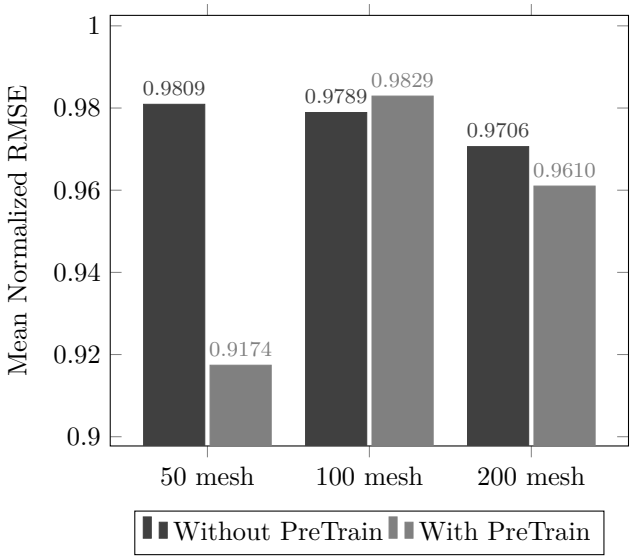


Figure 4.6: Pretrain Case 2

Three algorithms based on ML techniques, i.e. ANN, SVR, and ELM, have been studied. The results obtained with experimental data demonstrate that the proposed solution provides better performance than POWdER system, that has been used as reference, and that can represent a valid substitute for the sampling and sieving method. Among all those employed, the algorithm SVR returns the best results in terms of average error, while algorithms ANN and ELM have worse outcomes than SVR but in line with the reference POWdER values.

The application of heterogeneous information, coming from two POWdER datasets, leads to further improvement with two proposed approaches based on ANN. The obtained results show that the new ANN-based approaches have succeeded to improve the performance previously obtained with standard ANN, for both addressed datasets.

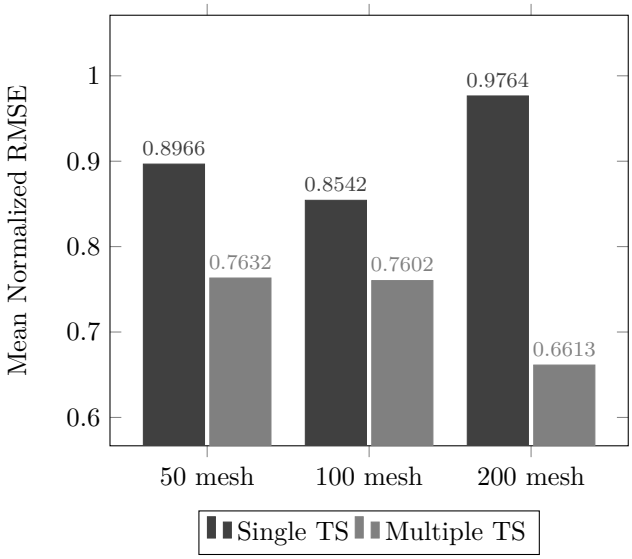


Figure 4.7: Multiple TS ANN Case 1

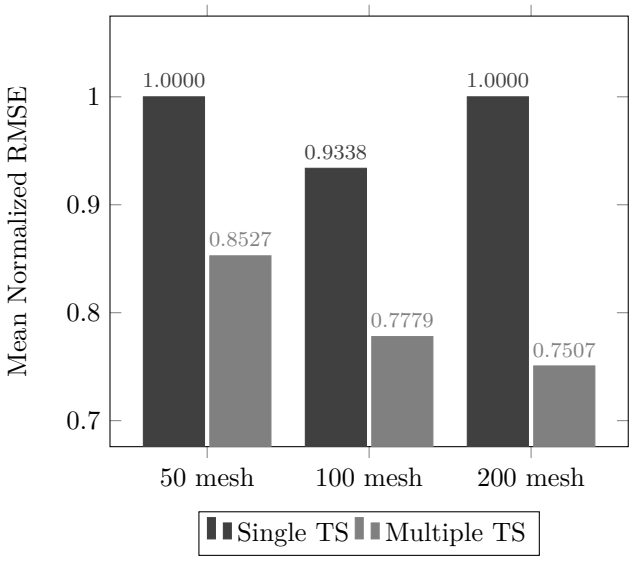


Figure 4.8: Multiple TS ANN Case 2

Chapter 5

A Condition Monitoring Technique for Industrial Power Plants

In this chapter, a condition monitoring approach suitable for coal fired power plant is proposed. 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.

In Chapter4, the effectiveness of ML techniques to train models for the estimation of PSD of coal powder has been demonstrated. Moreover, the AE based informations have proved to be suitable for the estimation of powder size in a non invasive way. The proposed regression approach relies on the reference PSD targets in order to compute supervised learning, and this information is not easily obtainable. 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 presented to solve specific tasks that involve pulverized material for both civil [57, 58] and industrial [59, 60] environments.

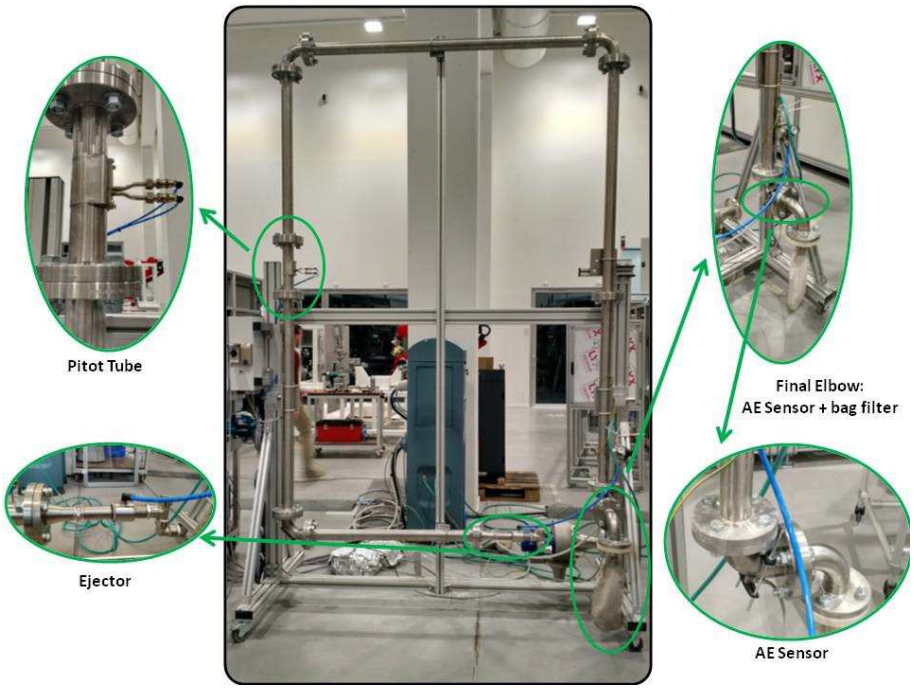


Figure 5.1: Test Rig Layout

For the task addressed in this investigation, it is important to prove the suitability of POWdER acquisition system in combination with a ML-based classifier for classification of pulverized material based on particles size. With this purpose, as first step for designing a CM technique based on particle size classification, a preliminary study has been carried out to investigate the efficacy of an algorithm based on ML technique for the classification of two kinds of powder with different particle size.

5.1 Food Powder Classification with POWdER System

For this test, the POWdER system has been used to measure the AE signals produced by a food powder that flowing inside a test rig set up for this experiment. The powder used for this test is composed by a mix made of sugar and dry milk usually involved in bakery and sweets industry. The test rig is made by an open ring piping where the powder mix is suspended in air and circulated. The average length of the piping is 10m while the diameter is DN50. Air circulation is operated by an ejector that allows changing air velocity between 10m/s and 18m/s by adjusting the pressure of ejector driving air. Air velocity

is measured trough a pitot tube. At the end of the piping there are two elbows so that there is a reasonable chance that the powder hits against the pipes, and a bag filter in order to collect launched powder at the exit. POWdER AE sensor is installed by means of a clamp on the external wall of one of the final elbows of the piping. In Figure 5.1 is showing the complete layout.

Two mixtures of sugar and dry milk with two different PSDs have been analysed, and a classification algorithm based on SVM has been used to distinguish between this two kinds of powder. The performance of the SVM model is evaluated in terms of accuracy by means of the procedure of K-folds CV. For this test, each fold contains the 15% of whole available dataset, thus 6 subsets are selected. The complete dataset includes 388 examples, divided into 3 classes, so each fold is made up by 65 examples. The following list shows the classes association:

- Class 0, associated with the signals generated by the only airflow inside the pipe;
- Class 1, associated with the signals generated by the powder mix 1;
- Class 2, associated with the signals generated by the powder mix 2.

The average accuracies, obtained from the CV, are reported in Table 5.1, both for Training set and Testing set, in terms of mean and variance. Another tool to better understand these results is the Confusion Matrix. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. In Table 5.2 is reported the mean confusion matrix, obtained by averaging the confusion matrices of Testing set coming from the CV.

Table 5.1: Accuracy Results the implemented SVM algorithm.

Mean Accuracy Train Set	**99.64%**
Variance Train Set	0.21
Mean Accuracy Test Set	**93.27%**
Variance Test Set	15.55

Table 5.2: Accuracy Results the implemented SVM algorithm.

		Predicted		
		Class 0	Class 1	Class 2
Actual	Class 0	188	38	46
	Class 1	174	35	42
	Class 2	198	40	48

Results show that the proposed approach provides an average accuracy, both in the test and training sets, over 90% that indicates the good robustness of this technique. More in details, around 93% of 65 AE signals, used as test set, have been classified correctly. By analysing the Table 5.2, it is possible to see that the air (Class 0) is misclassified only one time as mixture 1 (Class 1); mixture 1 is misclassified 1 time as air and 1 time as mixture 2, mixture 2 is misclassified both times as air.

These results demonstrate that the information extracted from the AE signal can be exploited to distinguish between two powder mixture, by means ML techniques.

5.2 The Proposed Condition Monitoring Approach

Moving from the results obtained with previous studies presented in Section 4.1 and Section 5.1, a new approach for condition monitoring of coal powder burned as fuel in power plants has been designed. The method proposed in this section 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, ANN, and ELM, to correlate AE signals acquired on the feeding ducts and PSD. The proposed method was tested on both available dataset coming from both the two industrial power plants, Plant A and Plant B. Differently from the previous tests, for this study only two out three meshes have been considered, i.e. 50MESH and 200MESH. These two represent the limits of the particle size range considered to discretize the PSD curve for this case study, for this reason they are the most useful to character-

ize the operating condition of the system. Three tests have been designed and carried out to show the response of the proposed approach to specific issues.

5.2.1 Binary Classification

The standard approach to the problem of PSD classification involves the use of only two classes to identify two particle 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. 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.

5.2.2 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 5.2 shows a comparison between the

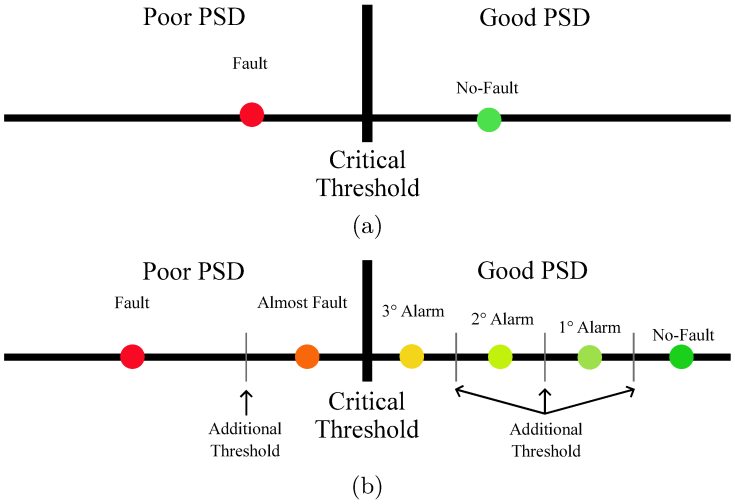


Figure 5.2: Comparison between the monitoring capabilities with (a) 1 threshold and 2 classes, and (b) 5 thresholds and 6 classes.

cases with 2 and 6 classes. Figure 5.2a 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 5.2b.

5.2.3 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 characterize it. As detailed in the Section 3, it is expensive, both in terms of time and resources, to get examples with associated targets needed for a supervised training by mean the sampling and sieving method. 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.

5.2.4 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 [61, 62, 63]. 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 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 False Positive Rate (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.

5.3 Computer Tests and Results

The tests have been performed with MATLAB® running on a PC with a processor i5 dual-core 2.3 GHz, 4 GB of RAM and Windows 7® OS.

The performance evaluation of the techniques discussed in Section 5.2 is carried out in terms of Mean and Standard Deviation of Accuracy array for the Testing set obtained at the end of the 6 iterations of the CV procedure.

Table 5.3: Binary Classification VS 4 and 6 classes classification for 50 mesh

	Plant A			Plant B		
	SVM	ANN	ELM	SVM	ANN	ELM
2 classes	96.59%	**97.77%**	96.50%	97.51%	97.34%	**97.55%**
4 classes	**94.47%**	94.26%	92.92%	93.08%	**93.37%**	92.53%
6 classes	**92.29%**	92.26%	90.54%	**90.21%**	88.72%	88.58%

5.3.1 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.

In order to make easier the comparison between the results of the various tests, Tables 5.3-5.4 show the average values of the accuracy 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%, if compared with 96.59% of SVM and 96,50% of ELM; for Plant B the performance is very similar for all the algorithms, 97.51% for SVM, 97.34% for ANN and 97.55% for ELM. The algorithm which uses the SVM achieves for both plants for the 200 mesh the best performance, providing an average accuracy of 97.13% for Plant A, while ANN and ELM obtain respectively the 96.59% and 96.75%, and an average accuracy of 94.02% for Plant B, higher than 92.86% and 93.28% respectively of ANN and ELM.

Table 5.4: Binary Classification VS 4 and 6 classes classification for 200 mesh

	Plant A			Plant B		
	SVM	ANN	ELM	SVM	ANN	ELM
2 classes	**97.13%**	96.59%	96.75%	**94.02%**	92.86%	93.28%
4 classes	**91.37%**	91.21%	90.95%	**87.98%**	85.30%	86.03%
6 classes	**89.07%**	88.91%	88.44%	**82.70%**	80.02%	80.50%

5.3.2 Multi-class Classification Results

In Tables 5.3-5.4 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 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. For the 200 mesh only for Plant A all the algorithms provide an accuracy higher than 90%, with Plant B data all the algorithms return lower accuracies. By increasing the number of classes to 6, there is a further overall reduction of accuracies 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%. Comparing the SVM, ANN and ELM models, the SVM-based algorithm achieve the best results for almost all the datasets.

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 5.5 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 4.3% for Plant A and of 7.30% for Plant B, while ANN and ELM have a greater reduction in 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 5.3.1 about the 50 mesh data and 200 mesh data.

Table 5.5: Accuracy variation between classification with 2 and 6 classes.

		PLANT A	*PLANT B*
50 mesh	SVM	**-4.30%**	**-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%

Table 5.6: Dataset reduction for 50 mesh

	Plant A			Plant B		
%Dataset	SVM	ANN	ELM	SVM	ANN	ELM
100%	96.59%	**97.77%**	96.50%	97.51%	97.34%	**97.55%**
90%	97.40%	97.39%	**97.50%**	**97.51%**	97.02%	97.38%
75%	84.61%	79.83%	**93.73%**	84.98%	**85.46%**	83.82%
50%	76.12%	74.07%	**81.63%**	79.68%	**80.92%**	80.42%

5.3.3 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 classification accuracy on the Testing set in relation with the progressive reduction of the observations number used for training the models with the developed algorithms. The studies were performed by using binary classification and three sub-sets of data, respectively with 90%, 75% and 50% of the whole dataset.

According to Table 5.6 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%

Table 5.7: Dataset reduction for 200 mesh

	Plant A			Plant B		
%Dataset	SVM	ANN	ELM	SVM	ANN	ELM
100%	**97.13%**	96.59%	96.75%	**94.02%**	92.86%	93.28%
90%	**96.57%**	96.13%	96.37%	**94.86%**	94.39%	94.27%
75%	91.19%	**92.44%**	90.47%	**95.15%**	94.98%	94.85%
50%	81.72%	81.55%	**87.57%**	89.14%	90.60%	**91.40%**

Table 5.8: 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%**

with SVM and ANN algorithms.

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%. 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.

Table 5.8 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.

5.3.4 False Positive Reduction Results

The metric used to evaluate the performance of the techniques proposed to reduce the FP, is the False Positives Rate (FPR). 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. It then defines the FPR as:

$$FPR = \frac{FP}{FP + TN} \tag{5.1}$$

In the Figures 5.3, 5.4, 5.6 and 5.5 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.(5.1) for the original classifier (STD), and the proposed Decision Threshold Technique (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 5.3 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 5.4, 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 5.5 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 5.6 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.

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

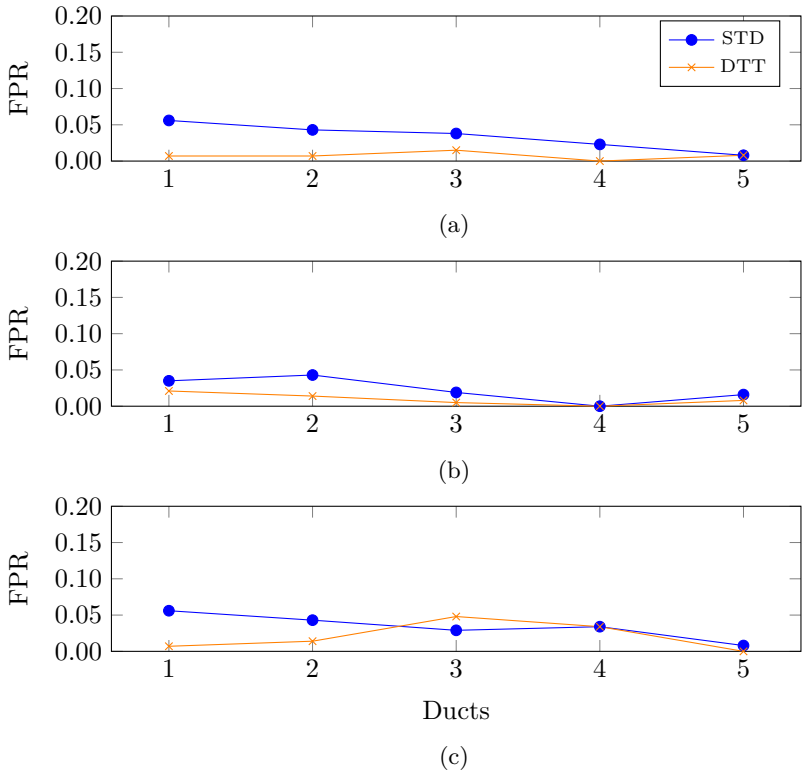


Figure 5.3: FPRs comparison for data 50 mesh, Plant A: (a) SVM, (b) ANN, (c) ELM.

in the Section 5.3.1, 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.

5.4 Remarks

In this chapter a new solution for non-invasive condition monitoring has been proposed. This solution is designed for the monitoring of burners devoted to coal powder combustion in an industrial power plant. The core of the solution is the classification algorithm that classifies the coal powder on the basis of its PSD as it enters the burner. The final goal is to monitor the boiler burner operating condition with the purpose of holding the optimal powder particle size that ensures the maximum combustion efficiency. With this solution it is possible detect non-optimal system set-up before the begin of the combustion process, and so act consequently to avoid drops in energy production efficiency.

Three classification algorithms based on ML techniques, i.e. SVM, ANN and

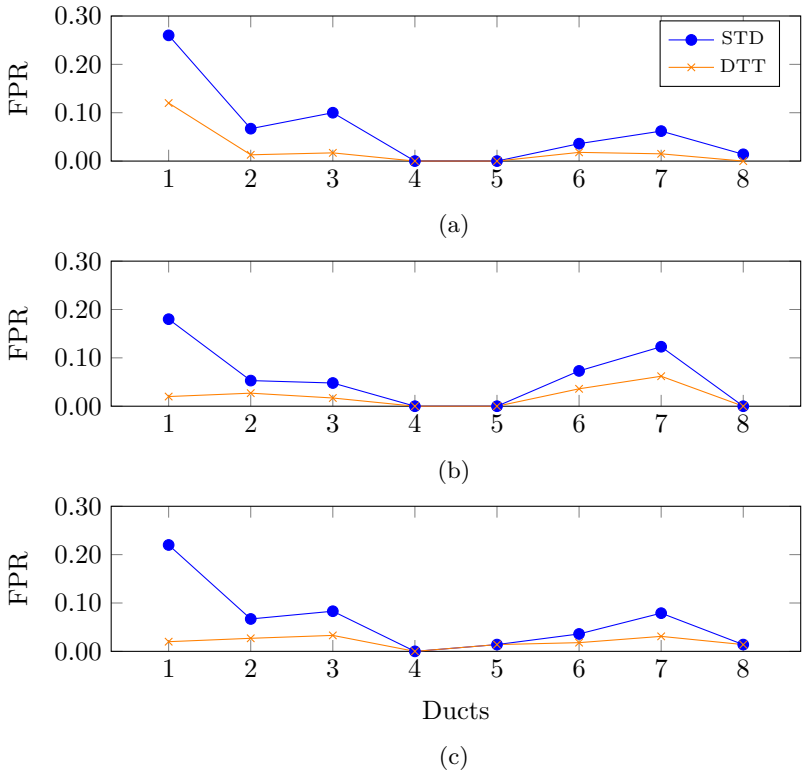


Figure 5.4: FPRs comparison for data 50 mesh, Plant B: (a) SVM, (b) ANN, (c) ELM.

ELM, have been implemented and compared. 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. 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.

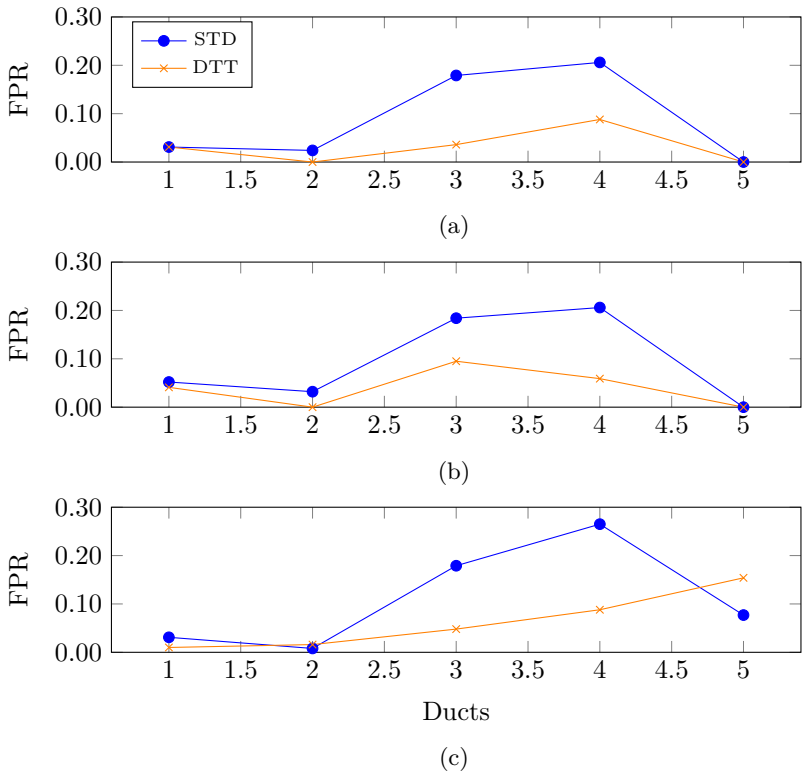


Figure 5.5: FPRs comparison for data 200 mesh, Plant A: (a) SVM, (b) ANN, (c) ELM.

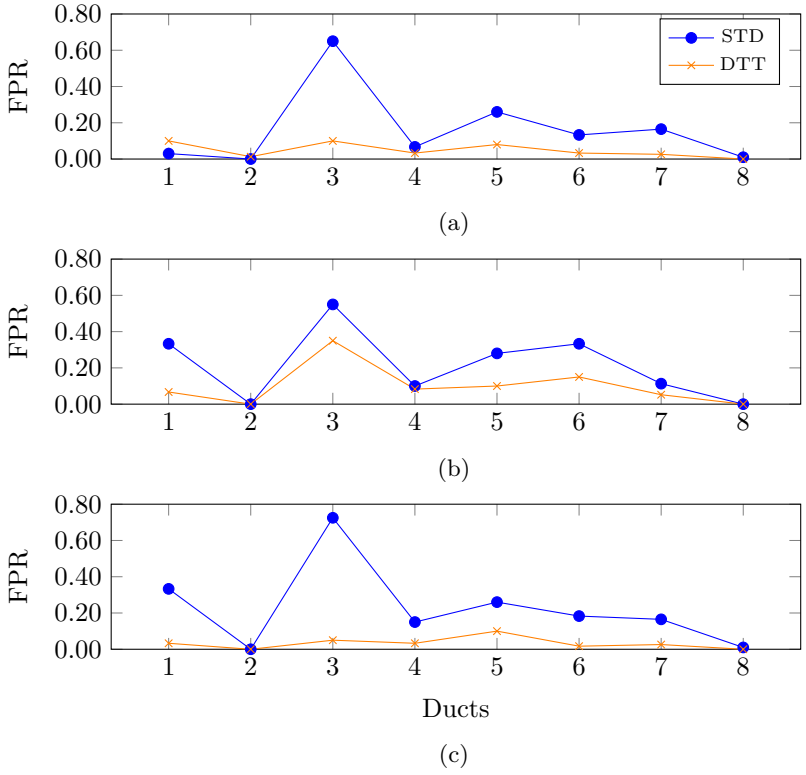


Figure 5.6: FPRs comparison for data 200 mesh, Plant B: (a) SVM, (b) ANN, (c) ELM.

Chapter 6

Faults Detection for Rolling Bearings

The detection and the prediction of faults in industrial machines is a crucial aspect to enhance reliability and reduce maintenance cost. Bearings are widely used in induction motors and rotating machines, so they are common components in many mechanical systems. Due to their diffusion, many of the failures that occur in industrial machines are ascribed to these components, in fact statistical studies show that bearing failures account for around 50% of the total failures of these devices [64, 65, 66]. A bearing failure can accelerate machinery deterioration and bring dangerous consequences for the machine operators present, so bearing condition monitoring becomes an important measure to ensure machine safety. Due to the importance of this topic, a lot of works and studies were published during past years by providing many different approaches to improve the detection and the classification of the faulty bearings [67]. Fault diagnosis of bearings is usually based on vibration signals, and a set of features is extracted from these signals in order to classify the faults. The features could be in the time domain, frequency domain or time-frequency domain, and the prevalent techniques include Kurtosis [68, 69], Wavelet Transform [70, 71], Cepstrum [72], or Envelope Spectrum [73].

The bearing behaviour depend on several parameters and their diagnosis during regular operation will often involve analysis of non-stationary signals as the rotating speed, loads, and environmental conditions vary with time. Therefore, the algorithms for fault detection and classification should allow analysis of non-stationary or quasi-stationary signals. Zvokelj et al. [74] realized the non-stationarity of the data collected from the monitoring of bearings, and so proposed an approach that combine the Principal Component Analysis (PCA) and the Ensemble Empirical Mode Decomposition (EEMD) methods. PCA reduces the dimensionality of the data and EEMD decomposes the signals into various time scales to allow the extraction of appropriate characteristics from the faulty bearing signals. Both time and frequency domains provide some features that can be used as characteristics of these non-stationary vibration

signals, useful to determinate the healthy state of the bearing. A means to extract these characteristics is by using images, which can provide a comprehensive description of vibration signals, including information about the various bearings faults. In the work presented by Wei Li et al. [75], the spectrum images are processed with two-dimensional PCA to reduce the dimensions, then a minimum distance method is applied to classify the bearing faults. Liang Hua et al. [76] presented a method where the acquired mechanical vibration signals are converted into colour time-frequency spectrum images by the processing of pseudo Wigner-Ville distribution. Then a feature extraction method based on quaternion invariant moment is used, combining image processing technology and multi weight neural network technology. Klein et al. [77] provided a method for the bearing fault diagnosis by applying the image processing techniques such as ridge tracking and related algorithms on the time-frequency representation, available by STFT or wavelet, of the non-stationary bearing signals. All the works in literature deal with the analysis and the diagnosis of the bearing faults by trying to get as close as possible to the real working conditions of a bearing. The main problems for this type of analysis are the non-stationarity of the signals or the variation of the operating parameters.

The method presented in this chapter provides an innovative monitoring approach that uses time-frequency analysis and image processing with a machine learning technique for the bearing faults classification; dealing with the aforementioned problems simultaneously. The proposed method involves a feature extraction technique that decomposes the vibration signals with Empirical Mode Decomposition (EMD) and Principal Component Analysis. Then Spectrograms and Images Moments are used to extract characteristics that allow the discrimination between different bearing faults by using a classification algorithm. This approach provides a classification of damaged bearings for different and variable set-ups in stationary and non-stationary conditions. Due to its characteristics, the method can be used to obtain a single control model suitable to monitor various bearings that operating under different conditions.

6.1 Theoretical Principles

This section gives a brief overview of the theoretical principles of the techniques used for the development of the proposed faults detection and classification method.

6.1.1 Vibration signals generated by bearing defects

The vibration response of a defective bearing consists of a series of impulses that are generated every time a running roller passes over the surface of the

bearing damages. The amplitude of the impulses are a measure of the intensity of the shock. Their size depends on the speed, the spatial extent of the damage as well as on the load conditions on the bearing. The frequency at which these impulses are produced is called the Bearing Characteristic Frequency (BCF), which depends on the shaft speed, the geometry of the bearing, and the site of defects. Generally, four types of BCFs are encountered for ball bearings: bearing pass frequency of outer race, bearing pass frequency of inner race, ball spin frequency, and the fundamental train frequency, which correspond to defects on the outer race, the inner race, the roller, and the cage, respectively. The amplitude of these BCFs is characteristically a sign of defect severity, and the presence of harmonics of the BCFs is an indication of the defect origin.

6.1.2 Empirical Mode Decomposition

Empirical Mode Decomposition [78] is based on the direct detection of local signal extrema at a variety of intrinsic time scales to decompose the signal. The resulting Intrinsic Mode Functions (IMFs) are considered as the most important characteristics of the signal, and since the decomposition is based on localized time-scales, is readily applicable to non-linear and non-stationary signals. An IMF of a signal from EMD satisfies the following two conditions: (1) the number of extrema and the number of zero crossings are equal, or their difference is no more than 1, and (2) its local mean is zero. Specifically, given a signal $x(t)$, the constituent IMFs, $c_i(t)$, can be obtained and summed such that:

$$x(t) = \sum_{i=1}^{K} c_j(t) \cdot r_K(t), \tag{6.1}$$

where K is the number of IMFs, $r_K(t)$ the final residue, which is the mean trend of the signal, and $c_i(t)$ represents the IMFs that are nearly orthogonal to each other and whose mean is close to 0.

6.1.3 Principal Components Analysis

Principal Components Analysis [79] is a feature extraction method that reduces the dimensionality J of data $\mathbf{X}_{MJ}$, where M is the number of observations and J the number of features, with a minimum loss of information by projecting the data into a lower dimensional subspace, which contains most of the variance of the original data. PCA thus represents data as the product of the mutually orthogonal data called scores $\mathbf{T}_{Mn} = [t_1, t_2, \ldots, t_n]$ and transposed linear transformation matrix $\mathbf{P}_{Jn} = [p_1, p_2, \ldots, p_n]$ also called the principal

component matrix, as shown in Eq.6.2:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T = \sum_{i=1}^{n} t_i \cdot p_i^T, \tag{6.2}$$

with $n < J$.

To create the best subspace for our signals, it is crucial to determine the appropriate number of principal components n to select. Due to the importance of this aspect, many studies have been presented in literature with different techniques proposed. For this study, the selection of the number of principal components has been performed by a comparative test. The procedure is based on the evaluation of classification accuracy provided by each different number of principal components in order to select the number that provides the best performance.

6.1.4 Spectral Analysis

Spectrograms are a visual representation of the spectrum of frequencies in a sound or other signal as they vary with time. They provide a way to recognize fault patterns in the time-frequency domain characteristics. In this work, spectral analysis is employed to create images that characterize the signals. A spectrogram is obtained by calculating the Short Time Fourier transform (STFT) [80] of the signal:

$$STFT\{x(t)\}(\tau, \omega) = \int_{-\infty}^{\infty} x(t) \cdot w(t - \tau) \cdot e^{-j\omega t} dx, \tag{6.3}$$

where $x(t)$ is the original signal, and $w(t)$ is the window function, for this work a Hamming window is used, and then:

$$spectrogram(t, \omega) = |STFT(t, \omega)|^2, \tag{6.4}$$

The spectrogram provide an image where the darkest colours represent higher amplitude of energy and show the relation between the signal in time domain and frequency domain.

6.1.5 Image Moments

In image processing, moments are scalar quantities used to characterize an image and to capture its significant features. Given an function $I(x, y)$, that represent the pixel intensity of an image, if this function is piecewise continuous and has non-zero values only in a finite region of the (x, y) plane, then the moments sequence M_{ij} is uniquely determined by $I(x, y)$. By considering that an image segment has finite area, or at least is piecewise continuous, moments of

all orders exist and a complete moment set can be computed and used uniquely to describe the information contained in the image. In the case of a grey-scale image, the image moments M_{ij} are calculated by:

$$M_{ij} = \sum_x \sum_y x^i \cdot y^j \cdot I(x, y), \tag{6.5}$$

where x and y are the pixel abscissa and ordinate. The central moments μ_{ij} are invariants with respect to translation and are defined as:

$$\mu_{ij} = \sum_x \sum_y I(x, y) \cdot (x - \overline{x})^i \cdot (y - \overline{y})^j, \tag{6.6}$$

where $\overline{x} = \frac{M_{10}}{M_{00}}$ and $\overline{y} = \frac{M_{01}}{M_{00}}$ are the components of the centroid. The 10 central moments up to order 3 are used to calculate the 7 Hu moments [81], that are invariant to translation, changes in scale, and rotating. For this work, all the 10 central moments and the 7 Hu moments are used as features set for the images.

6.2 Bearing Faults Simulator

The data used for this study were obtained from a Bearing Faults Simulator, Figure 6.1, a laboratory device produced by GUNT that mimics the behaviour of a standard rotary machine. An electric motor transmits rotation via a shaft to the bearing located in its housing. A micrometer screw generates a radial force that acts on the movably mounted housing with the stationary outer bearing race, thereby exerting a radial load on the bearing. The vibration sensor is a piezoelectric accelerometer with integral electronics mounted on the vertical axis that transforms vibrations into electrical signals collected by the control unit, then transferred into the PC.

Figure 6.1: Bearing Faults Simulator.

Table 6.1: Description of Bearing Faults Simulator Setups.

Stationary State	**Parameters**	
	Rotating Speed (rpm)	*Load (mm)*
Setup 1	1700	4
Setup 2	1700	6
Setup 3	3000	4
Setup 4	3000	6
Non-Stationary State	**Parameters**	
	Rotating Speed (rpm)	*Load (mm)*
Setup 1	1700	Variable Load from 4 to 6 and from 4 to 6
Setup 2	3000	Variable Load from 4 to 6 and from 4 to 6
Setup 3	Variable Speed from 1700 to 3000 and from 3000 to 1700	4
Setup 4	Variable Speed from 1700 to 3000 and from 3000 to 1700	6

Three bearings are used for this work, each one characterized by a fault associates to a single BCF: bearing with fault on the outer race (BOF), bearing with fault on the inner race (BIF) and bearing with fault on the rolling element (BRF). A fourth bearing with no faults (BNF) is used as reference.

Rotating machinery usually works under different loads and speeds, for this reason the presented technique is aimed to be suitable under stationary and non-stationary conditions. Hence, two rotating speeds and two loads are used to create 4 different simulator set-ups with which demonstrate the effectiveness of the proposed method. A first set of signals was recorded under constant conditions, in order to simulate stationary or quasi-stationary behaviour. To simulate a more real situation that involves variable speed and load, a second acquisition was carried out by using simulator set-ups with the same speeds and loads of the previous case, but by varying one of the two parameters during the acquisition. All acquired signals have duration of 8 seconds and they were acquired with a sample frequency of 32 KHz. The characteristics of the used set-ups are listed in Table 6.1, for both the stationary and non-stationary case. The loads 4 and 6 mm, corresponding to a force applied on the cage of approximately 5 and 32 N respectively.

6.3 The Proposed Fault Detection Method

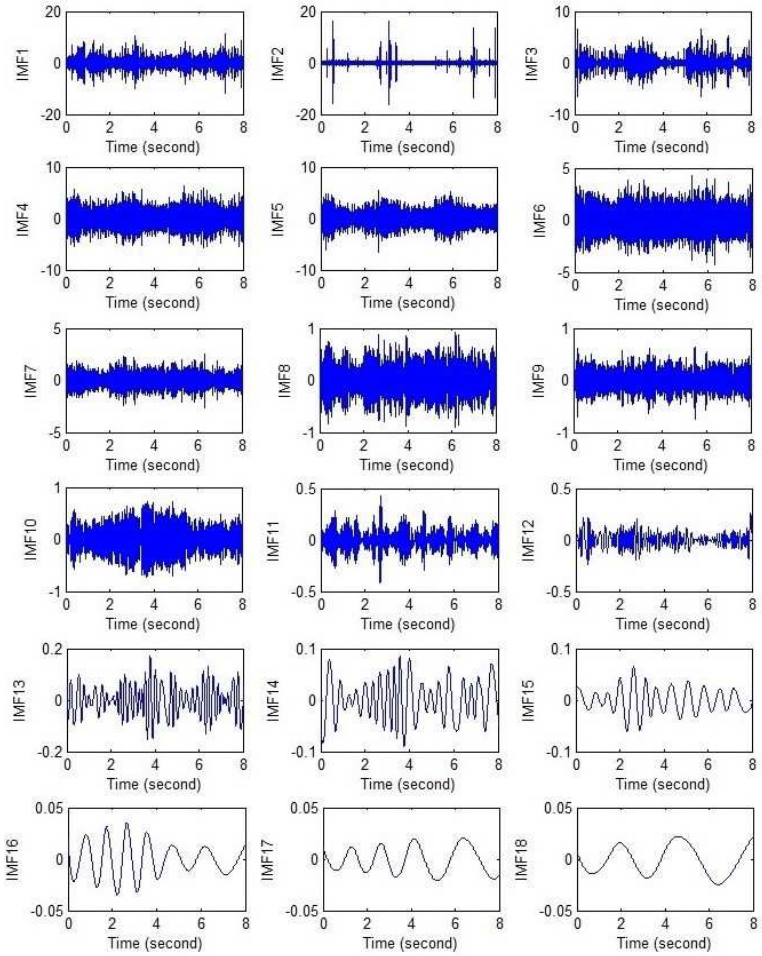


Figure 6.2: Example of IMFs of bearing vibration signal.

The proposed method combines EMD, which adaptively decomposes signals into various time scales, with PCA, which provides a mechanism to extract useful information from a signal by reducing the number of involved variables. The EMD technique is suitable for analysing non-stationary signals recorded from non-linear systems, which is often the case of real systems that use bearings. Furthermore due to its adaptive empirical nature appropriate for the processing of signals exhibiting non-linear characteristics.

In the second step, PCA gets a new representation of the IMFs matrix in a new space created by the Principal Components (PCs). Figure 6.3 shows the PCs obtained from the IMFs shown in the Figure 6.2.

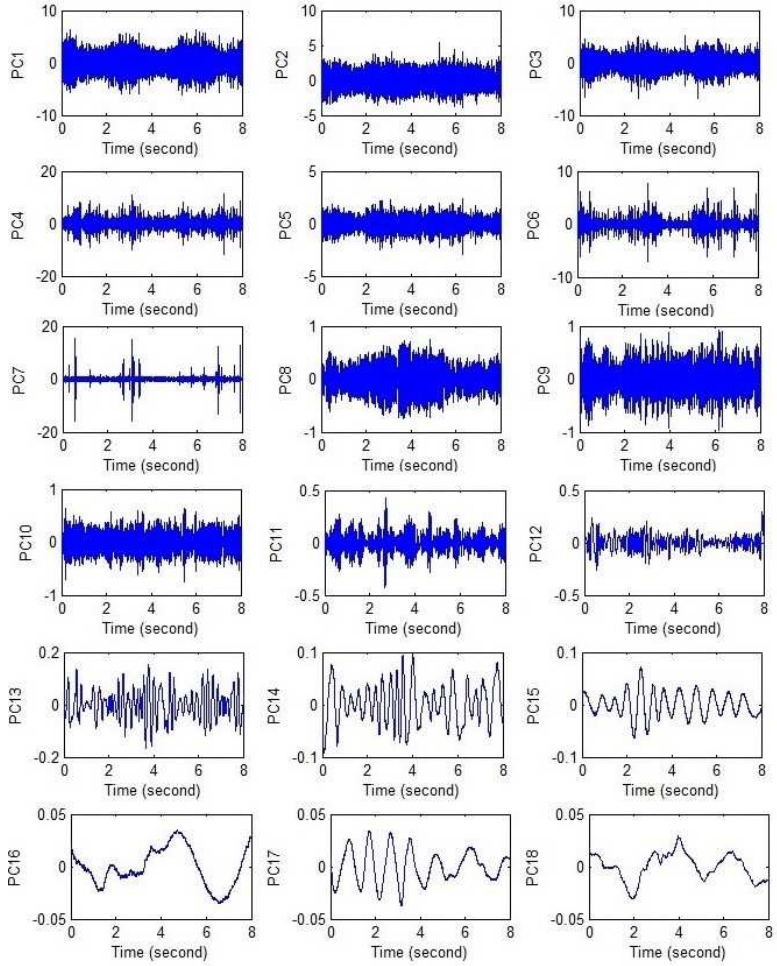


Figure 6.3: PCs of the IMFs after performing EMD on the original signal.

In the first step of the presented technique, the vibration signal is decomposed into the IMFs through the EMD algorithm. For example, Figure 6.2 shows the extracted IMFs for a signal decomposition case; the residual is not shown because it will not be considered in the next step. All IMFs, less the residual, are arranged in a matrix so that each column corresponds with an IMF, and each row represents a different time instant. This matrix is used for the PCA in the following step.

Each PC is taken as a standalone signal to obtain a representation in the time-frequency domain through the spectrogram. The spectrogram is saved and used as an image to characterize the PC. By calculating the spectrograms of all the PCs, a series of images that characterized the time-frequency relationship of the original signal is obtained, as shown in Figure 6.4. A selection is performed

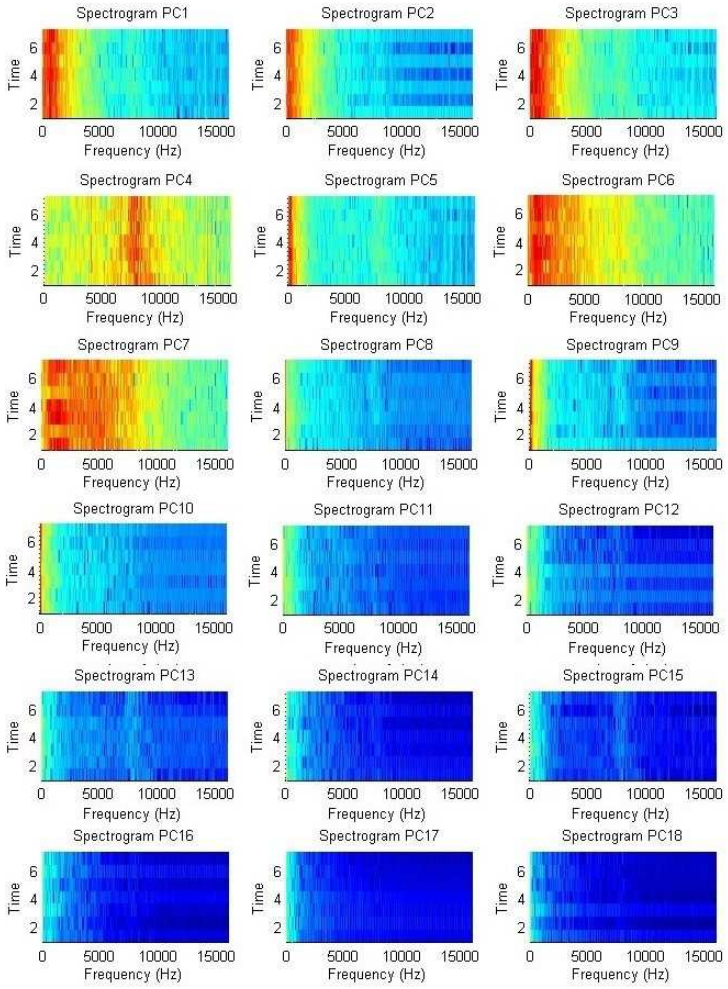


Figure 6.4: Spectrograms of the PCs.

to keep only those images that contain information useful to differentiate the signals exhibiting different faults. The choice of the number of images, thus the number of PCs, used to characterize a signal has been made by evaluating the classification accuracy provided by the classification algorithm that will be introduced later. The algorithm needs the same number of features for each observation, for this reason, to find the best number of PCs, the accuracy is evaluated on the overall dataset, and for each signal, the same number of images has been chosen.

As expressed earlier, the behaviour of a bearing is dependent on many factors, some under the control of the operator, such as the rotating speed and the load, while others are due to internal and external sources to the machine,

independent of human control. A monitoring system that must provide information about the state of a device must be robust to minimal changes of the system and take into account all factors that may affect the system, including those that are difficult to control. The use of a machine learning algorithm is suitable for a black box approach to create a model that takes into account all the factors that influence the phenomenon, both controllable and uncontrollable. In order to use the algorithm based on SVM to classify faulty bearings, it is necessary convert the images into a series of values that can be handled by the algorithm. To obtain this new representation, the selected images are converted into grey-scale images in order to extract their image moments, 17 for each image. The moments related to the images of a single signal are gathered together to form the feature vector associated with that signal.

6.4 Computer Simulation and Results

To demonstrate the efficacy of the proposed method for identifying different faulty bearings for stationary and non-stationary conditions, two typologies of tests are presented. The first test shows the accuracy obtained with a dataset composed by stationary signals only. The second tests the possibility to use a model trained on stationary data to classify signals collected under non-stationary conditions.

The tests have been performed with MATLAB® running on a PC with a processor i5 dual-core 2.3 GHz, 4 GB of RAM and Windows 7® OS.

6.4.1 Bearing faults classification with stationary signals

All the observations related to the stationary signals are gathered together to form a unique dataset. The evaluation of the accuracy has been performed using a CV K-Folds technique with 6 non-overlapping folds. At each CV iteration, a different combination of Training and Test sets is selected from the dataset. The Validation set is extracted from the selected Training set and used to identify the best values for the parameters C and γ, necessary to train the SVM model. The final performance is evaluated in terms of mean accuracy and standard deviation of the Test set. The accuracy of a model trained with a machine learning algorithm is strongly influenced by the data used for training, because the data must be able to characterize the phenomenon under observation, providing examples of all operating conditions that must be monitored. For this reason, an important parameter to take into account in order to obtain an accurate classification, is the number of observations associated with each class. Therefore, various tests were carried out with an increasing number of observations to see how the accuracy varies and then select the appropri-

ate number of observations. Along with the study on number of observations, the evaluation of the number of PCs, to take into account to obtain the best classification accuracy, has been performed. The graph in Figure 6.5 shows the evolution of the average accuracy obtained on the Test set with CV as a function of the number of observations that each class contains as well as the number of PCs.

Figure 6.5: Accuracy evolution in function of the number of the observations and the number of PCs.

It is possible to see that with 8 PCs and 40 observation for each class, so 160 observations for the dataset, a mean accuracy of 85% is obtained for the Test set. As referred in the Section 6.1.5, 17 moments are extracted from each image so the selection of 8 PCs produces a features vector with 136 elements. Table 6.2 shows in detail the mean accuracies and the corresponding standard deviations for each bearing obtained for the different number of observations per set-up with 8 PCs. The detailed results provide a complete view on the performance of the method for each bearing and show how increasing the number of observations per class leads to an increase of accuracy for each bearing. It is observable a common trend on the single bearing results, with an accuracy improvement at each observation increment. Only one case for BIF shows an opposite trend, an accuracy decrease in the step from 20 to 40 observation. This behavior is due to the specific observations added in the last increment that introduced some information that lead to an incorrect classification for this fault. In parallel with the overall increase of the mean accuracy, the standard deviation decreases, confirming the best performance is obtained with 40 observations per class.

Table 6.2: Classification Accuracy on Test Set for Each Bearings Under Stationary Conditions.

					Test Accuracy	
$n^{(a)}$	**BNF**	**BOF**	**BIF**	**BRF**	*Mean*	*STD*
8	37.50%	50.00%	37.50%	12.50%	34.38%	23.83
12	58.33%	50.00%	75.00%	75.00%	64.58%	21.53
20	85.00%	90.00%	85.00%	75.00%	83.75%	10.14
40	87.50%	90.00%	80.00%	85.00%	85.63%	9.99

(a) n is the number of observations per class in the Dataset

Table 6.3: Classification Accuracy on Test Set for Each Bearings Under Non-Stationary Conditions.

$n^{(a)}$	**BNF**	**BOF**	**BIF**	**BRF**	**Test Accuracy**
8	50.00%	81.25%	56.25%	37.50%	56.25%
12	87.50%	0.00%	93.75%	56.25%	59.38%
20	68.75%	68.75%	93.75%	25.00%	64.06%
40	75.00%	93.75%	87.50%	31.25%	71.88%

(a) n is the number of observations per class in the Dataset

6.4.2 Bearing faults classification with non-stationary signals

The proposed bearing faults classification technique is also suitable for providing accurate classification for signals recorded under variable rotating speed and load.

In this case, the classifiers trained with the stationary datasets introduced in the previous test are used to classify the signals measured under non-stationary conditions. Due to the best performance provided in the previous test, 8 PCs are maintained in this test to characterize the signals. For the training phase, the 4 datasets with a different number of observations for each class, respectively 8, 12, 20 and 40 observations, are used. The Test set is composed by all those signals measured under variable speed and load, in this case 16 observations are chosen for each class. This test does not involve a CV procedure since the Training and Test sets are fixed, so the performance is evaluated only in terms of accuracy on the Test set, and no standard deviation values are reported as in the stationary case study. Table 6.3 details the results obtained for each bearing.

This test confirms the good performance of the proposed technique for the

classification of faulty bearings. Obviously, the classification performance using non-stationary data for the test is inferior to that using the stationary data for both train and test, but the dataset with the highest number of observation per class in the Train set provide however an accuracy on Test set of 71%. By looking in detail the results for each bearing, it is not possible observe a common trend among the results. For each characteristic bearing fault the accuracy does not follow the increase of the number of observations. This behavior is different from the stationary case where a common trend is visible. The main reason for this difference is that with the non-stationary case the CV is not used and therefore the classification accuracies are not averaged. The lack of average causes the fluctuation of the single accuracies but at the same time it is possible to see that the overall accuracy shows an improvement at each increment of observations. It is also important to note that for BNF, BOF and BIF the performance are comparable with those obtained with the test on stationary data. For the BRF instead, the method provides an accuracy remarkably low and this accuracy drags down the overall accuracy. The changing of the operating conditions affects the frequency response of bearings and in the BRF case this leads the classification algorithm to misclassifying the fault.

6.5 Remarks

In this chapter, a condition monitoring solution for faulty bearings is exposed. The proposed solution applies the SVM classifier to classify the different faults that may occur on bearing. This solution has been developed to effectively work with bearing working under variable set-up of speed and load, for this reason it has been tested on experimental data coming from a bearing fault simulator and recorded with different speeds and loads. The proposed method has achieved a Test set accuracy of 85% with stationary signals and an accuracy of 75% with non-stationary signals. The obtained results demonstrate that the developed method is able to detect and to classy the faults present on roller bearing. Moreover, it has been demonstrated that this method is suitable for multiple rotating speeds and loads conditions as well as stationary and non-stationary signals.

Chapter 7

Quality Control for Electric Motors

DC electric motors are components widely used in many manufacturing industries, from home appliances to automotive. Given the widespread diffusion of commonly used objects and instruments, it is essential to ensure low levels of vibration and noise in order to provide high quality standards to the final consumer. For this reason, in the past few years, the standards and limits for the homologation of component and equipment have become more stringent in every industrial sector. Such components are produced in large quantities and with low production costs, it is therefore impossible to guarantee all the quality tests set by law for each engine produced. Testing takes time and often requires proper structures and equipment, making them unsuitable for integration within a production line.

The most common solution for the end-of-line quality control is to rely on the experience of qualified operators who manually check the status of the motors. However, this type of test is characterized by little repeatability as the evaluation is influenced by the operator's sensitivity and perception, by the background noise of the environment, and by his psycho-physical state. In order to ensure a more reliable quality control and to test each element produced by the line, it is necessary to replace the human operator's control with an automatic system that allows an automatic test of the component in a limited time interval to impact as less as possible on the production time of the line.

Due to the importance of the subject, over the years, many solutions have been presented to face the task of fault detection and quality control on electric motor. The solutions adopt a wide range of very different techniques but that can be divided into two main categories: data-driven methods and model-based methods. Model-based methods use mathematical models that describe motor behavior by using the physical relationship between the stimuli applied to the motor and its responses. An exact model allows to characterize every condition of the component and to identify if there are any faults, on the base of the measured parameters. The main problem of this approach is the a priori knowledge required to describe the motor with a proper mathematical model. It is a

complex system, subject to many variables that change its characteristics over time due to wear and changing operating conditions. It is therefore difficult to find a fixed relationship between the available parameters and the motor status. The second class of methods uses the data extracted from the motor to characterize its state. These methods use signal analysis and ML techniques to train expert systems able of detecting faulty motors. The training of the expert systems use the signals measured from the device under test, and needs only the knowledge of the motor condition, faulty or non-faulty, that is easily derivable from test performed by human operators. The main drawback for using data-based methods lies on the amount of needing data that must be collected to properly train the system. This difficult has been partially overcome in recent years thanks to the introduction of more accurate and economical sensors and measurement systems integrated within the production lines. The availability of more data facilitated the implementation of data-based methods that use both self-supervised and non-supervised techniques for the training. Many existing work in the field of fault detection for electric motors and rotary machines apply supervised classification techniques for detecting faulty motors and classification of fault types [82, 83, 10]. However, the impossibility to know any type of fault that may occur during the lifetime of an electric motor moves the interest in seeking un-supervised approaches for error detection, such as clustering and one-class classification [84, 85, 86]. Among the non-supervised methods, an alternative is the Novelty Detection approach. This approach has the goal of identifying new or unknown data that the machine learning model has not been trained with. The novelty approach has been used to develop many systems in different application fields,and in general in those contexts where the amount of normal data consists of a very large set, and the normal class can be accurately modelled, whereas the events not belonging to the normality are considered as novel events.

7.1 A Novelty Approach for Electrical Motors Fault Detection

The work presented in this chapter utilizes an un-supervised approach to provide a monitoring tool for the end-line quality control. The goal is to identify faulty motors and then discard them. The developed technique is based on Denoising Autoencoders (DEA) with Long Short-Term Memory (LSTM) Recursive Neural Network. The network is used as a one-class classifier to detect faulty motors, regardless of the type of defect. The approach is based on the evaluation of the reconstruction error committed by the trained network when it attempts to reconstruct a signal that does not belong to the dataset used dur-

ing the training phase. The training dataset has been populated with signals collected from motors that have been classified un-faulty by human operators, in order to create a reference background model.

The presence of construction defects in the motor causes an abnormal vibration if compared with the normal vibration emitted by a non-defective motor. For this reason, the vibration signals have been measured and used to extract useful information about the status of the motor. Measured signals have been appropriately processed and two sets of features have been extracted for use in network training. Log Mel Coefficients and Mel Frequency Cepstrum Coefficients (MFCC) are set of parameters widely used for speech analysis [1, 87, 88], but some works presented in literature have successfully applied these coefficients for the study of vibration signals [89, 90, 91]. A set of tests have been designed to compare the results obtained with each of these two feature sets.

7.2 Mel Spectrum and Mel-Frequency Cepstral Coefficients

The MFCC [92] are a representation of the short-term power spectrum of a signal, based on a linear cosine transform of a log power spectrum on a non-linear mel scale of frequency. The first step for the coefficients extraction is the computation of the frequency domain representation of the input signal by a Discrete Fourier Transform:

$$X^i(k) = \sum_{n=0}^{N-1} x^i(n) \cdot exp(-i2\pi kn/N), \tag{7.1}$$

for the i^th frame of the signal, where N is number of sampling points for the signal $x(n)$, and $k = 0, 1, ..., (N/2) - 1$.

The second processing step is the computation of the mel-frequency spectrum. Therefore, the spectrum is filtered with N_f different band-pass filters and the power of each frequency band is computed. This processing step is described by:

$$S^i(j) = \sum_{k=0}^{(N/2)-1} X^i(k) \cdot F_j(k), \tag{7.2}$$

where $F_j(k)$ is the amplitude of the band-pass filter with the index j, and $j = 0, 1, ..., N_f$. For this study $N_f = 26$. Then the logarithm of mel-frequency spectrum is computed:

$$L^i(j) = log[S^i(j)] \tag{7.3}$$

The results of this step are the Log Mel coefficients used as first features set

for the proposed work.

The final step converts the logarithmic Mel spectrum back to the time domain. This conversion is achieved by taking the Discrete Cosine Transform of the spectrum as:

$$C_m^i = \sum_{j=0}^{N_f} L^i(j) \cdot cos\left\{\frac{m(2j-1)\pi}{2N_f}\right\}, \tag{7.4}$$

where $m = 0, 1, ..., N_c$, and N_c is the number of chosen cepstral coefficients, in this study case $N_c = 13$. Together with the MFCC also the delta and the delta delta coefficients are computed and used as features:

$$\Delta C_m^i = C_m^{i+1} - C_m^{i-1}, \tag{7.5}$$

$$\Delta\Delta C_m^i = \Delta C_m^{i+1} - \Delta C_m^{i-1}, \tag{7.6}$$

7.3 The Acquisition System and Signal Processing

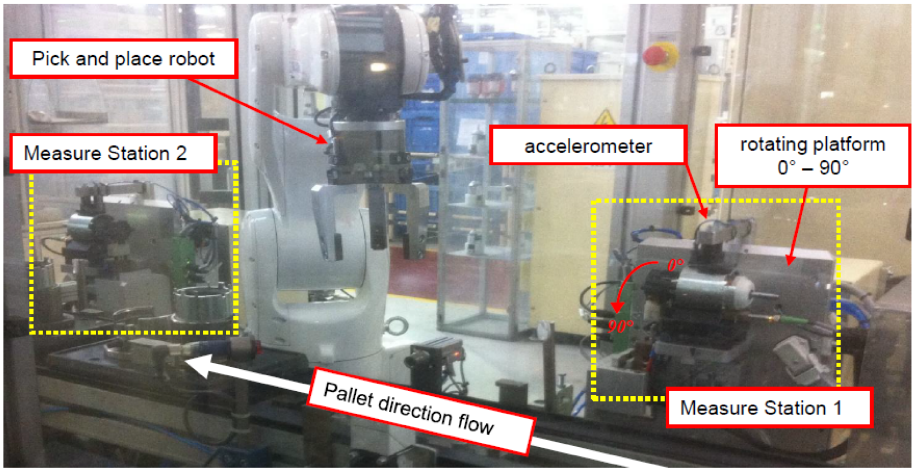


Figure 7.1: Measuring bench for the end-line motors quality check.

The signals used for this study derive from an actual industrial control system placed at the end of a DC motor production line of a global automotive components manufacturer. The system is composed by a robotic arm that picks and moves the motor under test and two measuring stations that work in parallel. The Figure 7.1 shows the actual measuring bench installation. The standard procedure for the motor quality check that involves the human

operator is: taking a motor from a quality inspection bin; performing visual inspection of the motor and check if all components are assembled correctly; connecting power cable; checking that the shaft rotates; rotating the motor in order to assess noise and vibration at different tilt angles; disconnecting the power cable; assigning a verdict (pass or fail). The weight of the rotor creates an unbalance force when the rotor axis is not aligned with gravity, so it is important to test the motor in different tilt angles, paying particular attention to the angle at which the motor will be mounted in the end product. The automatic control system mimic the human operator behaviour and follows the following steps for its measurement procedure:

1. The robot takes the piece from the line and places it horizontally on the "cradle" of one of the 2 measuring stations;
2. The clamp is closed that is lowered by rotating 90° to keep the engine in the cradle. In the upper part of the clamp there is a single-axis PCB 352C33 accelerometer (sensitivity of 100 mV/g) with IEPE technology. The pressure of the pneumatic cylinder that moves the clamp is monitored and kept constant in order to have a constant motor-cradle coupling force and thus increasing the repeatability of the measure;
3. During the step 2, the entire measurement area is released so that it remains floating on pneumatic suspension so as to avoid vibration transmission from the floor;
4. A 3V power supply is supplied with a TDK lambda ZUP 20-20 power supply, driven by an analogue output of a National Instrument PCI express 6351 board. At this voltage, the motor accelerate up to 1000 rpm. A PID controller regulates the voltage so that it has a speed of 1000 ± 5 rpm. It is therefore possible to assume that the tests take place at constant speed of 1000 rpm;
5. Once the speed is stable, the actual test in which vibrations are measured starts. During the test, the motor is rotated 90°, from the horizontal to the vertical position in 3s. This rotation is necessary because some motors are noisy only horizontally, others only in vertical, it depends on the defect;
6. After the test is concluded, the motor returns quickly horizontally, the clamp is opened, the robot takes it and puts it back on the line. In the case of a faulty motor, the motor is retested, if the fault is confirmed it is definitely classified as faulty.

The acceleration signals are acquired for each station with the NI PCIe 6351 with a sample rate of 25,6 kS/s, moreover the accelerometer channel passes

through an signal conditioner IEPE MMF M32, which has an anti-aliasing filter at 10kHz.

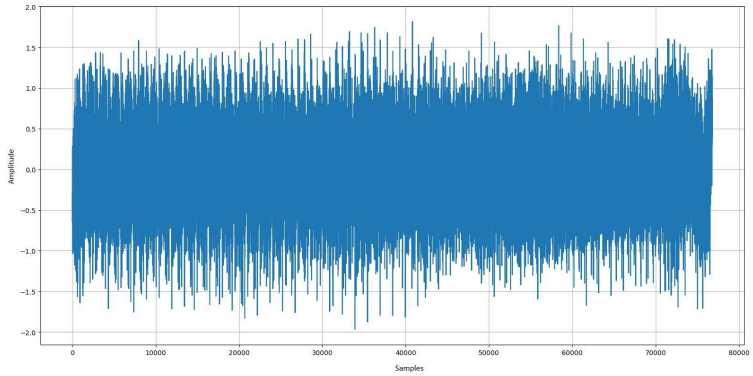


Figure 7.2: Example of Time Vibration Signal.

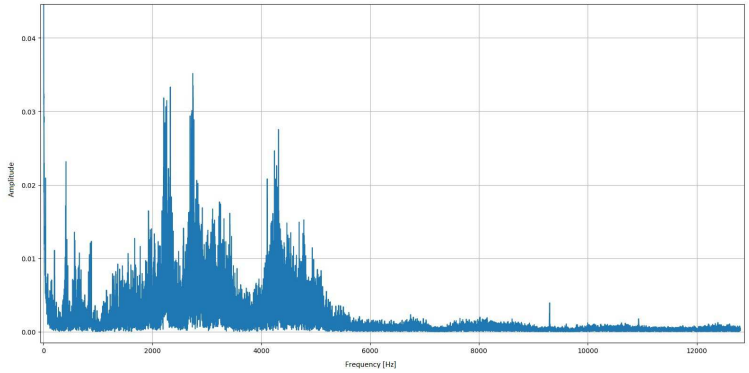


Figure 7.3: Example of Frequency Spectrum.

From each acquisition, the signal of 3s corresponding to the time interval during which the motor kept constant speed is extracted. An example of time signal is shown in Figure 7.2. Instead, in Figure 7.3 the frequency spectrum of the same signal is shown. How it is possible to appreciate, useful component up to a frequency of 6 KHz, for this reason the signals were down-sampled at 12.8 KS/s in order to maintain only the portion of spectrum containing information. From each raw signal, 3 s have been extracted and framed into frame of 600ms, with a 50% overlap, so as to get 10-frame set for each acquisition. The feature extraction is made frame by frame. For the Log Mel coefficients, 26 Mel filters

have been selected so 26 features characterize each frame. For the MFCC, 13 coefficients have been extracted together with delta and delta-delta coefficients for a total of 39 features.

7.4 Experimental Setup

Several tests have been carried out to find the most suitable network topology for the problem dealt with. The networks have been trained with the ADAM algorithm [30] by keeping the configuration parameters suggested by the original paper: *learning rate* = 0.001, $beta1 = 0.9$, $beta2 = 0.999$ and $\epsilon = 10^{-8}$. Given the use of DAE networks, different noise sigma values $\sigma = [0.1, 0.25, 0.5]$ have been evaluate. Different network topologies have been tested, by using 2 hidden layers as basic structure, and varying the LSTM units from 26-52 to 78-104 for each layer with Log Mel coefficients and from 39-78 to 117-156 with MFCC. The input and output layers of the network have a constant number of units, N = 26 equal for Log Mel feature set and N = 39 for MFCC. This allows the network to reconstruct the input feature set and to detect faulty motor signals from the reconstruction error. The reconstruction error is evaluated by computing the Euclidean distance between the original normalized feature vector and the network's output, the distances are summed up and the sum is averaged by the number of features. The result is a single parameter representing the error for each frame. The novelty detection have been done for each frame sequence, so a decision threshold has been set in order to obtain a binary signal. The classification of a sequence as faulty or non-faulty depends on the number of frames that exceed the threshold. Although each sequence is associated with a single condition (faulty or non-faulty), it is unlikely that all frames in the sequence have the same error. During the signal recording phase, the robotic arm turns the motor and certain abnormal vibrations can be highlighted only for few time instants. For this reason, some tests have been carried on to identify the minimum number of frames over the error threshold that identify a faulty motor. This parameter, identifies as F for brevity, has been varied from 2 up to 8 frames. The results obtained with the Denoising Autoencoder network were compared with those obtained with OC-SVM. OC-SVM models have been trained with Radial Basis Function as kernel and different values of γ and ϵ parameters: $\gamma = [2^0, ..., 2^{-5}]$, and $\epsilon = [2^{-1}, ..., 2^{-4}]$.

7.5 Dataset

The dataset used for this study contains signals measured at the end of a motors production line. Each motor is checked and a vibration signal is acquired to

evaluate its condition. The dataset contains 1178 signals associated with non-faulty motors and 14 signals that human operators have classified as defective. The available data have been subdivided into a Train dataset containing only non-defective motor signals, and a Test set containing signals of both categories. The Train set consists of 1170 non-defective signals, while the Test set consists of 14 faulty signals and 8 non-faulty signals.

7.6 Computer Simulations and Results

The tests used to evaluate the performance of the proposed method have been performed with the KERAS tools implemented in PYTHON and running on a PC with a processor i7 quad-core 2.3 GHz, 8 GB of RAM and Windows 10® OS.

Figure 7.4 shows the reconstruction error for 4 example sequences. As it is possible to appreciate, the level of reconstruction error is higher for faulty sequences and many frames have an error above the decision threshold. This shows how the DAE network commits a higher reconstruction error by trying to reconstruct faulty motor signals.

Figure 7.4: Example of residual error for 4 different sequences associated to 2 faulty motors and 2 normal motors.

For this study, the Receiver Operating Curve (ROC) has been used to evaluate the behavior of the trained networks at the variation of the decision threshold and the corresponding Area Under Curve (AUC) has been used as evaluation metric. The set of thresholds used to create the ROCs corresponds to the set of network reconstruction errors get as result by the background model.

In Table 7.1 the results obtained with different values of error sigma are compared. The network topology is kept same for all test as well as the minimum

Table 7.1: AUC for the two features set varying the amount of noise σ used in the training phase.

		Noise Variance		
Features Set	**Topology**	**0.1**	**0.25**	**0.5**
Log Mel	52-52	**0.95**	0.83	0.83
MFCC	78-78	**0.66**	0.62	0.63

Table 7.2: Comparison of AUC values for different network topology.

Features Set	**Topology**	**AUC**
Log Mel	26-52	0.84
	52-52	**0.95**
	52-78	0.84
	78-104	0.74
MFCC	39-78	0.65
	78-78	**0.66**
	78-117	0.62
	117-156	0.63

number of frames to have above the decision threshold to identify a faulty motor. The results show that for both the feature sets the best performance are achieved with $\sigma = 0.1$, with the Log Mel that outperforms the MFCC for each noise value.

A comparison among different topologies of the network is reported in Table 7.2. For both the features set, the best configurations use for each hidden layer a number of LSTM units equal to double of feature number.

The results obtained for different thresholds for the minimum number of frames F for detecting faulty motor is given in Table 7.3, for both DAE and OC-SVM. Also from this set of results, it is possible to appreciate better performance with Log Mel feature. With the DAE, the MFCC reach its best performance of AUC = 0.79 with $F \geq 8$, while for the Log Mel set $F \geq 4$ and $F \geq 6$ provide the best result of AUC = 0.95. The OC-SVM provides AUC = 0.68 with $F \geq 8$ for the MFCC feature set, and AUC = 0.82 with $F \geq 2$ with the Log Mel coefficients. The results demonstrate that with the Log Mel features it is possible to extract information from the raw signal that highlight the difference between faulty and non-faulty signals, and this lead to both an overall best performance and minor number of needing frames to distinguish faulty sequences from the background.

Table 7.3: Comparison of AUC values for different values of the F for DAE and OC-SVM.

		F = Number of Frames						
Feature Set	**Algorithm**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Log Mel	DAE	0.84	0.85	**0.95**	0.81	**0.95**	0.82	0.80
MFCC	DAE	0.70	0.71	0.63	0.67	0.66	0.71	**0.79**
Log Mel	OC-SVM	**0.82**	0.75	0.73	0.71	0.71	0.76	0.76
MFCC	OC-SVM	0.63	0.62	0.64	0.63	0.59	0.57	**0.68**

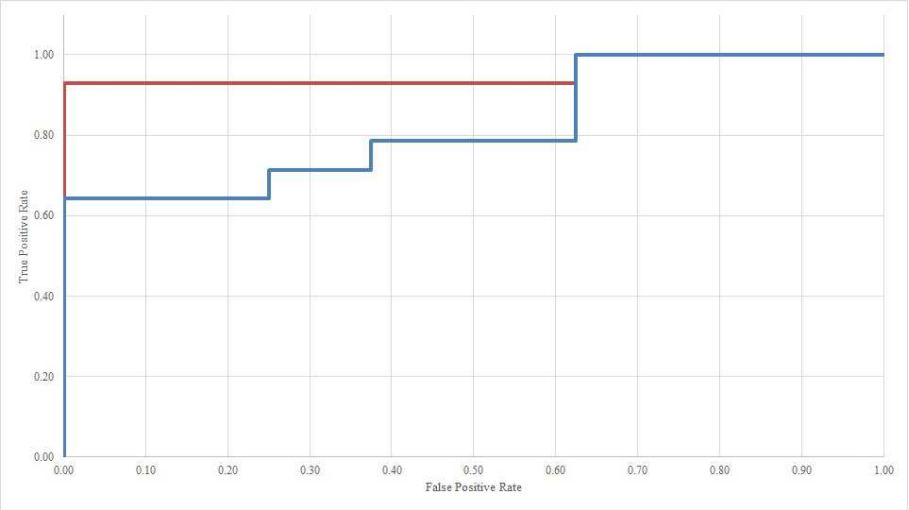


Figure 7.5: Comparison af two ROCs obtained with DAE network (red) and OC-SVM (blue).

Figure 7.5 shows the ROCs obtained with the bets configuration of DAE and OC-SVM. How it is possible to see, the DAE outperforms the OC-SVM, by providing an AUC of 0.95, instead the OC-SVM obtains an AUC of 0.82. The better performance of a DAE is due to in its ability of encoding the input by preserving the information about the input itself and simultaneously undoing the effect of a corruption process applied to the input of the auto-encoder. The combination of these two learning processes seems to be effective in this task.

7.7 Remarks

In this chapter a new solution for the end-line quality check for DC motor has been proposed. This solution has been designed to identify the faulty motors,

regardless the kind of faults, by using the novelty detection approach. DAE network has been employed to train a suitable model that characterize the normality of the system, i.e. the normal motors. Differently from a classification approach, the proposed solution uses less a-priori information because it does not need the knowledge of every kind of motor fault. Two features set have been compared, the first one containing Log Mel coefficients and the second composed by MFCC. Different tests have been carried out to find the best configuration for the DAE network, and a further comparison with an un-supervised method widely used in literature for novelty approach has been performed, i.e. OC-SVM. The proposed algorithm with its best configuration and the Log Mel features set achieves an AUC of 0.95, demonstrating that this novelty method is suitable for the detection of faulty DC motors. Moreover, in the comparison with the DAE outperforms the OC-SVM that achieves for its best configuration an AUC of 0.82.

Chapter 8

Conclusions

This thesis explored possible methodologies to solve industrial relevant problems related to quality checks and condition monitoring by using a machine learning approach. The variety of problems that have been solved by means of the developed algorithms demonstrates that machine learning techniques are effective tools to achieve the goal of monitoring quality of products along the production line and to monitor condition of industrial equipments during working condition in order to optimize their performances. With the purpose of demonstrate the effectiveness of the proposed algorithms on data coming from real situations, the datasets used for the tests were composed by experimental data collected from industrial working systems and devices.

The first case is related to the estimation and classification of particle size of powder employed in industrial processes. This problem has been analysed under different aspect with the final goal of designing a methodology for the combustion optimization of coal burner in a coal fired power plant. The standard procedure for the PSD estimation within an industrial power plant is the sampling and sieving method that can provide with a precise measure of the PSD. However, this method is time consuming and it is not suitable for a continuous on-line monitoring of the powder size. In Chapter 4, the author propose a new solution that applies a ML technique to perform a continuous non-invasive PSD monitoring, and so overcoming the limits of the sampling and sieving method. Different kind of ML algorithms, i.e. ANN, SVR and ELM, have been tested for this solution. The obtained results have been compared with those obtained with the system POWdER, an industrial system that perform the PSD monitoring, and the new solution has proven to obtain better performance than the reference system. Moreover, a further improvement for the PSD estimation is reached by using heterogeneous data coming from different industrial plant. The proposed solution has proven to be able to exploit data from multiple source, this aspect is particularly important in an industrial context where it can be not easy collecting enough data for the training of the expert monitoring system. The Chapter 5 proposes a new scheme for a non-invasive condition monitoring approach based on classification for burners

devoted to coal powder combustion in the same industrial power plant of the previous case study. The usual procedure to perform the efficiency estimation the burner/boiler system considers the measurement of the produced energy and the produced ash to asses if the system is operating under optimal condition. With this procedure it is possible to act only after the end of the energy production process. The new proposed solution uses the algorithms previously implemented for the PSD estimation modified to fulfil a classification task. The size of the powder that feeds the burners is continuously monitored and classified in order to immediately identify non-optimal operating conditions. The proposed method has the great advantage of using a-priori knowledge already available in an industrial power plant, as is the energy efficiency of the burner combustion, to avoid further data collection. This aspect is particularly relevant because represents a substantial time saving during the phase of system set-up. Moreover, despite this method is applied on a specific task 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.

In Chapter 6 a maintenance tool for roller bearings is proposed with the objective of detecting and classifying different fault types that occur on this device. The available techniques that deal with the problem of bearing fault identification perform the analysis on vibration signals measured under stationary conditions of speed and load. The new proposed technique applies images analysis together with the SVM classifier and it is designed and tested on roller bearing operating in different working conditions. The suitability of this method for multiple rotating speeds and loads conditions as well as stationary and non-stationary signals represents an important feature for a monitoring system devoted to control of the bearing status, because it is able to asses the device conditions under every operating set-up.

The last addressed problem is illustrated in Chapter 7. The Novelty Detection approach is applied for the end-line quality control of DC motors. Conversely to the standard classification techniques that require the knowledge of all the different faults that may occur, the proposed novelty approach use the characterization of the normality, i.e. the motors without faults, a task relatively simple due to the great amount of working motor available during the production phase, to detect the faulty motors among all the produced motor. This aspect makes the proposed quality control system able to distinguish the faulty motors regardless the type of fault present on the device, and without the informations about the possible faults. A Denoising Autoencoder network with Long Short-Term Memory modules has been used to implement the algorithm that exploits features extracted from vibration signals for the model training.

In conclusion, the solutions proposed in this thesis for actual industrial issues have demonstrated, trough the tests on proper experimental data, that ML techniques can be effectively applied for the implementation of industrial CM systems, suitable to be used for system monitoring in maintenance and quality control processes.

8.1 Future Works

With the spread of more advanced monitoring and control system in manufacturing, following the trend of automation and data exchange of Industry 4.0, the author is confident that more and more amount of data will be available in the future. The more experimental dataset the more practical issues could be possible to address by means of expert systems based on ML tools

With this idea, future works will be oriented mainly toward two direction.

For the first one, new cases study of the same kind of those addressed in this thesis will be analysed with the purpose of understand the generalization degreed of the proposed solutions. Additionally, a further in-depth study is needed to find optimal ML technique and parameters for each specific issue.

A second research direction will seek to find diverse industrial and manufacturing problems where to develop advanced monitoring solution able to improve the performance of the systems in terms of reliability, security and quality.

List of Publications

[1] **D. Rossetti**, S. Squartini, and S. Collura, "Machine Learning Techniques for the Estimation of Particle Size Distribution in Industrial Plants", *Advances in Neural Networks*, vol. 54, pp. 355 – 343, 2016.

[2] S. Collura, M. Dumont, **D. Rossetti**, and T. Urbank, "Optimized ATEX Acoustic Emission Measuring Chains for Particle Size Monitoring in Industrial Plants", *19th World Conference on Non-Destructive Testing*, 13-17 June 2016, Munich, Germany.

[3] **D. Rossetti**, S. Squartini, S. Collura, and Y. Zhang, "Exploiting Heterogeneous Data for the Estimation of Particles Size Distribution in Industrial Plant", *17th International Conference on Mechatronics*, 7–9 December 2016, Prague, Czech Republic.

[4] **D. Rossetti**, Y. Zhang, S. Squartini, and S. Collura, "Classification of Bearing Faults Through Time-Frequency Analysis and Image Processing", *17th International Conference on Mechatronics*, 7–9 December 2016, Prague, Czech Republic.

[5] **D. Rossetti**, S. Squartini, S. Collura, and Y. Zhang, "Power Plant Condition Monitoring by Means of Coal Powder Granulometry Classification", *Journal of the International Measurement Confederation*, minor revisions required.

Bibliography

[1] Erik Marchi, Fabio Vesperini, Florian Eyben, Stefano Squartini, and Björn Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks," *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, 04 2015.

[2] Joel Levitt, *The Complete Guide to Preventive and Predictive Maintenance*, Industrial Press, New York, New York, NY, USA, 2003.

[3] Ilya B. Gertsbakh, *Reliability Theory with Applications to Preventive Maintenance*, Springer-Verlag, Berlin, Berlin, Germany, 2000.

[4] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483 – 1510, 2006.

[5] R. Yam, P. Tse, and L. Li, "Intelligent predictive decision support system for condition-based maintenance," *The International Journal of Advanced Manufacturing Technology*, vol. 17, no. 5, pp. 383–391, February 2001.

[6] Patricia Melin and Oscar Castillo, "An intelligent hybrid approach for industrial quality control combining neural networks, fuzzy logic and fractal theory," *Information Sciences*, vol. 177, no. 7, pp. 1543 – 1557, 2007, Fuzzy set applications in industrial engineering.

[7] Liqun Hou and Neil W. Bergmann, "Novel Industrial Wireless Sensor Networks for Machine Condition Monitoring and Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 10, pp. 2787 – 2798, 2006.

[8] C. Ruiz-Cárcel, V.H. Jaramillo, D. Mba, J.R. Ottewill, and Y. Cao, "Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions," *Mechanical Systems and Signal Processing*, vol. 66-67, no. Supplement C, pp. 699 – 714, 2016.

[9] Tshilidzi Marwala, *Condition Monitoring Using Computational Intelligence Methods*, Springer-Verlag London, London, United Kingdom, 2012.

[10] V. N. Ghate and S. V. Dudul, "Cascade neural-network-based fault classifier for three-phase induction motor," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1555 – 1563, 2011.

[11] J. Rafiee, F. Arvani, A. Harifi, and M.H. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1746 – 1754, 2007.

[12] Xianghui Huang, T.G. Gabetler, and R.G. Harley, "Detection of rotor eccentricity faults in closed-loop drive-connected induction motors using an artificial neural network," *IEEE Transactions on Power Electronics*, vol. 22, pp. 1552 – 1559, July 2007.

[13] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72-73, no. Supplement C, pp. 303 – 315, 2016.

[14] M. Fast and T. Palmé, "Application of artificial neural networks to the condition monitoring and diagnosis of a combined heat and power plant," *Energy*, vol. 35, no. 2, pp. 1114 – 1120, 2010, ECOS 2008.

[15] Zhigang Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *Journal of Intelligent Manufacturing*, vol. 23, no. 2, pp. 227–237, Apr 2012.

[16] Achmad Widodo and Bo-Suk Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560 – 2574, 2007.

[17] Bo-Suk Yang, Won-Woo Hwang, Dong-Jo Kim, and Andy Chit Tan, "Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines," *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 371 – 390, 2005.

[18] Liu Han, Liu Ding, Jin Yu, Qi Li, and Yanming Liang, *Power Plant Boiler Air Preheater Hot Spots Detection System Based on Least Square Support Vector Machines*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[19] L.B. JACK and A.K. NANDI, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 16, no. 2, pp. 373 – 390, 2002.

[20] Uwe Kruger, Q Chen, D J. Sandoz, and R C. McFarlane, "Extended pls approach for enhanced condition monitoring of industrial processes," *AIChE Journal*, vol. 47, pp. 2076 – 2091, 09 2001.

[21] W. H. Allen, A. Rubaai, and R. Chawla, "Fuzzy neural network-based health monitoring for hvac system variable-air-volume unit," *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 2513–2524, May 2016.

[22] S. A. Ethni, S. M. Gadoue, and B. Zahawi, "Inter-turn short circuit stator fault identification for induction machines using computational intelligence algorithms," *2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 757–762, March 2015.

[23] Jesus A. Carino, Miguel Delgado Prieto, Daniel Zurita, Marta Millán, Juan Antonio Ortega Redondo, and René de Jesús Romero-Troncoso, "Enhanced industrial machinery condition monitoring methodology based on novelty detection and multi-modal analysis," *IEEE Access*, vol. 4, pp. 7594–7604, 2016.

[24] D. P. Filev, R. B. Chinnam, F. Tseng, and P. Baruah, "An industrial strength novelty detection framework for autonomous equipment monitoring and diagnostics," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 4, pp. 767 – 779, 2010.

[25] Guang Wang and Jianfang Jiao, "A kernel least squares based approach for nonlinear quality-related fault detection," *IEEE Transactions on Industrial Electronics*, vol. PP, pp. 1–1, 12 2016.

[26] L. Clifton, D. A. Clifton, Y. Zhang, P. Watkinson, L. Tarassenko, and H. Yin, "Probabilistic novelty detection with support vector machines," *IEEE Transactions on Reliability*, vol. 63, no. 2, pp. 455–467, June 2014.

[27] Warren S. McCulloch and Walter Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[28] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, ," in *Neurocomputing: Foundations of Research*, James A. Anderson and Edward Rosenfeld, Eds., chapter Learning Representations by Back-propagating Errors, pp. 696–699. MIT Press, Cambridge, MA, USA, 1988.

[29] Yoshua Bengio, "Learning deep architectures for ai," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[30] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[31] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, New York, NY, USA, 2008, ICML '08, pp. 1096–1103, ACM.

[32] P.W.D. Charles, "Project title," 2013.

[33] Corinna Cortes and Vladimir Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273 – 297, 1995.

[34] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Secaucus, NJ, USA, 2006.

[35] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

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

[37] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, July 2001.

[38] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme Learning Machine: Theory and Applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[39] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 2, pp. 513–529, April 2012.

[40] Stefano Collura, D Possanzini, M Gualerci, L Bonelli, and D Pestonesi, "Coal mill performances optimization through non-invasive online coal fineness monitoring," in *Powergen, Wien*, 2013.

[41] Mallat Stéphane, *Chapter 8 - Wavelet Packet and Local Cosine Bases*, Academic Press, Boston, third edition edition, 2009.

[42] Alessandro Bastari, Cristina Cristalli, Roberto Morlacchi, and Eraldo Pomponi, "Acoustic emissions for particle sizing of powders through signal processing techniques," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 901 – 916, 2011.

[43] M.F. Leach, G.A. Rubin, and J.C. Williams, "Particle size determination from acoustic emissions," *Powder Technology*, vol. 16, no. 2, pp. 153 – 158, 1977.

[44] M.F. Leach, G.A. Rubin, and J.C. Williams, "Particle size distribution characterization from acoustic emissions," *Powder Technology*, vol. 19, no. 2, pp. 157 – 167, 1978.

[45] Alberto Boschetto and Fabrizio Quadrini, "Powder size measurement by acoustic emission," *Measurement*, vol. 44, no. 1, pp. 290 – 297, 2011.

[46] M. Uher and P. Benes, "Measurement of particle size distribution by the use of acoustic emission method," in *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, May 2012, pp. 1194–1198.

[47] Yonghui Hu, Xiaobin Huang, Xiangchen Qian, Lingjun Gao, and Yong Yan, "Online particle size measurement through acoustic emission detection and signal analysis," in *Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2014 IEEE International*, May 2014, pp. 949–953.

[48] E. Pasolli, F. Melgani, N. Alajlan, and Y. Bazi, "Active learning methods for biophysical parameter estimation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 10, pp. 4071–4084, Oct 2012.

[49] Wenbin Cai, Ya Zhang, and Jun Zhou, "Maximizing expected model change for active learning in regression," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, Dec 2013, pp. 51–60.

[50] Huimin Chen and X.R. Li, "Distributed active learning with application to battery health management," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, July 2011, pp. 1–7.

[51] Xiaohang Ren, Kai Chen, Xiaokang Yang, Yi Zhou, Jianhua He, and Jun Sun, "A new unsupervised convolutional neural network model for chinese scene text detection," in *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*, July 2015, pp. 428–432.

[52] Christian Plahl, Tara N Sainath, Bhuvana Ramabhadran, and David Nahamoo, "Improved pre-training of deep belief networks using sparse encoding symmetric machines," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4165–4168.

[53] S.R. Gangireddy, F McInnes, and S Renals, "Feed forward pre-training for recurrent neural network language models," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 2620–2624, 01 2014.

[54] Luca Pasa and Alessandro Sperduti, "Pre-training of recurrent neural networks via linear autoencoders," in *Advances in Neural Information Processing Systems*, 2014, pp. 3572–3580.

[55] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2524, 2013.

[56] Darrin P Lewis, Tony Jebara, and William Stafford Noble, "Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure," *Bioinformatics*, vol. 22, no. 22, pp. 2753–2760, 2006.

[57] K. Nurzynska, M. Kubo, and K. i. Muramoto, "Snow particle automatic classification with texture operators," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, July 2011, pp. 2892–2895.

[58] K. K. Pahalawatta and 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)*, Nov 2013, pp. 1–8.

[59] F. Previdi, S. M. Savaresi, P. Guazzoni, and L. Zetta, "Light charged particle classification using subspace identification methods and neural networks," in *2006 American Control Conference*, June 2006, pp. 6 pp.–.

[60] C. Yibo and 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"*, Oct 2006, vol. 2, pp. 2010–2014.

[61] D. Breitgand, M. Goldstein, E. Henis, and O. Shehory, "Performance management via adaptive thresholds with separate control of false positive and false negative errors," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*, June 2009, pp. 195–202.

[62] S. Yusuf, D. J. Brown, A. Mackinnon, and R. Papanicolaou, "Fault classification improvement in industrial condition monitoring via hidden markov models and naive bayesian modeling," in *2013 IEEE Symposium on Industrial Electronics Applications*, Sept 2013, pp. 75–80.

[63] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, May 2016.

[64] Pat O'Donnell, "Report of large motor reliability survey of industrial and commercial installations, part i," *IEEE Transactions on Industry Applications*, vol. IA-21, no. 4, pp. 853–864, July 1985.

[65] O. V. Thorsen and M. Dalva, "A survey of faults on induction motors in offshore oil industry, petrochemical industry, gas terminals and oil refineries," in *Petroleum and Chemical Industry Conference, 1994. Record of Conference Papers., Institute of Electrical and Electronics Engineers Incorporated Industry Applications Society 41st Annual*, Sep 1994, pp. 1–9.

[66] O. V. Thorsen and M. Dalva, "Failure identification and analysis for high-voltage induction motors in the petrochemical industry," *IEEE Transactions on Industry Applications*, vol. 35, no. 4, pp. 810–818, Jul 1999.

[67] Akhand Rai and S.H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, no. Supplement C, pp. 289 – 306, 2016.

[68] Dong Wang, Peter W. Tse, and Kwok Leung Tsui, "An enhanced kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1, pp. 176 – 199, 2013.

[69] Yanxue Wang and Ming Liang, "An adaptive sk technique and its application for fault detection of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1750 – 1764, 2011.

[70] Hongkun Li, Fujian Xu, Hongyi Liu, and Xuefeng Zhang, "Incipient fault information determination for rolling element bearing based on synchronous averaging reassigned wavelet scalogram," *Measurement*, vol. 65, no. Supplement C, pp. 1 – 10, 2015.

[71] Peter W. Tse and Dong Wang, "The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "two automatic vibration-based fault diagnostic methods using the novel sparsity measurement – parts 1 and 2"," *Mechanical Systems and Signal Processing*, vol. 40, no. 2, pp. 499 – 519, 2013.

[72] Choon-Su Park, Young-Chul Choi, and Yang-Hann Kim, "Early fault detection in automotive ball bearings using the minimum variance cep-

strum," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 534 – 548, 2013.

[73] Yuh-Tay Sheen, "An envelope analysis based on the resonance modes of the mechanical system for the bearing defect diagnosis," *Measurement*, vol. 43, no. 7, pp. 912 – 934, 2010.

[74] Matej Žvokelj, Samo Zupan, and Ivan Prebil, "Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1049 – 1067, 2010.

[75] Wei Li, Mingquan Qiu, Zhencai Zhu, Bo Wu, and Gongbo Zhou, "Bearing fault diagnosis based on spectrum images of vibration signals," *Measurement Science and Technology*, vol. 27, no. 3, pp. 035005, 2016.

[76] Liang Hua, Yujian Qiang, Juping Gu, Ling Chen, Xinsong Zhang, and Hairong Zhu, "Mechanical fault diagnosis using color image recognition of vibration spectrogram based on quaternion invariable moment," *Mathematical Problems in Engineering*, vol. 2015, pp. 11, 2015.

[77] Renata Klein, Eyal Masad, Eduard Rudyk, and Itai Winkler, "Bearing diagnostics using image processing methods," *Mechanical Systems and Signal Processing*, vol. 45, no. 1, pp. 105 – 113, 2014.

[78] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[79] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.

[80] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck, *Discrete-time Signal Processing (2Nd Ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.

[81] Ming-Kuei Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962.

[82] A. Adouni, A. Abid, and L. Sbita, "A dc motor fault detection, isolation and identification based on a new architecture artificial neural network," *2016 5th International Conference on Systems and Control (ICSC)*, pp. 294–299, May 2016.

[83] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (svms)," *Applied Soft Computing*, vol. 11, no. 6, pp. 4203 – 4211, 2011.

[84] Roozbeh Razavi-Far, Maryam Zanjani, Shokoofeh Zare, Mehrdad Saif, and Jafar Zarei, "One-class classifiers for detecting faults in induction motors," *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–5, April 2017.

[85] Abdenour Soualhi, Guy Clerc, and Hubert Razik, "Detection and Diagnosis of Faults in Induction Motor Using an Improved Ant Clustering Technique," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 4053 – 4062, Sep 2013.

[86] Hyun Cheol Cho, Jeremy Knowles, Sami Fadali, and Kwon Soon Lee, "Fault detection and isolation of induction motors using recurrent neural networks and dynamic bayesian modeling," *Control Systems Technology, IEEE Transactions on*, vol. 18, pp. 430 – 437, 04 2010.

[87] Michael Seltzer, Dong Yu, and Yongqiang Wang, "An investigation of deep neural networks for noise robust speech recognition," *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, pp. 7398–7402, 10 2013.

[88] Seyed Reza Shahamiri and Siti Salwah Binti Salim, "Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of mfcc parameters and studying a speaker-independent approach," *Advanced Engineering Informatics*, vol. 28, no. 1, pp. 102 – 110, 2014.

[89] Michal Borsky, Marion Cocude, Daryush D. Mehta, Matías Zañartu, and Jon Guðnason, "Classification of voice modes using neck-surface accelerometer data," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5060–5064, March 2017.

[90] Yukitoshi Kashimoto, Manato Fujimoto, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto, "Floor vibration type estimation with piezo sensor toward indoor positioning system," *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 10 2016.

[91] Fulufhelo Nelwamondo and Tshilidzi Marwala, "Faults detection using gaussian mixture models, mel-frequency cepstral coefficients and kurtosis," *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC '06.*, vol. 1, pp. 290 – 295, 11 2006.

[92] Paul Mermelstein, *Distance Measures for Speech Recognition–Psychological and Instrumental*, New York : Academic Press, 1976.