Exploiting Heterogeneous Data for the Estimation of Particles Size Distribution in Industrial Plants

Damiano Rossetti and Stefano Squartini
Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
Emails: d.rossetti@pm.univpm.it
s.squartini@univpm.it

Stefano Collura
Research for Innovation
Loccioni Group
Angeli di Rosora, Ancona, Italy
Email: s.collura@loccioni.com

Yu Zhang
School of Engineering
University of Lincoln
Lincoln, United Kingdom
Email: yzhang@lincoln.ac.uk

Abstract—In industrial environments, it is often difficult and expensive to collect a good amount of data to adequately train expert systems for regression purposes. Therefore the usage of already available data, related to environments showing similar characteristics, could represent an effective approach to find a good balance between regression performance and the amount of data to gather for training. In this paper, the authors propose two alternative strategies for improving the regression performance by using heterogeneous data, i.e. data coming from diverse environments with respect to the one taken as reference for testing. These strategies are based on a standard machine learning algorithm, i.e. the Artificial Neural Network (ANN). The employed data came from measurements in industrial plants for energy production through the combustion of coal powder. The powder is transported in air within ducts and its size is detected by means of Acoustic Emissions (AE) produced by the impact of powder on the inner surface of the duct. The estimation of powder size distribution from AE signals is the task addressed in this work. Computer simulations show how the proposed strategies achieve a relevant improvement of regression performance with respect to the standard approach, using ANN directly on the dataset related to the reference plant.

I. Introduction

The particle size of the powder is an important parameter in many industrial processes, as it is related to the physical and chemical properties of materials. In most cases the powder particles have irregular shapes and speed, and travel within structures changing their characteristics over time. Generally, it is not interesting to describe the size of single particle but the size of an ensemble of particles. This means that cumulative parameters, such as the Particle Size Distribution (PSD), are usually employed on purpose. The PSD is a list of values, usually expressed in terms of percentage, denoting the relative amount, typically by mass, of particles present according to size. This size typically lies within predefined ranges, sorted in ascending or descending order.

In industrial plants, the evaluation of the PSD is usually performed by collecting a sample of powder and analysing it in laboratory. This method produces an accurate estimation of the PSD for a given time instant, but it is time consuming and impossible to use for continuous monitoring. In order to have a continuous monitoring of powder size, it is necessary to use a system that carries out the estimation in a non invasive way, for the whole time period of interest.

The main challenge is to find a 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. Acoustic Emission (AE) signals produced by the impact of powder on a metallic surface have been identified as meaningful quantities in order to obtain the PSD. Leach *et al.* [1], [2] 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, is is possible to determine their average diameter and the related size range. This method gave satisfactory results just for regularly shaped particles, i.e. spheres and cylinders. Unfortunately, it is impractical for most industrial applications where fluxes of irregular shaped of particles are involved. However, the same authors demonstrated that a particle impacting on a metallic surface generates an AE signal containing the information about its size. Many applications of this theoretical result have appeared in the literature [3], [4], [5], confirming the suitability of acoustic emissions for PSD measurement in engineering problems.

Moving from these premises, the authors [6] have recently used diverse machine learning techniques to train models for the estimation of PSD of powder by exploiting AE based information. Training and test data from a single duct was used and the results showed good performance for the employed algorithms, i.e. ANN, Support Vectors Regression (SVR) and Extreme Learning Machine (ELM). The training was performed by using suitable supervised learning algorithms and a single dataset related to a specific industrial plant.

Several tests conducted with this experimental setup 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 collecting 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 [7], [8], [9] 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) [10], Deep Belief Networks (DBN) [11], Recurrent Neural Networks (RNN) [12], [13], 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 paper proposes an alternative approach 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 recently investigated in the field of image processing with CNN [14] and in a biological context with Support Vectors Machine (SVM) [15]. 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 authors want 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.

This paper is organized as follows: Section II presents the algorithms and the datasets used for computer simulations, the methodology is presented in Section III, Section IV describes the experiments performed and presents results obtained in the diverse operating conditions addressed, finally, conclusions are drawn in Section V.

II. ALGORITHM AND DATASETS

A. Artificial Neural Networks

Generally, a Neural Network (NN) [16] is composed of neurons organized in layers, denoted as input, output, and hidden layers.

At the beginning of the training process, the weights of the neurons are initialized with initial weights and biases. In this work, for initial values used both random numbers and pre-trained values. During the training process, the initial weights are updated with the Backpropagation algorithm [17], using 900 epochs. Different activation functions can be chosen for the neurons: hyperbolic tangent, radial basis functions, and unipolar and bipolar sigmoid. In the experiments, the unipolar sigmoid has been used.

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

B. Datasets

The data used in this work are related to an industrial system employed in plants for the production of energy, namely POWdER [18]. This system continuously monitors the PSD of coal powder conveyed in ducts from grinding mills up to the boiler. Monitoring is done by using sensors installed on the outer surface of the ducts, detecting the acoustic emissions produced by the powder. The sensors are installed near a duct curve because this point has the highest probability of particle impact, generating the AE. The curve is the final part of a feeding duct that carries the coal powder from the mill to the burners. Figure 1 shows the structure of the plant: composed of the mill that grinds the coal, the feeding ducts that carry the coal powder, and the boiler with the burners where the coal combustion occurs.

The signals are acquired with a sample rate of 2 MS/s for 100k samples, thus each measured signal has duration of 50 ms. From those signals 64 features that characterize the energy of the acoustic emission signal are extracted. Each group of features corresponds to a vector of 3 targets. They represent the values of the PSD associated with the AE signal from which the features are extracted. For this system, the PSD is classified in three decreasing dimension sizes; corresponding to 50 MESH, 100 MESH and 200 MESH. The outcomes are the three numerical values, corresponding to the percentage of coal particles in the single powder sample which dimensions are lower than 300, 150 and 75 μm respectively. Therefore, each dataset item is composed by 64 features and 3 targets, and contains the information of a single AE acquisition.

Two distinct datasets related to two different industrial plants are involved in this work. The two plants are identified
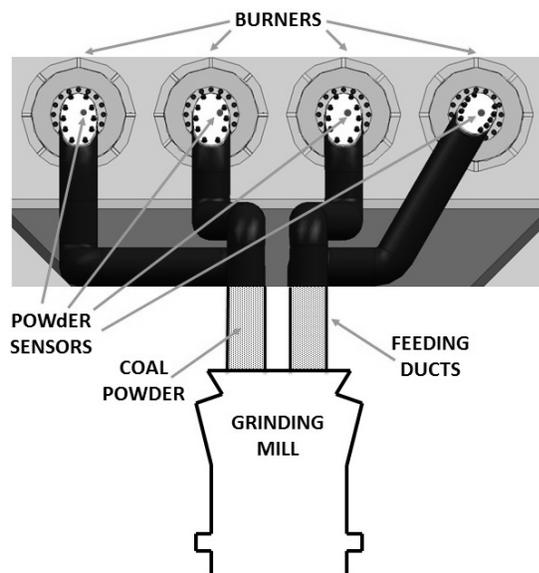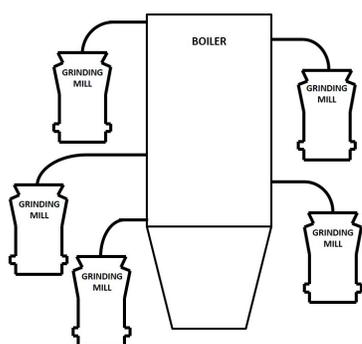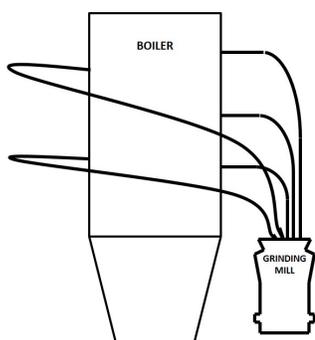
Fig. 1: Sketch of the power plant structure

(a) Plant A

(b) Plant B

Fig. 2: Different pipeline structure of two plants

as Plant A and Plant B. The plants differ in terms of pipeline layout, type of coal employed, and structure of mills, 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 2a), whereas for Plant B the five ducts are connected with the same mill (Figure 2b).

Moreover, 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 rates). The two datasets can be used to evaluate the generalization capability of the expert system for a certain duct. The number of data in each Primary dataset is different for each duct and varies from a minimum of 250 observations to a maximum of 330. The number of examples in the Secondary dataset is fixed for the two plants; 36 for plant A and 99 for plant B. It must be noticed that this distinction does not necessarily correspond to the standard Training/Test one. 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.

III. THE PROPOSED SUPERVISED TECHNIQUES FOR PSD ESTIMATION

In this Section, the two proposed supervised techniques exploiting heterogeneous data for PSD estimation are described in detail.

A. Pre-training

The 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.

The overall algorithm is depicted in Figure 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 Cross Validation (CV) with 6 not-overlapping folds. At each CV iteration, a different combination of Training and Test sets is selected from one
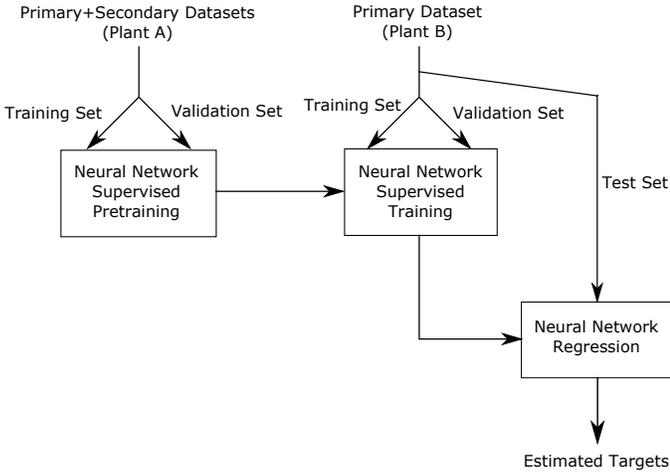
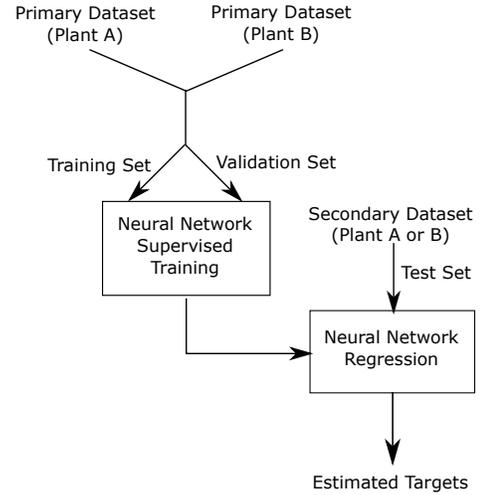Fig. 3: Supervised Pre-training Approach

Fig. 4: The Multiple Training Sets Approach

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.

B. 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. 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.

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.

IV. COMPUTER SIMULATION AND RESULTS

All simulations have been performed in Matlab 2013a® running on a Windows 7® OS. The performance evaluation of techniques discussed in Section III 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 I an example of normalization for the ANN Pre-training Case 1, for the single Duct 1, is shown.

TABLE I: 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 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 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.
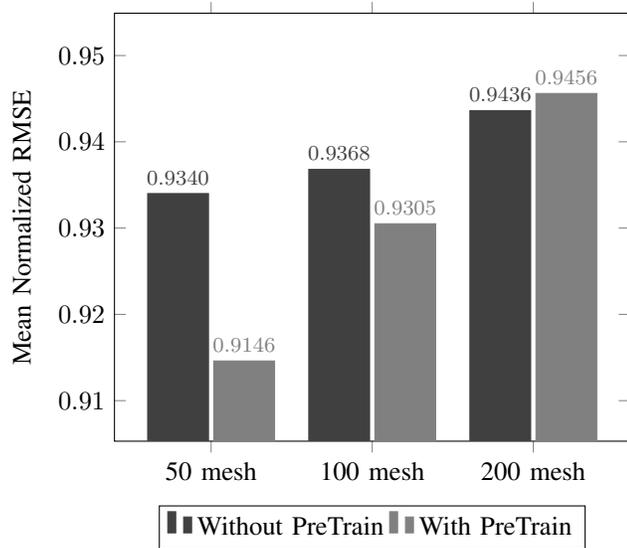
was tested. The models are trained by using all data contained in the Primary datasets of two plants. This means having three distinct models, one per mesh, which are finally tested against the Secondary dataset of the reference plant for each duct. In this case, the comparison is made with the standard approach involving ANN as an alternative. Figure 7 and Figure 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.

Fig. 5: Pretrain Case 1

Fig. 7: Multiple TS Case 1

Fig. 6: Pretrain Case 2

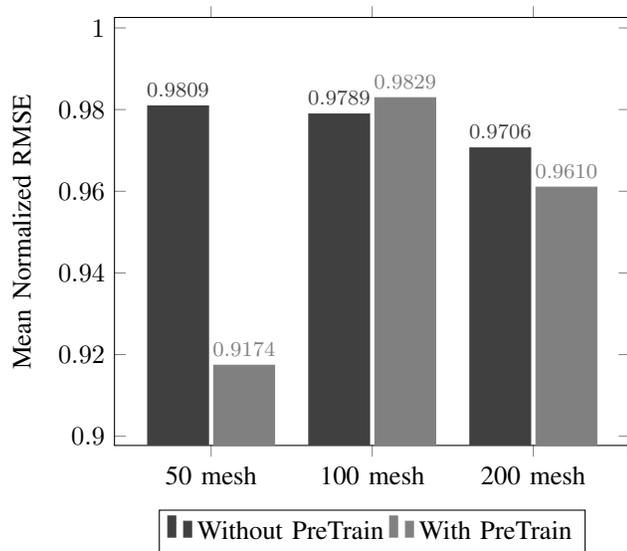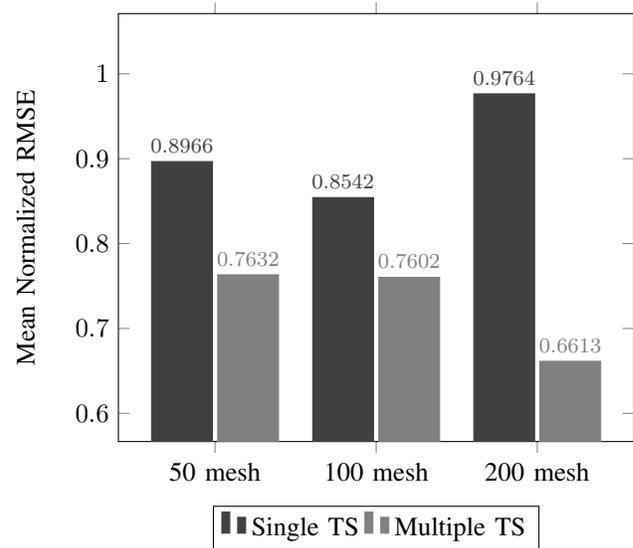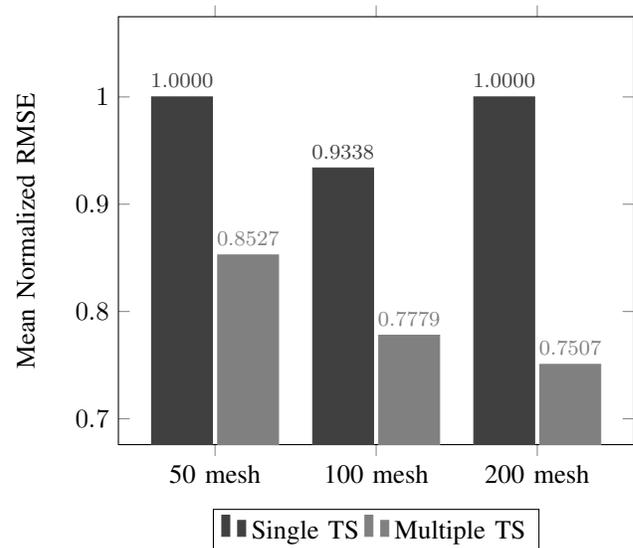
Fig. 8: Multiple TS Case 2

In the second set of experiments, the effectiveness of the second proposed supervised method described in Section III

In Table II 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

TABLE II: 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**

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 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.

V. CONCLUSION

In this work, two alternative strategies to use heterogeneous data for improving the performance of PSD estimation in industrial plants are proposed. In the first one, an ANN-based approach is used to implement a supervised pre-training procedure in order to initialize the free network parameters; using data from diverse environments with respect to the one taken as reference for fine-tuning and testing. In the second one, the use of multiple datasets from diverse plants for training ANN parameters has been investigated. with the aim to create general mesh-based models to test them on data of any plant.

The obtained results show that the pre-trained ANN has better performance then the ANN with random initialization of network parameters, for both addressed case studies. Like the previous one, also the second approach, by using a single model trained on the datasets of both plants collected together, provide enhanced performance if compared with those obtained with the train on a training set of a single duct. Concluding, the two proposed supervised strategies, based on the employment of already available heterogeneous data, deliver improved PSD estimation performance compared to standard algorithms. Comparing the two techniques, the Multiple Training Sets approach is more performing than the Pre-training approach, because it provides better results for all Test sets and a greater error reduction.

Future works will be targeted to the application of the proposed strategies to other types of industrial plant, in which the estimation of the powder size is used for monitoring purposes. Moreover, different machine learning techniques will be also employed in substitution of the ANN addressed in this paper, including Deep Learning architectures and related algorithms. Finally, unsupervised pre-training approaches and active learning strategies will be implemented and evaluated for the PSD estimation problem, and also combined with the supervised ones here proposed, in relationship with the characteristics and availability of suitable datasets.

REFERENCES

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

[2] ——, "Particle size distribution characterization from acoustic emissions," *Powder Technology*, vol. 19, no. 2, pp. 157 – 167, 1978.

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

[4] 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.

[5] Y. Hu, X. Huang, X. Qian, L. Gao, and Y. 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.

[6] D. Rossetti, S. Squartini, and S. Collura, "Machine learning techniques for the estimation of particle size distribution in industrial plants," 2015, (in press).

[7] 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.

[8] W. Cai, Y. Zhang, and J. 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.

[9] H. Chen and X. 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.

[10] X. Ren, K. Chen, X. Yang, Y. Zhou, J. He, and J. 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.

[11] C. Plahl, T. N. Sainath, B. Ramabhadran, and D. 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.

[12] S. R. Gangireddy, F. McInnes, and S. Renals, "Feed forward pre-training for recurrent neural network language models," 2014.

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

[14] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2524, 2013.

[15] D. P. Lewis, T. Jebara, and W. S. 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.

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

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

[18] S. 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.