Integration of Empirical Mode Decomposition and Machine Learning for operating condition classification: a proposal

L. Leoni*, A. Cantini *, F. De Carlo*, M. Tucci*

** Department of Industrial Engineering (DIEF), University of Florence, Viale Morgagni, 50135– Florence – Italy (leonardo.leoni@unifi.it, alessandra.cantini@unifi.it, filippo.decarlo@unifi.it, mario.tucci@unifi.it*

Abstract: During the last decades, advances related to the Internet of Things technologies have led to the widespread adoption of condition monitoring approaches for failure diagnosis. This fundamental vision has resulted in the introduction of advanced maintenance techniques such as Condition-Based Maintenance and Predictive Maintenance. Within this context, one of the main challenges arises from pre-processing the data acquired through sensors and, more specifically, the denoising procedures. Choosing a solid tool to separate the noise signal from the true one is usually considered a difficult task. In addition, data are usually collected from several different sensors, each of which monitors a specific process variable. Reducing the number of features analyzed could be helpful to improve the accuracy of the subsequent diagnosis and, at the same time, in making it less time-consuming. As a result, this paper aims at presenting a diagnosis approach based on Empirical Mode Decomposition (EMD) and Principal Component Analysis (PCA) for noise removal and feature reduction respectively. Indeed, EMD is very suitable for highly dynamic and non-stationary signals, while PCA is a good renowned approach for reducing the dimension of a multi-variate signal. Subsequently, a supervised machine learning approach is employed to classify the acquired signal. To demonstrate the applicability of the methodology, a compressor operating within a geothermal plant is considered as a case study, while the selected failure mode is the surge. The developed approach could be exploited by maintenance engineers and asset managers to perform diagnoses on relevant equipment.

Keywords: Machine Learning; Empirical Mode Decomposition; Principal Component Analysis; Operating condition classification

I. INTRODUCTION

In the last decades, the technological evolution has led to a vast diffusion of the Condition Monitoring (CM) as a support tool for making decisions on maintenance activities [1]. In turn, the diffusion of CM resulted in an ongoing effort to develop Condition Based Maintenance and Predictive Maintenance approaches [2-5]. Condition monitoring is defined as the process of measuring relevant Process Variables (PVs) to detect possible changes from a normal state to a faulty one [6]. A given PV is usually monitored through a specific sensor, whose acquired signal could contain high-noise components. Since there could be many monitored PVs, reducing the noise and the dimension of the data is of prominent importance to improve the accuracy of the subsequent classification techniques, while alleviating eventual time-consuming issues.

Given the relevance of the topic, there is a great deal of research on signal denoising across different fields [7-10]. De Faria et al. [11] successfully implemented a Discrete Wavelet Transform (DWT) to denoise a dataset concealing information related to the humidity, temperature, and light intensity of a lab. At first, the signal is decomposed into levels, and then the most energetic parts are extracted to rebuild the signal. Srivastava et al. [12] developed a DWT method based on the introduction of a user-automatically adjusted threshold level for denoising. Moreover, the authors introduced two thresholds for positive and negative wavelet coefficients.

Meanwhile, Empirical Mode Decomposition (EMD) and derivative approaches have become quite popular to deal with noise reduction [7, 13, 14]. Quite recently, BahooToroody et al. [15] applied EMD to reduce the noise concealed within a monitored pressure of a pressure regulator. Subsequently, the authors exploited a Hierarchical Bayesian Regression and a Generalized Linear Model to predict the number of pressure exceedances. A more recent work by Dwivedi et al. [16] proposed an integration between EMD and Stationary Wavelet Transform (SWT) for denoising the powerline interference of an electrocardiogram signal. Specifically, the authors decomposed the signal through EMD and subsequently performed another step of noise removal via SWT. The attractiveness of the EMD is related to its features, which make it very suitable for dealing with nonstationary data and complex time-dependent autocorrelation structure [17].

Within the context of condition monitoring, data reduction also plays a pivotal role. To perform this task,

many techniques could be adopted, such as Neighborhood Component Analysis (NCA) [18], Linear Discriminant Analysis (LDA)[19] and Random Forest (RF) [20]. Among all data reduction techniques, PCA is possibly the most common multivariate analysis technique in different disciplines [21]. Recent relevant examples related to PCA can be found in Parhizkar et al. [22] and Omuya et al. [23].

The final step of condition monitoring consists in applying an algorithm capable of identifying the state of the monitored equipment. This task is usually accomplished through Machine Learning (ML) techniques, such as Support Vector Machine (SVM), Decision Tree (DT), RF, K-Nearest Neighbors (KNN), and Neural Network (NN). Within this context, an important difference is given between supervised and unsupervised learning. The first case is obtained when the acquired data are provided with a class, while the second one is related to a scenario characterized by unlabelled observations.

Despite all the ongoing efforts, there is still space to propose a tool capable of classifying the operating condition of a device in case of high dimensional data with a strong non-stationary and non-linear nature. Indeed, as far as the authors know, the engineering related works on the application of EMD either involve a reduced set of PVs or they consider one single PV and EMD is applied after the extraction of relevant features such as the kurtosis. Consequently, the main objective of this paper is to present a framework capable of performing failure diagnosis in case several non-stationary and non-linear PVs are monitored for the same component. Compared to other similar studies, EMD and PCA are directly applied to the monitored signals, rather than a set of extracted features. Furthermore, different kinds of PVs are considered such as thermodynamic (e.g., temperature) and physical (e.g., position of a valve).

After this brief introduction, the remainder of this paper is organized as follows; Section 2 presents the core methods adopted for the study, while Section 3 describes the developed methodology. In Sections 4 and 5 the results are presented and discussed, respectively. Finally, in Section 6 the conclusions are drawn.

II. Material and Methods

In this work, the acquired signals are denoised through EMD and, subsequently, the data reduction is performed through PCA. Finally, the denoised and reduced signals are classified through a NN. The developed methodology is validated on real data acquired from 27 sensors placed on different parts of a compressor operating in a geothermal plant.

An overview of the main tools adopted to conduct this study is provided in the following subsections.

A. *Empirical Mode Decomposition*

Within the context of condition monitoring, a pivotal task is to reduce the noise of the signals monitored through sensors. The EMD is a filtering technique based on the Hilbert-Huang Transform [24] and it adopts a sifting process to decompose the signal into a series of Intrinsic Mode Functions (IMFs) and a residual term as depicted by Eq. 1.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r(t) \quad (1)$$

where n is the number of IMFs, while $r(t)$ and $c_i(t)$ identifies the residual and the *i*-th IMF, respectively. Since any identified IMF could either belong to the noisy signal or the true signal, Eq. 1 could be rewritten accordingly (see Eq. 2).

$$x(t) = \sum_{i=1}^{n} c_{i,TS}(t) + \sum_{i=1}^{m} c_{i,N}(t) + r(t) \quad (2)$$

In Eq. 2 $c_{i,TS}$ and $c_{i,N}$ denote a true signal IMF and a noisy IMF, respectively. To determine the noisy IMFs, a Statistical Significance Test (SST) is adopted. The SST requires the estimation of the mean period and the energy density for each IMF as illustrated in Eq. 3 and Eq. 4, respectively [17].

$$T_i = \frac{n}{P_i} \quad (3)$$

$$E_i = \frac{1}{n}\sum_{t=1}^{n} |c_i^2(t)| \quad (4)$$

where n is the number of observations, while P_i represents the number of peaks of the *i*-th IMF. Subsequently, the null hypothesis of Eq. 5 is tested.

$$ln\left(\frac{1}{3}E_1\right) + lnT_1 < lnE_i + lnT_i < ln(3E_1) + lnT_1 \quad (5)$$

The null hypothesis is that every IMF is a noisy IMF, thus, the IMFs for which the null hypothesis is accepted are considered as noisy, while the others are true signals.

B. *Principal Component Analysis*

PCA is a dimension reduction approach that transforms high-dimensional data into a lower-dimensional space [25]. Data reduction is performed through the identification of the so-named principal components that are characterized by the highest variation in data [26]. Let MS be a multivariate signal and PC the matrix of principal components, then PC could be expressed as a linear transformation mapping as illustrated by Eq. 6.

$$PC = W^T \cdot MS^T \quad (6)$$

where $W = [w_1, w_2, \dots, w_j \dots, w_n] \in R^{nxn}$, indeed each w_j is a column vector in R^n denoting the coefficient of the linear transformation. PC is a matrix where each component could be represented by Eq. 7.

$$PC_j = [PC_j(t_1), PC_j(t_2), \dots, PC_j(t_m)]^T \quad (7)$$

where t_1 is the first time instant, while t_m is the last one. The variance of a signal is proportional to its norm,

which is the energy of the signal. The norm of the j-th component is calculated through Eq. 8.

$$\|PC_j\|^2 = w_j^T CM \cdot w_j \quad (8)$$

where $CM = MS^T MS$ is the covariance matrix. The principal components are found iteratively by solving the eigenvalue problem shown in Eq. 9.

$$\begin{cases} CM \cdot w_1 = \lambda_1 w_1 \\ max_w = \{w^T CM \cdot w\} = \lambda_1 \end{cases} \quad (9)$$

where $\lambda_1 = max\{\lambda_j\}$ is the largest eigenvalue.

C. Neural Network

A NN is a renowned machine learning technique inspired by human neurons. It is characterized by an input layer,, an output layer and one or more hidden layers. Each layer is formed by nodes. The number of nodes within the input layer is equal to the number of inputs that are fed to the NN, while the nodes number of the output layer is set by users based on the desired output. The number of hidden layers and the nodes of each hidden layer are usually defined through an optimization process. The NN generates the outputs after combining the inputs through a series of activation functions and biases introduced through the hidden nodes. An example of NN is shown in Fig. 1.

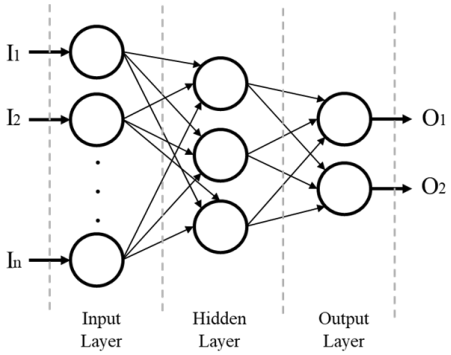

Fig. 1 schematic example of a NN

III. Developed Methodology

The sequence of the proposed methodology is shown in Fig. 2.

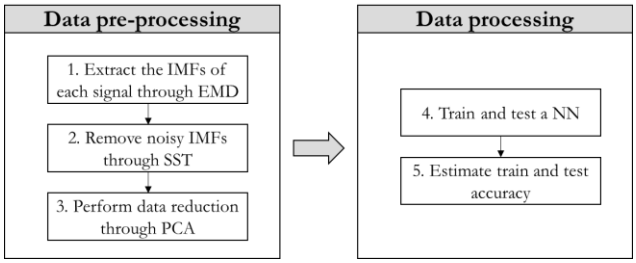

Fig. 2 sequence and steps of the developed methodology

The presented approach has two major stages named data pre-processing and processing, respectively. The first stage is divided into three steps, while the second stage has two steps. The first step consists of applying EMD for each acquired signal to determine the related IMFs (step 1). Subsequently, every IMF is screened through the null hypothesis of the SST (see Eq. 5), which allows us to detect and remove the noisy IMFs (step 2). To improve the accuracy of the subsequent classification approach and reduce the calculation time, a PCA is adopted on the denoised signals (step 3). The third step ends the data preprocessing stage. Next, the reduced dataset is split into a training and a test set, which is a process required to train and test the NN (step 4). Finally, the accuracy of the NN is estimated for both the training and test set (step 5).

IV. Results: Application of the Methodology

A. Case study

To describe an implementation of the framework, a compressor operating in a geothermal plant is chosen as case study (see Fig. 3).

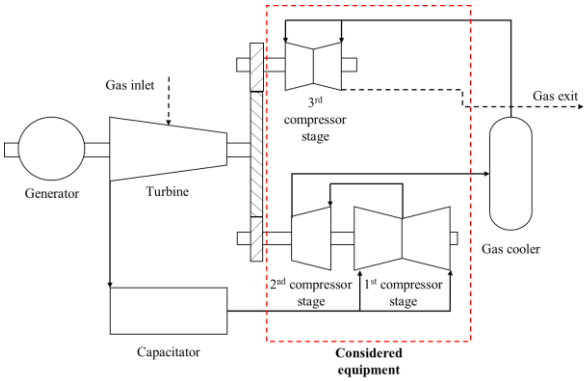

Fig. 3 schematic representation of the system concealing the considered compressor

It is a three-stage compressor whose task is to extract non-condensable gas. The first stage is characterized by two opposing impellers keyed on the same shaft. The second stage has a single impeller which is attached to the same shaft of the first stage. The last stage has two opposing impellers as the first one, but it is built on a separate shaft. The plant operates in a single flash condensation cycle, and it can develop up to 20 MW. The extracted gas is processed together with the steam coming from a turbine and a condenser.

The considered compressor could work under three distinct operating conditions identified as follows: I) regime or good working, II) anomaly working and III) surge condition. Specifically, the surge is an undesired operating condition which is considered as the failure mode throughout all this study. A total of 29 time series were extracted for 27 different PVs (see Table 1), each of which is monitored by a distinct sensor. The time series were classified into regime, anomaly, and surge through expert judgements. Specifically, the experts identified the operating condition through the analysis of the inlet pressure of the compressor.

B. Noise Removal and Data Reduction

Some of the acquired signals, especially the pressure ones, are very dynamic and present high non-stationary

and non-linear features. Thus, it is highly recommended to remove noisy components to obtain a more stable signal, which could be analysed more easily. It is worthwhile mentioning that the non-stationarity and nonlinearity are mainly associated with the surge events.

TABLE I
MONITORED PROCESS VARIABLES

#	Monitored process variable
1	Net active power
2	Wet bulb temperature
3	Flow rate - low pressure stage
4	Flow rate - high pressure stage
5	Suction gas pressure - low pressure stage
6	Suction gas pressure - medium pressure stage
7	Suction gas pressure - high pressure stage
8	Outlet high pressure stage gas pressure
9	Exhaust gas pressure
10	Interstage pressure gas extractor
11	Interstage pressure gas extractor
12	Interstage pressure gas extractor
13	Suction gas temperature - low pressure stage
14	Suction gas temperature - low pressure stage
15	Suction gas temperature - high pressure stage
16	First stage temperature
17	Second stage temperature
18	Third stage temperature
19	Outlet capacitator temperature
20	Outlet third stage temperature
21	Interstage gas temperature
22	Interstage gas temperature
23	Interstage gas temperature
24	Interstage gas temperature
25	Position of the first anti-surge valve
26	Position of the second anti-surge valve
27	Capacitator absolute pressure

For each signal of a given time series the EMD was applied, and the signal was divided into its corresponding IMFs. The sifting process to obtain the IMFs is stopped as soon as the residual does not present any peak (i.e., is monotonic) or when the maximum number of IMFs is reached. For this study, the maximum number of IMFs was set equal to 20. As an example, a signal from a surge event is considered. After decomposition, 11 IMFs were identified. Next, each IMF was tested through the null hypothesis of the SST. Specifically, the first and second IMFs were identified as noise, while the rest was depicted as true signal. The results are shown in Fig. 4 where the red dotted lines represent the lower and upper bound of null hypothesis acceptance, while the black solid line is the mean value. Finally, the violet dots are the recognized noisy IMFs, whereas the green dots denote the true-signal IMFs.

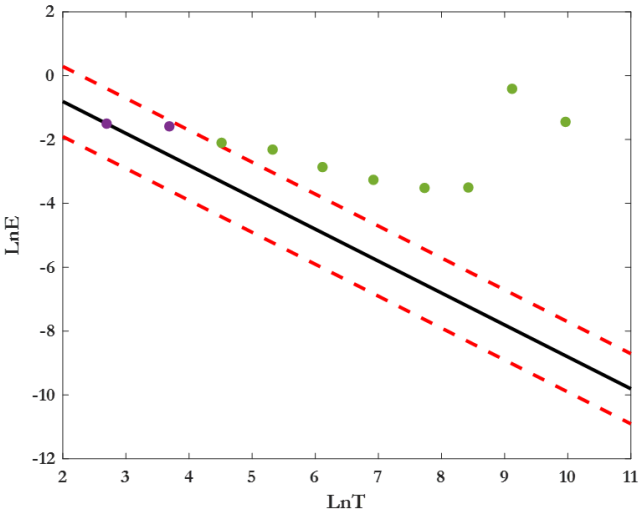

Fig. 4 example of the EMD application and subsequent SST

Next, the denoised signal was reconstructed through the sum of the residual and the detected true signal IMFs. Considering the previous example, the original signal and the denoised signal are represented by the black and red lines in Fig. 5 respectively.

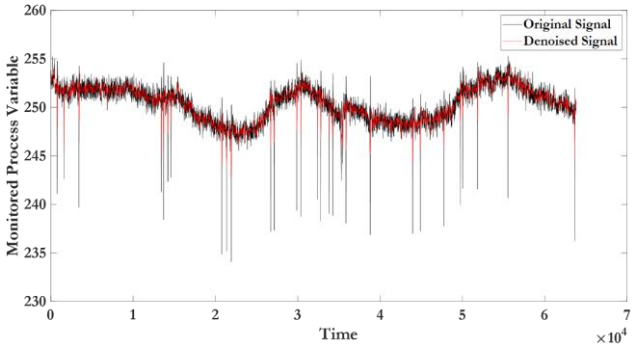

Fig. 5 example of original and denoised signal

The starting data are strongly unbalanced since there are many more observations related to the regime operating condition. Specifically, 4,932,153 observations were monitored for the regime conditions, while only 358,452 and 391,893 data points are associated with the anomaly and surge, respectively. Therefore, before applying the PCA, it is strongly recommended to balance the dataset. The lowest number of observations belongs to the anomaly operating condition, which is used as reference to obtain a homogeneous data set. Specifically, 75% of the observations were randomly

sampled from the anomaly conditions (that is, 268,839 observations). Subsequently, the same number of data points was also randomly extracted from the surge and regime conditions. Accordingly, 806,517 observations are considered for the PCA. Moreover, they were later exploited as a training set, while the discarded observations (ie 4,875,981) were used as test set.

After the balancing of the dataset, PCA was carried out. It emerged that more than 95% of the variability is explained by the first five principal components, as depicted in Fig. 6 and Table 2.

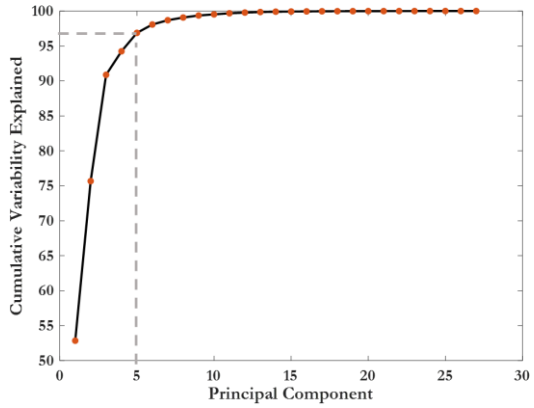

Fig. 6 cumulative explained variability for each principal component

TABLE II
EXPLAINED VARIABILITY AND CUMULATIVE EXPLAINED VARIABILITY FOR THE FIRST FIVE PRINCIPAL COMPONENTS

Principal Component	Explained Variability	Cumulative Explained Variability
1	52.85	52.85
2	22.82	75.67
3	15.21	90.88
4	3.36	94.24
5	2.64	96.88

It is worth mentioning that the first three principal components are characterized by much more explained variability. To better underline this variability, Fig. 7 shows the data represented in the space of the first three principal components.

As depicted by Fig. 7, the largest variability is associated with the first principal component which ranges between -250 and 150. On the other side, the variability associated with the second and third principal components are lower, since they are between -100 and 200, and -150 and 150 respectively. The second and third principal components have a similar variability, which is also denoted by the weight associated with them by the PCA (see Table 2).

To conclude this section, data reduction was performed by extracting the first five principal components of both the training and test set. Accordingly, from now on, the considered dataset has five inputs (i.e., the first five principal components) and one output (i.e., the operating condition).

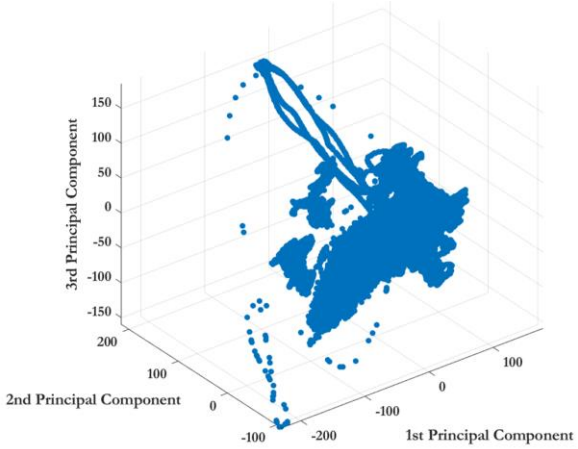

Fig. 7 training set observations represented in the space of the first three principal components

C. *Classification via NN*

At first, the NN was trained, thus the training set was fed into a NN with the characteristics reported in Table 3. Specifically, the adopted NN has one fully connected hidden layer with 10 nodes. The Rectified Linear Unit (ReLU), shown in Eq. 10, was chosen as the activation function. x denotes the input that is passed to a neuron.

$$f(x) = max\{0, x\} \tag{10}$$

TABLE III
CHARACTERISTICS OF THE ADOPTED NN

Variable	Value
Number of fully connected layers	1
Layer size	10
Activation function	ReLU
Iteration limit	1,000
Regularization strength parameter	0

The training was conducted through a 5-fold cross validation, and it resulted in the confusion matrix illustrated in Table 4, where the correct classifications are highlighted through a light blue cell.

TABLE IV
OBTAINED CONFUSION MATRIX FOR THE TRAINING SET

		Predicted		
		Regime	Anomaly	Surge
True	Regime	250,470	17,969	400
	Anomaly	8,581	250,152	10,106
	Surge	443	5,813	262,583

It emerged that 763,205 observations were correctly classified, while the remaining 43,412 data points were

associated with a wrong class. Thus, the overall accuracy of the training was estimated at 94.63%. Based on this value, it is possible to state that the NN was successfully trained. However, it is necessary to verify how the NN performs on data that were not used for the training process. To this end, the built NN is exploited to predict the class of the test dataset that was excluded during the training. The confusion matrix obtained from the test procedure is shown in Table 5.

TABLE **V**
OBTAINED CONFUSION MATRIX FOR THE TEST SET

		Predicted		
		Regime	Anomaly	Surge
True	Regime	4,452,882	206,578	3,854
	Anomaly	1,515	85,558	2,540
	Surge	154	3,637	119,263

Based on Table 5, it is possible to state that the developed NN has good generalizability. In fact, 4,657,703 observations were correctly classified, while a misclassification occurred for 218,278 data points. Accordingly, the test accuracy is equal to 95.52%.

V. DISCUSSION

From the previous section, it emerged that the NN has a high accuracy for both the training and the test set; thus, it could be adopted for diagnosis tasks. To have a better grasp on the performance of the NN, it is required to analyze the confusion matrices obtained more in-depth. For this purpose, let us define a relative accuracy for each operating condition as revealed by Eq. 11.

$$Acc_{rel,i} = \frac{\#Good\ classification,i}{\#Observations,i} \quad (11)$$

where $\#Observations,i$ and $\#Good\ classification,i$ identifies the total number of observations and the correctly classified observations of the *i*-th operating condition. For instance, considering the training dataset of the regime condition, the total number of observations is equal to 268,839, while the correctly classified data points are 250,470. The aforementioned values result in a relative accuracy equal to 93.17%. The relative accuracy estimated for the three operating conditions is listed in Table 6. It emerged that the highest relative accuracy is related to the surge operating condition for both the training and the test data. Furthermore, the test relative accuracies of the regime and anomaly state are higher compared to the training ones. This is an indicator of good generalization.

TABLE VI
TRAINING AND TEST ACCURACY FOR EACH OPERATING CONDITION

Operating condition	**Training accuracy (%)**	**Test accuracy (%)**
Regime	93.17	95.49
Anomaly	93.05	95.47
Surge	97.67	96.92

Another relevant discussion is related to misclassifications. Indeed, it is pivotal to understand which are the relationships between true and predicted class. Considering the regime state, it is mostly misclassified as anomaly. Specifically, 98% of the classification errors predicted a regime observation to be an anomaly, while only 2% resulted in a surge misclassification. With regards to the anomaly state, there is almost no difference in misclassification between anomaly and surge. Finally, the surge working condition is mostly misclassified as anomaly (93% and 96% of the time for training and test respectively). The aforementioned considerations ensure that it is unlikely to classify a regime condition as a surge and vice versa.

VI. CONCLUSION

This paper presents a framework for classifying the operating condition of a monitored equipment. The proposed approach can be used for classifying the operating condition of a device even in case it is characterized by highly dynamic and non-stationary PVs. The methodology is based on the application of EMD and PCA as core tools for noise removal and data reduction, respectively. Furthermore, as a classification tool, a NN is adopted. The proposed approach is tested on data coming from a compressor, whose operating conditions are identified as regime, anomaly, and surge. The results are quite promising, as the training and test accuracy are equal to 94.63% and 95.52%, respectively. Furthermore, the probability of incorrectly classifying the regime state as surge and vice versa is quite low. Further works can consider different case studies with a higher number of monitored signals. Finally, the adoption of distinct data reduction techniques (e.g., NCA, LDA) and ML approaches (e.g., SVM, DT, RF) can be tested coupled with EMD.

REFERENCES

[1] Tiboni, M.; C. Remino; R. Bussola;C. Amici A Review on Vibration-Based Condition Monitoring of Rotating Machinery. *Applied Sciences*. (2022). *12* p. 972.

[2] Leoni, L.; A. BahooToroody; M.M. Abaei; F. De Carlo; N. Paltrinieri;F. Sgarbossa On hierarchical bayesian based predictive maintenance of autonomous natural gas regulating operations. *Process Safety and Environmental Protection*. (2021). *147* p. 115-124.

[3] Navicelli, A.; F. De Carlo;M. Tucci Prognostics of industrial plant components in the absence of fault data: comparison between multivariate control charts and one-class SVM. *Proceedings of the XXV Summer School Francesco Turco*. (2021), p. 1-7.

[4] Bautista, L.; I.T. Castro;L. Landesa Condition-based maintenance for a system subject to multiple degradation processes with stochastic arrival intensity. *European Journal of Operational Research*. 2022.

[5] Oakley, J.L.; K.J. Wilson;P. Philipson A condition-based maintenance policy for continuously monitored multi-component systems with economic and stochastic dependence. *Reliability Engineering & System Safety*. (2022), p. 108321.

[6] Nithin, S.; K. Hemanth; V. Shamanth; R.S. Mahale; P. Sharath;A. Patil Importance of condition monitoring in

mechanical domain. *Materials Today: Proceedings*. (2021).

[7] Zhang, Y.; J. Ji;B. Ma Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode decomposition-convolutional deep belief network. *Measurement*. (2020). *156* p. 107619.

[8] Guo, X.; C. Shen;L. Chen Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Applied Sciences*. (2017). *7* p. 41.

[9] Antczak, K. Deep recurrent neural networks for ECG signal denoising. *arXiv preprint arXiv:1807.11551*. 2018.

[10] Bajaj, A.;S. Kumar A robust approach to denoise ECG signals based on fractional Stockwell transform. *Biomedical Signal Processing and Control*. (2020). *62* p. 102090.

[11] De Faria, M.L.L.; C.E. Cugnasca;J.R.A. Amazonas Insights into IoT data and an innovative DWT-based technique to denoise sensor signals. *IEEE Sensors Journal*. (2017). *18* p. 237-247.

[12] Srivastava, M.; C.L. Anderson;J.H. Freed A new wavelet denoising method for selecting decomposition levels and noise thresholds. *IEEE access*. (2016). *4* p. 3862-3877.

[13] Bin, G.; J. Gao; X. Li;B. Dhillon Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*. (2012). *27* p. 696-711.

[14] Žvokelj, M.; S. Zupan;I. 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*. (2010). *24* p. 1049-1067.

[15] BahooToroody, A.; F. De Carlo; N. Paltrinieri; M. Tucci;P. Van Gelder Bayesian regression based condition monitoring approach for effective reliability prediction of random processes in autonomous energy supply operation. *Reliability Engineering & System Safety*. (2020). *201* p. 106966.

[16] Dwivedi, A.K.; H. Ranjan; A. Menon;P. Periasamy Noise reduction in ECG signal using combined ensemble empirical mode decomposition method with stationary wavelet transform. *Circuits, Systems, and Signal Processing*. (2021). *40* p. 827-844.

[17] BahooToroody, A.; M.M. Abaei; F. BahooToroody; F. De Carlo; R. Abbassi;S. Khalaj A condition monitoring based signal filtering approach for dynamic time dependent safety assessment of natural gas distribution process. *Process Safety and Environmental Protection*. (2019). *123* p. 335-343.

[18] Raghu, S.;N. Sriraam Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. *Expert Systems with Applications*. (2018). *113* p. 18-32.

[19] Sharma, A.; K.K. Paliwal; S. Imoto;S. Miyano A feature selection method using improved regularized linear discriminant analysis. *Machine vision and applications*. (2014). *25* p. 775-786.

[20] Menze, B.H.; B.M. Kelm; R. Masuch; U. Himmelreich; P. Bachert; W. Petrich;F.A. Hamprecht A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*. (2009). *10* p. 1-16.

[21] Abdi, H.;L.J. Williams Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*. (2010). *2* p. 433-459.

[22] Parhizkar, T.; E. Rafieipour;A. Parhizkar Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. *Journal of Cleaner Production*. (2021). *279* p. 123866.

[23] Omuya, E.O.; G.O. Okeyo;M.W. Kimwele Feature selection for classification using principal component analysis and information gain. *Expert Systems with Applications*. (2021). *174* p. 114765.

[24] Huang, N.E.; Z. Shen; S.R. Long; M.C. Wu; H.H. Shih; Q. Zheng; N.-C. Yen; C.C. Tung;H.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. Series A: mathematical, physical and engineering sciences*. (1998). *454* p. 903-995.

[25] Kherif, F.;A. Latypova, *Principal component analysis*, in *Machine Learning*. (2020), Elsevier. p. 209-225.

[26] Ringnér, M. What is principal component analysis? *Nature biotechnology*. (2008). *26* p. 303-304.