

Anomaly Detection in Electromechanical Systems by means of Deep-Autoencoder

Francisco Arellano-Espitia
MCIA Research Center
Department of Electronic Engineering
Technical University of Catalonia
Terrassa, Spain
francisco.arellano@upc.edu

Miguel Delgado-Prieto
MCIA Research Center
Department of Electronic Engineering
Technical University of Catalonia
Terrassa, Spain
miguel.delgado@upc.edu

Víctor Martínez-Viol
MCIA Research Center
Department of Electronic Engineering
Technical University of Catalonia
Terrassa, Spain
victor.martinez.viol@upc.edu

Ángel Fernández-Sobrino
MCIA Research Center
Department of Electronic Engineering
Technical University of Catalonia
Terrassa, Spain
angel.fernandez.sobrino@upc.edu

Roque Alfredo Osornio-Rios
HSPdigital CA-Mecatronica
Engineering Faculty
Autonomous University of Queretaro
San Juan del Rio, Mexico
raosornio@hspdigital.org

Abstract—Anomaly detection in manufacturing processes is one of the main concerns in the new era of the Industry 4.0 framework. The detection of uncharacterized events represents a major challenge within the operation monitoring of electrical rotatory machinery. In this regard, although several machine learning techniques have been classically considered, the recent appearance of deep-learning approaches represents an opportunity in the field to increase the anomaly detection capabilities in front of complex electromechanical systems. However, each anomaly detection technique considers its own data interpretability and modelling strategy, which should be analyzed in front of the specificities of the data generated in an industrial environment and, specifically, by an electromechanical actuator. Thus, in this study, a comparison framework is considered including multiple fault scenarios in order to analyze the performance of four representative anomaly detection techniques, that is, one-class support vector machine, k -nearest neighbor, Gaussian mixture model and, finally, deep-autoencoder. The experimental results suggest that the use of the deep-autoencoder in the task of detecting anomalies of operation in electromechanical systems has a higher performance compared to the state of the art methods.

Keywords—*anomaly detection, electromechanical systems, deep-autoencoder, deep-learning.*

I. INTRODUCTION

In recent years, the industrial sector has evolved to allow more sophisticated production processes with the aim of increase operational efficiency. Such smart manufacturing environment is constituted by a high interaction between production systems (*i.e.* operational technology), and computational systems and communications (*i.e.* information technology), leading to the so called Industry 4.0 [1]. Among all industrial departments, the maintenance's area represents one of the most important investments to increase multiple key performance indicators (*e.g.* overall equipment effectiveness). In this regard, the implementation of predictive maintenance schemes is a desired strategy to act based on factual information rather than preventative recommendations. However, considering complexity and dynamic operating conditions of electrical rotatory based machinery and processes, current predictive maintenance schemes must accomplish with two main features. First, the capability of multiple patterns characterization resulting from such variability of the system's operation (*i.e.* torque, speed and fault conditions).

Second, the capability of anomalies detection referred as unknown conditions of the asset in order to allow an effective application avoiding false positives or even false negatives outcomes.

Indeed, in the last decade, with the fast-growth of methods based on Artificial Intelligence (AI), numerous data-based models have emerged. These algorithms have the ability to identify and learn complex relationships from the data in regard with the considered conditions. The Artificial Neural Networks (ANNs), are one of the most significant examples of these models. Furthermore, advances in hardware, especially in graphics processing units, have driven the development of depth ANN-based models. In this regard, Deep Learning (DL) techniques based on the ANN principles have enhanced the possibilities of complex data patterns management, characterization and posterior recognition in multiples disciplines [2-4].

Thus, the impact and advances over the use of DL in other fields of research are influencing the proposals to face current industrial challenges, specifically related with the industrial maintenance, leading to numerous studies for the diagnosis of faults in rotating systems [5-7]. However, most of these fault detection and identification schemes are carried out under the assumption of having enough representative data in regard with the healthy condition and the set of eventual faults conditions that will appear in the machine under monitoring. This assumption does not hold in a real industrial environment, where only the availability of data corresponding to the healthy condition can be assured. Thus, in order to allow the detection of operating conditions different from those considered as a reference or even different from those characterized as faults, the so called novelty detection or anomaly detection concept is being introduced.

In this regard, there are multiple studies around the detection of novelties applied to rotating electrical machinery, especially based on machine learning (ML) models. For example, in [8] a novelty detection and fault diagnosis scheme is carried out on an electromechanical system. In [9], different ML algorithms focused on the detection of anomalies are evaluated. These works, although they represent a significant advance in the area of anomaly detection, do not include the evolution towards DL technology.

In addition, many researches, such as [10], point out that using assembly strategies of different ML-based models represents one of the best options to increase performance when considering multiple operating and failure conditions.

In this sense, with the emergence of DL, the opportunity appears to explore new approaches where different independent models are not considered, but rather that the characterization of all the patterns is carried out by the same model, which offers a potential to improve the capacity of characterization and prevent diagnostic errors. Some recent works are pointing in this direction, however, studies that consider different technologies and novelty detection techniques are still needed to discuss their performance, limitations and thus allow making the best decision during the design of detection and diagnostic systems applied to the industrial sector.

Therefore, two main contributions are proposed in this study. First, a quantitative and qualitative comparison of the different anomaly detection approaches, including ML-based techniques such as k -nearest neighbors (k NN), one-class support vector machine, (OC-SVM) and Gaussian mixture model (GMM), but also a novel DL-based technique, deep-autoencoder (DAE). Second, the proposal of a novelty detection methodology including the use of DAE considering numerical features extracted from different domains to allow an incremental learning approach by re-training the models. Thus, the DAE-based methodology is validated to start from data corresponding only to the healthy condition of the electromechanical system under monitoring. It should be noted that the experimental validation is based on a public scientific database, and different faults, severities and operating conditions have been considered.

This paper is organized as follows: Section II describes the theoretical aspects of DAE as an anomalies detection model; Section III describes the proposed anomaly detection based PMS methodology; Section IV includes the experimental system description used for validation and presents the experimental results and discussion; finally, Section V includes conclusion remarks.

II. DEEP-AUTOENCODER AS ANOMALY DETECTION MODEL

Autoencoders (AEs) are a type of symmetrical neural networks designed to learn a low dimensional representation given the input data. They consist of two stages: encoder, which learns to map input data to a low dimensional representation, and decoder, which learns to map this low dimensional representation back to the reconstruction of input data. It is possible to build a Deep-Neural-Network (DNN) by stacking several AEs and performing a layer-by-layer training, through the method proposed by Hinton *et al* [11]. In this, each AE is trained to reduce the reconstruction error (RE) between the original input x and reconstructed output produced by the decoder \hat{x} by measuring of a cost function. In general, the RE of a AE is commonly measured by mean square error (MSE). Average MSE of N data samples can be written as follows:

$$MSE = \frac{1}{N} \|x_k - \hat{x}_k\|^2 \quad (1)$$

where x_k represents the original data input vector, and \hat{x}_k represents the reconstructed vector.

A DNN structure based on stacked AE, as shown in Fig.1, has the ability to learn complex relationships from the data through a mapping of non-linear functions.

In the encoding stage, the input data are continuously compressed by several hidden layers to a more compact representation.

$$E_1 = f(W_{e1}x_k + b_{e1}) \quad (2)$$

$$E_2 = f(W_{e2}E_1 + b_{e2}) \quad (3)$$

⋮

$$E_n = f(W_{en}E_{n-1} + b_{en}) \quad (4)$$

where E_i denotes the data vector mapped by each network layer and E_n denotes the deepest hidden representation vector. f represents the non-linear activation function of each layer. W_{ei} and b_{ei} are weight parameters and the corresponding biases of each encoding layer respectively.

In the decoding stage, the deepest hidden representation obtained from the encoding process is gradually decoded by several network layers. In the last layer of network, the model outputs a reconstruction vector of the initial input.

$$D_1 = f(W_{d1}E_n + b_{d1}) \quad (5)$$

$$D_2 = f(W_{d2}D_1 + b_{d2}) \quad (6)$$

⋮

$$\hat{x}_k = f(W_{dn}D_{n-1} + b_{dn}) \quad (7)$$

where D_j represents the corresponding the decoding vector for each of the encoding stages. W_{di} and b_{di} , are weight parameters and the corresponding biases of each decoding layer respectively. \hat{x}_k is the final output vector of the network, which is also the reconstruction vector of x_k .

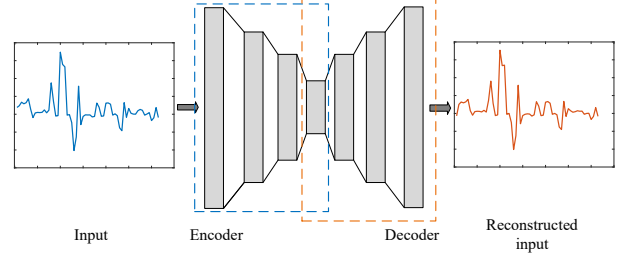


Fig. 1. Schematic structure of a deep-autoencoder.

The mapping function learned by DAE is specific to the training data distribution. Therefore, a DAE-based model typically fails at reconstructing data which is significantly different from data it has seen during training. This property of learning a specific mapping of a data distribution is particularly useful for the anomaly detection task. Applying a DAE-based model for anomaly detection follows the general principle of first training normal behavior and then generating an anomaly score for a new data sample. The cost function of the MSE used for AEs training allows it to be enabled as an anomaly measure.

In the field of fault detection in electromechanical systems, this ability to detect anomalies can be widely exploited. Especially when including new information to the initial model. Through network retraining, it can learn new patterns of new operating conditions and fault states, allowing an evolutionary diagnostic model.

III. METHODOLOGY

In this anomaly detection based PMS methodology proposed, a reconstruction-based model, as shown in Fig. 2. The first step is to process the data acquired from the system under monitoring to characterize the normal operating condition. Then, from the processed data, the estimation of numerical features in different domains is followed in order to better characterize and highlight possible anomalies in the electromechanical operation. Considering previous works related with predictive maintenance applied to electromechanical systems, the statistical time-domain features are: mean, maximum value, RMS, square root mean, standard deviation, variance, RMS shape factor, square root mean shape factor, crest factor, latitude factor, impulse factor, skewness, kurtosis, and normalized fifth and sixth. 15 statistical frequency-domain features are considered. In addition, 14 characteristic frequencies of faults are considered, characteristic frequencies of bearings and characteristic frequencies of gearbox [12-14]. In the case of features in the time-frequency domain, the decomposition of the signal is obtained through the Empirical mode decomposition (EMD) method. The first two IMF's of the decomposition are selected to obtain their Fourier amplitude coefficients and calculate 14 features of each one. Then, each of these features are concatenated to form a single feature vector that will serve to train the DAE with measurements of the normal operation of the machine. So, the input of the anomaly detection model is the information contained in the features, and the output is an anomaly score that determines how different is the new measurement analyzed compared to those that has been trained (the reference). Full details of how to build the SAE model can be found in [15]. Since the DAE tends to perform a poor reconstruction with data different from those used for its training, the reconstruction error of the input data is expected to be a score of anomaly detection. For the anomaly score, its set a simple threshold δ_{95} at the 95th percentile of $MSE(x_k, \hat{x}_k)$ for every x in the entire training set [16]. For any test data sample x_{Test} in online, its classify it as novel if

$$MSE(x_{Test}, \hat{x}_{Test}) > \delta_{95} \quad (8)$$

In contrast, the performance of three different classic methods for anomaly detection are analyzed. First, a distance-based method: k -nearest neighbors. Second, a domain-based method: one-class support vector machine. And third, a probabilistic-based method: Gaussian mixture model. All the details regarding these techniques can be found in [9], [17].

Distance-based methods are based on the assumption that normal data have close neighbours in the positive training set. The Euclidean distance is used to measure the proximity of a point to its close neighbors and establishes a k NN score. The probabilistic approaches estimate a probability density function of the positive data. The model will then classify points that lie outside of the high density region as a novelty. Domain-based methods construct a boundary using only the positive dataset. Domain-based approaches work by finding the optimal separation boundary between two classes, which is called a hyperplane. When getting a hyperplane, the output of this stage is binary (i.e. known or unknown).

In order to evaluate the anomaly detection performance of the DAE and the other techniques, it has been decided to use True Positive (TP) to represent the number of correctly classified positive samples, False Negative (FN) to represent the number of positive samples misclassified as negative, True Negative (TN) to represent correctly classified negative samples and False Positive (FP) to represent the number of negative samples misclassified as positive. The anomaly detection performance is evaluated considering the True Positive Rate (TPR), True Negative Rate (TNR) and the Balanced Accuracy:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{True Negative Rate (TNR)} = \frac{TN}{TN+FP} \quad (10)$$

$$\text{Balanced Accuracy} = \frac{(TPR + TNR)}{2} \quad (11)$$

The TPR and TNR are the scores correctly classified of the samples of the known classes and the samples of unknown classes, respectively.

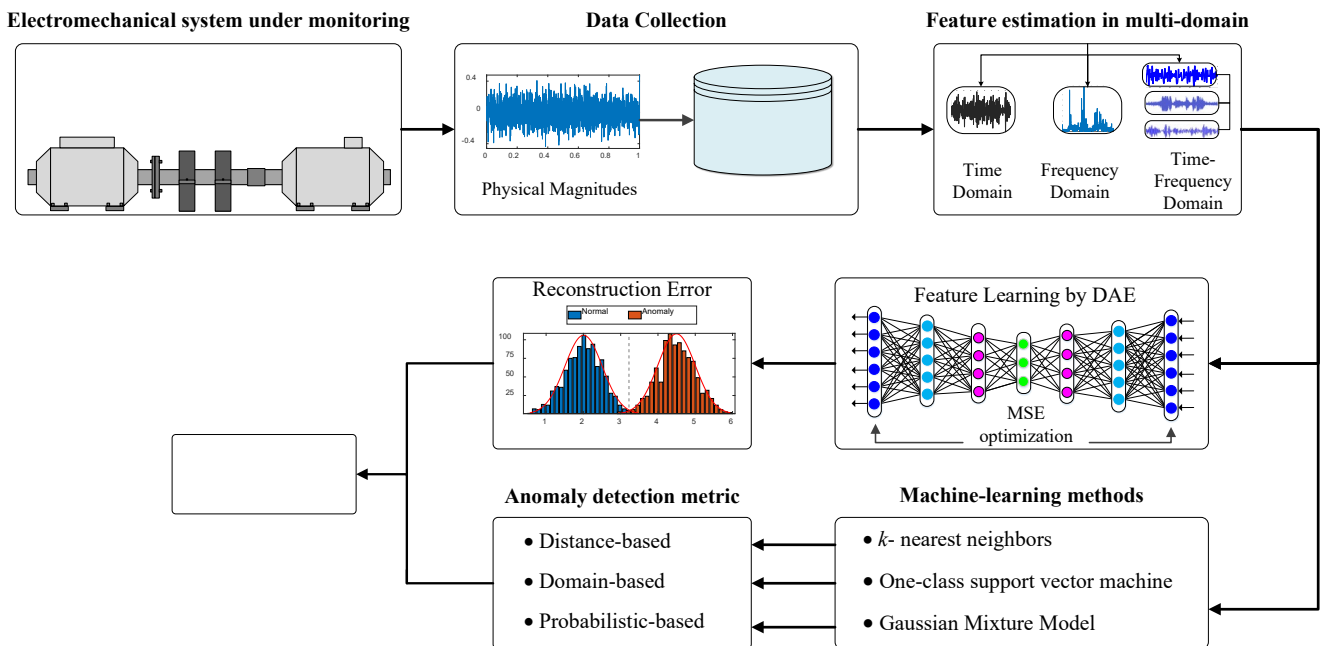


Fig. 2. Framework of the anomaly detection methodology.

IV. VALIDATION OF THE METHODOLOGY

The database provided by the Case Western Reserve University [17], is used to validate the proposed methodology and carry out the comparison with other methods. It was collected using accelerometers mounted at the drive end of motor which consists of normal data and faulty data. The faulty data set was generated by single point fault in the ball (FB), the inner race (FI) and the outer race (FO). For each one of the faults, various fault severities are considered, 0.007, 0.014 and 0.021 in, respectively. Also, the data were acquired at different operating conditions corresponding to various motor loads (0, 1, 2, and 3 hp). All data sets were collected with the sampling frequency of 12 kHz.

Seven scenarios for test are considered to evaluate the capability of the methodology to detect anomaly scenarios and the response of the models to the incorporation of new classes to the initially available information. The distribution of the classes for each scenario is showed in Table I. The four classes are grouped in three sets: training set, known set and unknown set. Each of the scenarios correspond to a progressing stage of the proposed approach, from an initial knowledge of only the normal condition (NC) with the four operating conditions, to a scenario where data of three classes is known. Posteriorly, one fault state is added to the training stage in each progressing stage. The developed DAE is formed by an encoder with an input layer of $x_k = 63$ neurons and $L = 3$ hidden layers of $D1 = 50$, $D2 = 10$, $D3 = 2$ neurons and a symmetric decoder. The hyperparameters of the DAE are set as following: L2-regularization = $5.0e-5$, sparsity regularization = $5.0e-5$ and sparsity proportion = 0.4.

TABLE I. EXPERIMENTAL SET FOR EACH TRAINING AND TESTING SCENARIO.

Label	Training Set	Testing Set	
		Known Set	Unknown Set
S1	NC	NC	FB, FI, FO
S2	NC, FB	NC, FB	FI, FO
S3	NC, FI	NC, FI	FB, FO
S4	NC, FO	NC, FO	FB, FI
S5	NC, FB, FI	NC, FB, FI	FO
S6	NC, FB, FO	NC, FB, FO	FI
S7	NC, FI, FO	NC, FI, FO	FB

The average MSE values obtained for the training stage of each of the scenarios are shown in Table II. Although the resulting range of MSE values differs for each scenario, the difference in these values does not imply a bad reconstruction of the model. In general, the MSE value obtained depends on the range of values of the input features. In this regard, Fig. 3 shows an example of reconstruction of the model for different health conditions. In Fig. 3(a), a real and reconstructed signal corresponding to the ball fault is shown, in this case a MSE value of 4.97 and a corresponding error percentage of 5.02% are presented. On the other hand, in Fig. 3(b) a real and reconstructed signal corresponding to the inner fault is presented with an MSE value of 172.26 and an error percentage of 7.32%. As can be seen, both figures are correctly reconstructed, however the high value of MSE in the case of inner fault compared to that of ball fault, is given by the value range of the input data. In this sense, a threshold is established at the 95th percentile of $MSE(x_k, \hat{x}_k)$. Fig. 4 illustrates a testing set distribution, in this the bars of the blue histogram represents the known data, while the orange bars are the data corresponding to unknown data or anomalies.

TABLE II. AVERAGE MSE CORRESPONDING TO EACH TRAINING SCENARIO.

Label	Training Set	Average MSE
S1	NC	0.35
S2	NC, FB	12.73
S3	NC, FI	202
S4	NC, FO	404
S5	NC, FB, FI	132
S6	NC, FB, FO	167
S7	NC, FI, FO	204

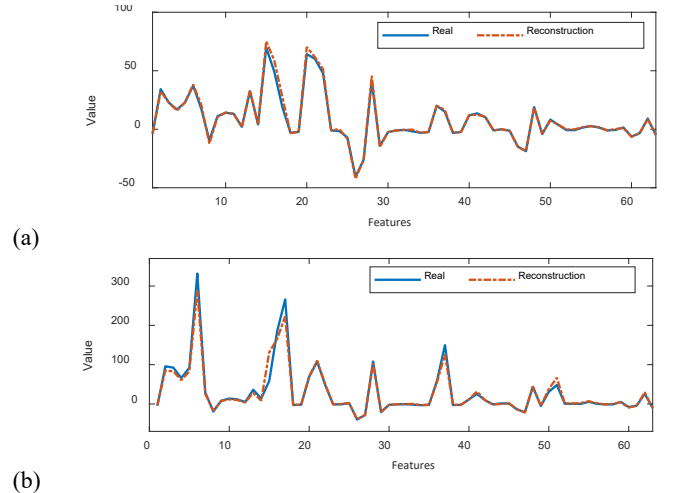


Fig. 3. Real feature signal and reconstructed signal by DAE. (a) Ball Fault; (b) Inner Fault.

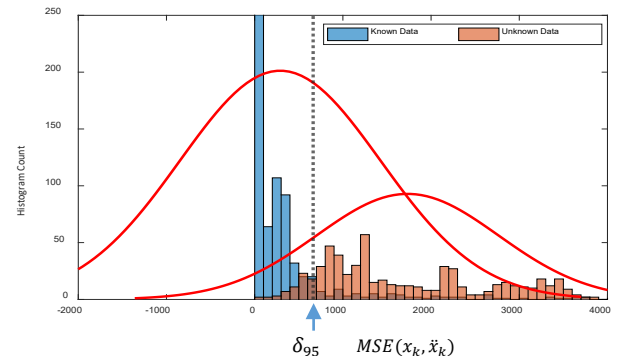


Fig. 4. Distribution of known data (blue) and unknown data (orange) and the corresponding threshold at the 95th percentile.

The orange distribution can be observed that can be correctly classified 94% of the samples as novel. While the known data is classified close to 95%.

In order to verify the effectiveness of the anomaly detection methodology, its compared with several successful classical anomaly methods: k -Nearest Neighbors (k NN), One-Class Support Vector Machine, (OC-SVM) and Gaussian Mixture Model (GMM). In the case of k NN, the algorithm was implemented by setting the number of neighbors nearest to $k=3$. For OC-SVM, the Radial Basis Function (RBF) kernel was used and is set to 0.08. While for GMM, the Gaussian number was established corresponding to the number of operating conditions considered. All the parameters were obtained through experimentation. The result of anomaly detection of the DAE and the methods used as a comparison is shown in Table III. In terms of TPR and TNR, the results show high variability.

On the one hand, it can be seen that k NN in general has a high performance in TPR but an undesirable performance in TNR. The opposite occurs with OC-SVM, in general it has a poor performance to generalize known cases, while the unknown data are detected correctly in some scenarios. On the other hand, GMM is the method with the greatest deviation. In this sense, DAE detection method presents a trade-off between the detection of known data and the deviations presented. Table IV shows the corresponding balanced accuracy. It can be noted that the DAE-based method is superior to the other methods. In this sense, DAE presents a higher performance in five of seven scenarios considered. k NN is superior in two, GMM in one and OC-SVM does not present a higher performance in any scenario.

TABLE III. COMPARISON OF PERFORMANCE OF ANOMALY DETECTION

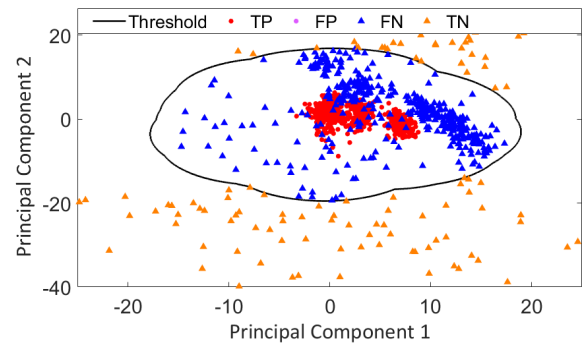
Label	KNN		OC-SVM		GMM		DAE	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
S1	100	75.1	80.5	98.2	57.5	98.7	99.8	100
S2	96.8	80.4	86.4	83.6	93.8	70.7	92.3	87.9
S3	96.9	98.5	86.2	99.6	84.5	98.5	95.2	97.2
S4	88.0	80.5	81.5	82.8	85.5	97.5	80.0	86.0
S5	96.7	66.1	89.8	66.6	94.7	67.0	95.2	67.8
S6	89.7	93.8	90.2	98.0	97.2	11.9	95.2	94.0
S7	90.2	59.3	88.8	37.4	83.5	55.8	97.3	38.9

TABLE IV. BALANCED ACCURACY OF ANOMALY DETECTION

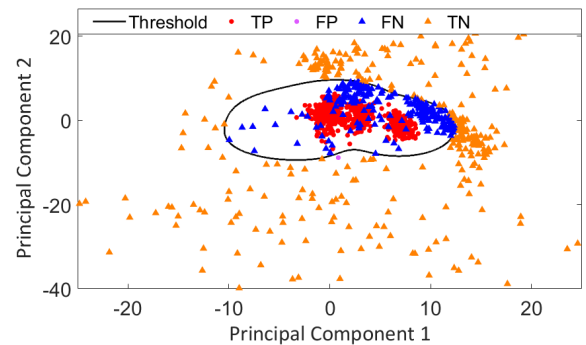
Label	Balanced Accuracy			
	KN N	OC-SVM	GMM	DAE
S1	87.5	89.4	78.1	99.9
S2	88.6	85.0	82.2	90.1
S3	97.7	92.9	91.5	96.2
S4	84.3	82.1	91.3	83.0
S5	80.9	78.2	80.8	81.5
S6	91.8	90.1	54.6	94.6
S7	74.8	63.1	69.7	69.6
Average accuracy	86.5	82.97	78.3	87.8

To better understand the functionality of each of the methods, the data is reduced to a 2-D space making is presented use of principal component analysis (PCA). PCA aims to find the linear projections that best capture the variability of the data and all the principal components are orthogonal to each other. After that, each of the methods used as an anomaly detection comparison is applied. Each of the 2-D projections and their corresponding anomaly detection method are shown in Fig. 5. Thereby, the projected space of features corresponds to a cumulative variance upper than 76%, which reflects a good concentration of the data to be qualitatively analysed. It should be noted that a good anomaly detection approach is expected to capture the high value of TP and TN , contrary a high value of FP and FN is undesirable.

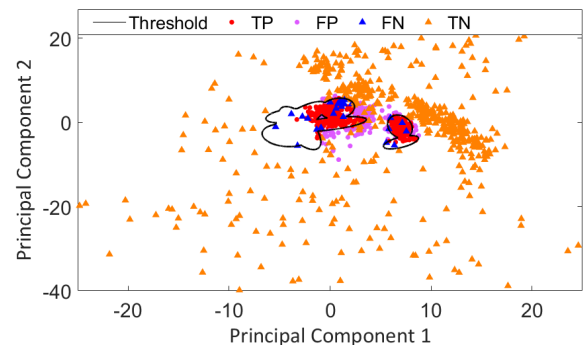
First, the detection of anomalies by the k NN model is presented in Fig. 5(a). In this, it can be seen that the boundary produced to determine the known data covers a large part of the feature space. Consequently, samples outside this boundary or threshold are considered anomalies. In this sense, the k NN method has a great capacity to capture the known data corresponding to the training stage, however, it produces a high number of FN , that is, incorrectly classified anomalous data. Second, the OC-SVM model is presented as an anomaly detector in Fig. 5(b).



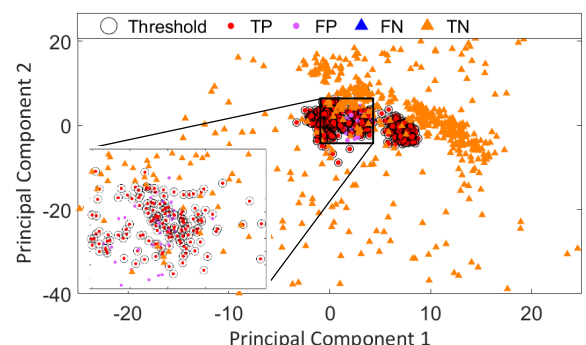
(a)



(b)



(c)



(d)

Fig. 5. Anomaly Evaluation of scenario S1 under different anomaly detection methods. (a) knn-method; (b) OC-SVM method; (c) Gmm-method; (d) DAE-method.

Third, the GMM-based anomaly detection model is presented in Fig. 5(c). As it can be seen, the shape and number of Gaussians greatly limits capture about the known data. Which interpret into a low performance with respect to the TPR values. Finally, the anomaly detection capacity of the DAE-based reconstruction model is presented.

This method is not based on delimiting a space of characteristics, but on establishing a membership of novelty based on the reconstruction error. Since the DAE learns a specific distribution from the training data, it will generally not be successful in reconstructing data that is significantly different from the data it has seen during training. To identify anomalies, the reconstruction error score is used as an anomaly score, marking samples with reconstruction errors above a threshold determined by the 95th percentile established. Fig. 5(d) shows the anomaly distribution by DAE obtained for the same projection obtained through PCA. The results of the projections shown correspond to scenario S1. However, the projections are representative of the application of each of the methods chosen to perform the anomaly detection comparison.

V. CONCLUSION

The proposed study presents the anomaly detection capabilities of a model based on Deep-Autoencoders applied to an electromechanical system compared to classical methods. The anomalies are interpreted as deviations from the healthy condition. In this work, the different bearing faults allow to validate the anomaly detection approach, where the healthy condition is trained first and fault cases are progressively added to the available knowledge. There are three main important aspects in this study. First, different study scenarios are presented to evaluate SAE's anomaly detection capabilities. Second, comparisons are carried out with different state-of-the-art techniques for anomaly detection. Third, each of them is evaluated in terms of detection of known cases and novel cases, which mean samples used during training and never-before-seen samples or unknown samples, respectively.

It should be noted that the anomaly detection study faced in electromechanical systems involves various requirements and considerations. Such as different operating conditions, complexity within electromechanical systems, different fault states, and data representation. Therefore, the application of adaptive methods and with higher calculation capacity is required to face the different problems present in the diagnosis and detection schemes of industrial systems. In this sense, the reconstruction-based approach, that means, the DAE-based approach, its superior in terms of performance in five out of seven scenarios. In addition, it presented the best balance between TPR and TNR of all the methods considered for the comparison. Although the results obtained by deep-learning are promising, it must be taken into account that the field of diagnosis and detection of anomalies is rapidly evolving, so are far from solved problems.

In future works, more advanced strategies will be studied, especially with the use of deep-learning tools, to improve anomaly detection schemes and make them more and more reliable. Taking into account the great challenges that emerge in smart manufacturing and in the industry 4.0 framework.

ACKNOWLEDGMENT

This work was partially supported by CONACYT, under doctoral scholarship number 710065, and partly co-financed by the European Regional Development Fund of the European Union in the framework of the ERDF Operational Program of Catalonia 2014-2020.

REFERENCES

- [1] W. Yu, T. Dillon, F. Mostafa, W. Rahayu and Y. Liu, "A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance," in *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 183-192, 2020.
- [2] A. Akay and H. Hess, "Deep Learning: Current and Emerging Applications in Medicine and Technology," in *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 906-920, 2019.
- [3] R. A. Khalil, N. Saeed, M. Masood, Y. M. Fard, M. -S. Alouini and T. Y. Al-Naffouri, "Deep Learning in the Industrial Internet of Things: Potentials, Challenges, and Emerging Applications," in *IEEE Internet Things J.*, 2021.
- [4] W. Li, H. Li, Q. Wu, X. Chen and K. N. Ngan, "Simultaneously Detecting and Counting Dense Vehicles From Drone Images," in *EEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9651-9662, 2019.
- [5] M. He and D. He, "Deep Learning Based Approach for Bearing Fault Diagnosis," in *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057-3065, 2017.
- [6] S. Shao, R. Yan, Y. Lu, P. Wang and R. X. Gao, "DCNN-Based Multi-Signal Induction Motor Fault Diagnosis," in *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658-2669, 2020.
- [7] X. Zhao, M. Jia, P. Ding, C. Yang, D. She and Z. Liu, "Intelligent Fault Diagnosis of Multichannel Motor-Rotor System Based on Multimanifold Deep Extreme Learning Machine," in *IEEE/ASME Trans. Mechatronics.*, vol. 25, no. 5, pp. 2177-2187, 2020.
- [8] M. Delgado-Prieto, J. A. Cariño, J. J. Saucedo-Dorantes, R. A. Osornio-Rios, L. Romeral and R. J. Romero Troncoso, "Novelty Detection based Condition Monitoring Scheme Applied to Electromechanical Systems," 2018 IEEE 23rd Int. Conf. ETFA, 2018, pp. 1213-1216.
- [9] M. Ramírez Chavez, L. Ruíz Soto, F. Arellano Espitia, J. J. Saucedo, M. Delgado Prieto and L. Romeral, "Evaluation of Multiclass Novelty Detection Algorithms for Electric Machine Monitoring," 2019 IEEE 12th Int. SDEMPED, 2019.
- [10] J. J. Saucedo-Dorantes, M. Delgado-Prieto, R. A. Osornio-Rios and R. d. J. Romero-Troncoso, "Industrial Data-Driven Monitoring Based on Incremental Learning Applied to the Detection of Novel Faults," in *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5985-5995, 2020.
- [11] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, Vol. 18, no. 7, pp. 1527-1554, Jul. 2006.
- [12] F. Arellano-Espitia, J. J. Saucedo-Dorantes, R. A. Osornio-Rios, M. Delgado-Prieto, J. A. Cariño-Corrales and R. J. Romero-Troncoso, "Statistical data fusion as diagnosis scheme applied to a kinematic chain," 2018 IEEE 23rd Int. Conf. ICIT, 2018, pp. 2111-2118.
- [13] Li YF, Liang XH, Zuo MJ. "Diagonal slice spectrum assisted optimal scale morphological filter for rolling element bearing fault diagnosis," *Mech. Syst. Signal Process* 2017; Vol. 85, February 2017, pp. 146-161.
- [14] S. H. Kia, H. Henaoui and G. Capolino, "A comparative study of acoustic, vibration and stator current signatures for gear tooth fault diagnosis," 2012 XXth Int. Conf. on Elec. Mach., Marseille, 2012, pp. 1514-1519.
- [15] F. Arellano-Espitia, M. Delgado-Prieto, V. Martinez-Viol, J.J. Saucedo-Dorantes, R.A. Osornio-Rios, "Deep-Learning-Based Methodology for Fault Diagnosis in Electromechanical Systems," *Sensors* 2020, 20, 3949.
- [16] A. Amini, W. Schwarting, G. Rosman, B. Araki, S. Karaman and D. Rus, "Variational Autoencoder for End-to-End Control of Autonomous Driving with Novelty Detection and Training De-biasing," 2018 IEEE/RISJ IEEE 23rd Int. Conf. IROS, 2018, pp. 568-575.
- [17] D. Oosterlinck, D. F. Benoit and P. Baecke, "From one-class to two-class classification by incorporating expert knowledge: Novelty detection in human behaviour," *Eur. J. Oper. Res.*, vol. 282, Issue 3, Pages 1011-1024, 2020.
- [18] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64-65, pp. 100-131, Dec. 2015.