

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Novel Methods Based on Deep Learning Applied to Condition Monitoring in Smart Manufacturing Processes

Francisco Arellano Espitia and Lucia Ruiz Soto

Abstract

The Industry 4.0 is the recent trend of automation and the rotating machinery takes a role of great relevance when it comes to meet the demands and challenges of smart manufacturing. Condition-based monitoring (CBM) schemes are the most prominent tool to cover the task of predictive diagnosis. With the current demand of the industry and the increasing complexity of the systems, it is vital to incorporate CBM methodologies that are capable of facing the variability and complexity of manufacturing processes. In recent years, various deep learning techniques have been applied successfully in different areas of research, such as image recognition, robotics, and the detection of abnormalities in clinical studies; some of these techniques have been approaching to the diagnosis of the condition in rotating machinery, promising great results in the Industry 4.0 era. In this chapter, some of the deep learning techniques that promise to make important advances in the field of intelligent fault diagnosis in industrial electromechanical systems will be addressed.

Keywords: Industry 4.0, condition-based monitoring, deep learning

1. Introduction

In recent years within the industrial sector, there is a trend toward the evolution to the Industry 4.0 paradigm, which implies the integration of multiple technologies for the start-up of intelligent factories capable of adapting to the needs and production processes. In these intelligent manufacturing systems, the diagnosis of the condition of the machine is of great importance to prevent failures and avoid monetary losses caused by work stoppages in production. The condition-based monitoring (CBM) schemes are the most accepted to carry out this task. However, one of the main challenges within CBM schemes is the construction of models capable of adapting to highly complex manufacturing systems, which are also subject to high variability of their operating conditions and under the presence of high noise.

Meanwhile, deep learning (DL), or also known as deep neural networks (DNN), has become an analytical tool that has attracted more and more attention from researchers in different areas of research in recent years. The main skill DNN has

the ability to learn and extract useful patterns from the data. Therefore, there is currently a tendency to make use of this ability of DNNs to extract significant features from complex manufacturing systems, in order to find the characteristic patterns of faults and thus be able to diagnose anomalies in a timely manner.

As a branch of machine learning, the DL appears from the learning capacity of the artificial neural networks (ANNs); however, the learning capacity of the ANN is limited and presents problems when making the adjustment of weights through error correction (backpropagation). Therefore, different DL architectures have been developed based on stacking multiple layers of ANN, such as auto-encoders, convolutional neural networks, or restricted Boltzmann machine. These architectures seek to obtain hierarchical representations and intrinsic relationships of the data.

The main reason for the application of techniques based on DL in the study of the condition of electromechanical systems is due to the limitation presented by the basic analysis schemes. A traditional diagnostic scheme consists in the extraction and selection of feature engineering from the acquisition data, followed by the application of a dimensionality reduction process and the training of a prediction model based on machine learning which includes support vector machines (SVM), simple neural networks (NN), or regression algorithms.

The main limitation of these traditional diagnostic models is the low capacity to adapt to complex electromechanical systems, and therefore, they have difficulties to adequately characterize all the variability of operation and the different condition states including faults. Unlike traditional schemes based on machine learning, DL schemes are not limited to characterizing systems with only a set of pre-established features, but, through the construction of structures based on neural networks, they are able to extract hierarchical representations of the data. These representations or extracted features have a greater representative capacity because the schemes for their extraction are through non-linear algorithms; with this, a structure based on deep learning is able to learn the adjacent non-linearities of faults and multiple operating conditions of modern manufacturing processes that integrate rotary systems among their components.

The purpose of this literature is to review the emerging research papers of DL focused on condition monitoring. After the brief summary of the DL tools, the main applications of deep learning are about the monitoring of the condition of electromechanical systems.

2. Deep neural networks

To solve binary classification problems, one of the algorithms inspired by the learning process of biological neural networks was called perceptron [1]. The perceptron consists of an input unit directly connected to an output node; the pattern learning process is performed through an operation called activation function. To solve more complex problems, multi-layer perceptron called artificial neural networks (ANN) are used. The training process of these ANNs is performed by executing multiple iterations each time a new measurement is presented, and the weights and biases are adjusted by following a training and error correction algorithm called backpropagation [2].

By adding more hidden layers to the network, it is possible to create a deep structure capable of extracting more complex patterns and finding more hidden data relationships. These deep architectures with multiple hidden layers are known as deep neural networks (DNN). However, a trivial problem, which arises in the training of DNN as more hidden layers are added to the network, is that the correction of the error does not propagate toward the first layer of the network, generating a problem of vanishing of the gradient, hindering the learning process.

2.1 Convolutional neural network

One of the main DNN-based architectures for feature extraction is convolutional neural networks (CNNs). A convolutional neural network is a kind of artificial neural network designed specifically for identifying patterns in the data [3]. This type of architecture uses a multi-channel input, such as an image or multiple combined signals. The central idea behind CNN is the mathematical operation of convolution, which is a specialized type of linear operation. Each CNN layer performs a transform domain, where the parameters to perform the transformation are organized as a set of filters that connect to the input and thus produce an output layer. The output of a CNN layer is a 3D tensor, which consists of a stack of arrays called feature maps; these features can be used as an input to a next layer of the CNN scheme. A simple CNN architecture is shown in **Figure 1**.

CNN has three main states: convolution, pooling, and fully connected. Convolution puts the input signal through a set of convolutional operators or filters, each of which activates certain features from the data. Pooling minimizes the output through performing a decrease in non-linear sampling, reducing the number of parameters that the network needs to learn. The last layer is a fully connected layer that produces a vector of N dimensions, where N is the number of classes that the network can predict. This vector contains the probabilities for each class of the data considered. Finally, the output of a convolutional network is connected to a classification stage, in order to obtain a diagnosis.

2.2 Auto encoders

The auto-encoder is a type of symmetrical neural network that tries to learn the features in a semi-supervised manner by minimizing reconstruction error. A typical structure of an auto-encoder is shown in **Figure 2**. This has three layers: input layer, hidden layer, and output layer. The learning procedure of AE consists in two stages: encoder and decoder stages. Input layer and the hidden layer are regarded as an encoder, and the hidden layer and the output layer are regarded as a decoder.

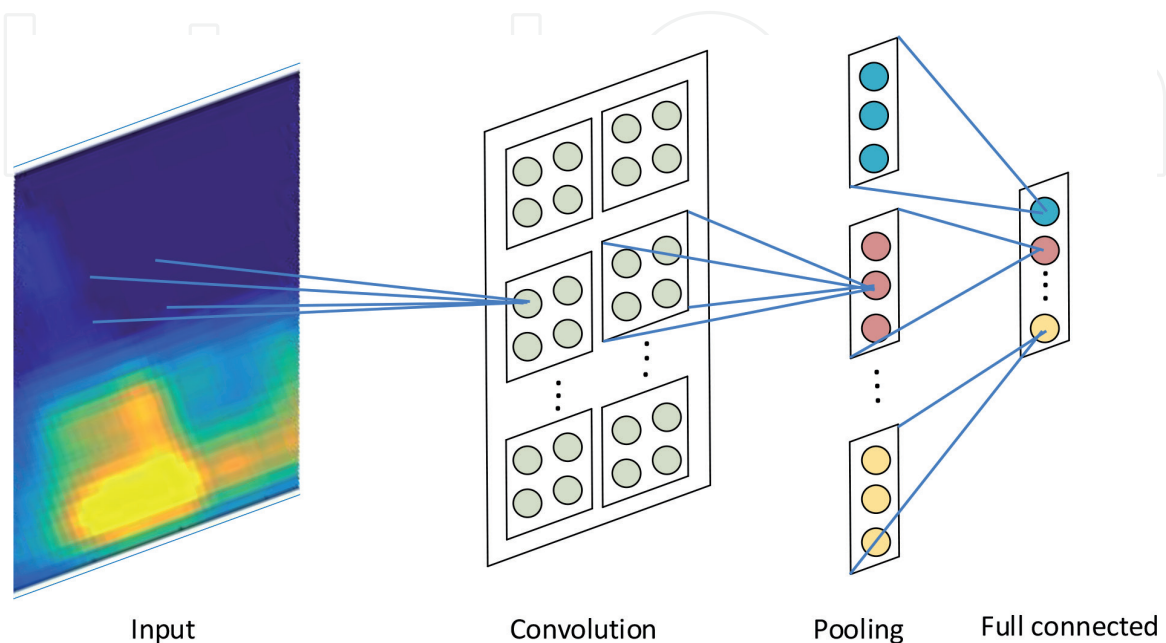


Figure 1.
Architecture of a CNN comprising commonly used layers as convolution and pooling.

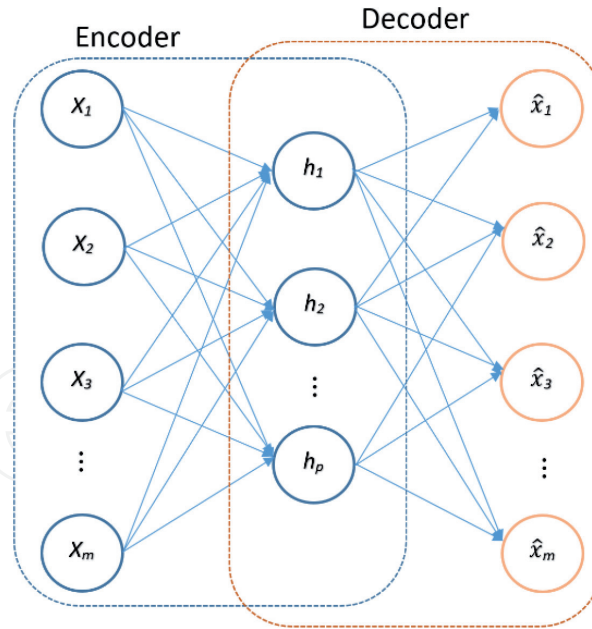


Figure 2.
Architecture of an auto-encoder.

The encoder process is described by $f_{\{W^1, b^1\}}(x) = s_f(W^{(1)}x + b^{(1)})$, and the decoder process is $g_{\{W^2, b^2\}}(x) = s_g(W^{(2)}x + b^{(2)})$, where s_f and s_g are the activation functions of the encoder and decoder, respectively, W is the weight vector between these different layers, and b is the bias. W and b are the trainable parameters of encoder and decoder stages. Furthermore, the sigmoid function is chosen as an activation function to the network. For any sample x^m from input data set $\{x^m\}_{m=1}^M$, where M is the number of samples, the encoder stage computing an encode vector $a^m = f_{\{W^1, b^1\}}(x^m)$. Also, a^m can be regarded as a feature representation that the encoder process learning from the input data.

To improve the performance of the traditional auto-encoder, a sparse restriction term is introduced, generating a variant known as sparse auto-encoder (SAE) [4–6]. The sparse restriction term works on the hidden layer to control the number of “active” neurons. In the network, if the output of a neuron is close to 1, the neuron is considered to be “active,” otherwise it is “inactive.” With the sparse restriction, SAE can obtain proper parameter sets by minimizing the cost function

$$J_{sparse}(W, b) = \frac{1}{M} \sum_{m=1}^M L(x^m, \hat{x}^m) + \lambda \cdot \|W\|_2^2 + \beta \cdot \sum_{j=1}^n KL(\rho \|\hat{p}_j) \quad (1)$$

where $L(x^m, \hat{x}^m) = \|x - \hat{x}^m\|$ is the average sum of squares error term, λ is the weight decay parameter, β is the sparsity penalty parameter, and ρ is the sparsity parameter.

2.3 Restricted Boltzmann machine

A restricted Boltzmann machine (RBM) is a type of neural network formed by two layers that consist of two groups of units including visible units v and hidden units h with the constraint that there only exists a symmetric connection between visible units and hidden units, and there are no connections between nodes with a same group, as shown in **Figure 3**. These networks are modeled by using stochastic units, habitually Gaussian.

The learning procedure includes several stages known as Gibbs sampling, which gradually modifies the weights to minimize the reconstruction error. These type of NNs is commonly used to model probabilistic relationships between variables.

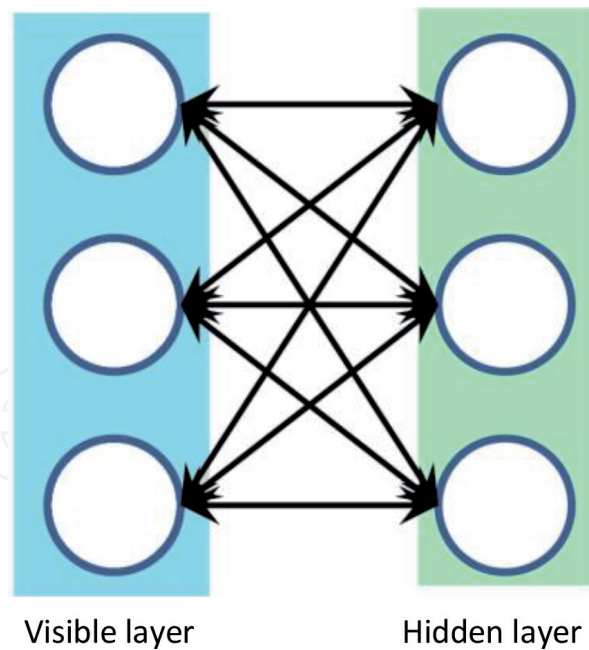


Figure 3.
Schematic illustration of restricted Boltzmann machine.

The most used algorithm to perform the training of an RBM is the contrastive divergence (CD) method [7]. Contrastive divergence is a type of unsupervised learning algorithm; it consists of two stages that can be called positive and negative stages. During the positive stage, the network parameters are modified to replicate the training set, while during the negative stage, it attempts to recreate the data based on the current network configuration.

Restricted Boltzmann machines can be used in deep learning networks in order to extract characteristic patterns from the data. For example, deep belief networks can be designed by stacking various RBM and performing a fine-tuning the resulting deep network with gradient descent and backpropagation. Like the CNN network, a classification stage is connected to the deep network output.

3. Applications of deep learning in condition-based monitoring

For several years, the best tools for monitoring electromechanical systems were data-driven schemes [8]. However, with the increase in the complexity of the systems, the increase in case studies, and the need to incorporate new operating conditions, traditional machine-based schemes are insufficient to characterize such complexity because their discriminative capacity is decreasing. Consequently, the study of the condition of the machine has been moving toward the incorporation of techniques based on deep learning.

Applications such as feature extraction, dimensionality reduction, novelty detection, and transfer learning are some of the tasks that can be carried out through the three deep learning techniques mentioned above: CNN, AE, and RBM.

3.1 Feature extraction

The schemes that are able to extract features effectively and have the ability to handle large data dimensions are needed. Automation of feature engineering has become an emerging topic of research in academia; in recent years, it has emerged deep learning (DL) techniques capable of dealing with the complexity presented

in many cases of study. DL is a branch of machine learning based on multi-layer neural networks or deep neural networks (DNNs), where the objective of each layer or level is to learn to transform your input data into a non-linear and more abstract representation. The transformation learned through DNN can contain information that preserves the discriminative features of the data, which helps distinguish the different classes. With the application of schemes based on deep learning, it has been possible to reduce the dependence on the design of functions and limit the manual selection of features; in this way, it is possible to dispense with human experience or great prior knowledge of the problem. With the emergence of deep learning, many fields of research have made use of these tools to facilitate the processing of massive data. In applications such as vision [9], image recognition [10], medical analysis [11], and other applications, the use of deep learning has obtained valuable results.

An example of application of schemes based on deep learning applied to industrial machines is presented in [12]; in this study, they implemented a structure of deep learning known as a stacked denoising auto-encoder to extract data characteristics from five data sets. Another application example is the approach proposed in [13]; in this study, they used a fully connected winner-take-all auto-encoder for the diagnosis of bearing faults, and the model is applied directly on temporary vibration signals without any time-consuming feature engineering process. The results indicate that the implemented method can learn from sparse features from input signals. In [5], they performed an unsupervised learning procedure for the automatic features extraction for the identification of bearing failures. First, they performed a non-linear projection to compress the information through a technique called compressed sensing, followed by the automatic feature extraction in transform domain using a DNN based on sparse stacked auto-encoders. The proposed approach highlights the effectiveness of extracting features automatically through the deep neural network, which demonstrate that they contain relevant information that helps the diagnostic process and thereby helps to reduce human labor. Another investigation in which CNN is applied for the diagnosis of faults in spindle bearings is presented in [14]. In this approach, the image is used as input for CNN to learn the complex characteristics of the system. Finally, the output is processed by a multi-class classifier. This method demonstrated a good classification efficiency regardless of the load fluctuation.

3.2 Dimensionality reduction

Deep learning has attracted attention in several fields of study because it allows the extraction of features from complex signals and the processing of large data. Although the application of deep learning in the diagnosis of faults in industrial machines has concentrated on the automatic extraction of features, the utility of these tools goes further; a clear example is the application of DNN structures for the compression or reduction of dimensionality of data. As we have seen above, structures based on DNN are able to learn intrinsic relationships of the data; however, during this learning process, it is possible to generate a reduced representation of the data. A structure based on DNN capable of learning a coded and reduced representation is the so-called auto-encoder. Unlike linear dimensional reduction techniques, such as principal component analysis (PCA) and linear discriminant analysis (LDA), a structure of stacked auto-encoders can provide a non-linear representation that was learned from the data provided. Therefore, a reduction of dimensionality based on the auto-encoder can provide a better representation that helps to discriminate between the conditions of the machine. An example of the difference between the application of a linear technique for the reduction of

dimensionality and one based on the auto-encoder is shown in **Figure 4(a)** and **(b)**, correspondingly.

The management of large data dimensions represents a problem and a challenge in different studies. This is reported in [15], where the generation of big data constitutes a challenge in schemes for protection against cyber-attacks. Therefore, they propose a methodology based on DNN for dimensionality reduction and feature extraction. The method is compared with other dimensionality reduction techniques. The results show that this approach is promising in terms of accuracy for real-world intrusion detection.

A research applied in monitoring the condition for diagnosis of rolling bearing is shown in [16]. In this study, they propose two structures of auto-encoder, a sparse auto-encoder (SAE) and denoising auto-encoder (DAE) for the dimensionality reduction and for the extraction of characteristics, correspondingly. The results show that the applied methodology can effectively improve the performance of fault diagnosis of rolling bearings.

3.3 Novelty detection

To avoid the incorrect evaluation of the health of the machinery, it is necessary to incorporate the current CBM schemes, the ability to classify data from novel scenarios or in test cases, where there is not enough information to describe anomalies. In this regard, research has been carried out to deal with the appearance of unknown scenarios in monitoring schemes. Novelty detection is the method used to recognize test data that differ in some aspects of the data available during training [17]. The study scenarios in which novelty detection schemes have been implemented include detection and medical diagnoses, damage detection in structures and buildings, image and video processing, robotics, and text data mining.

Recent contributions to novelty detection in CBM schemes have managed to combine the classic approaches of multi-faults detection and the ability to detect new operating modes [18]. This study has two main aspects; first, a new signal measurement is examined by a novelty detection model by one-class support vector machine (OC-SVM) method. If the measurement is cataloged as novel, the system is considered to be working under a new operation condition or a new fault. If the measurement is cataloged as known, the system is working under healthy or faulty condition, previously trained.

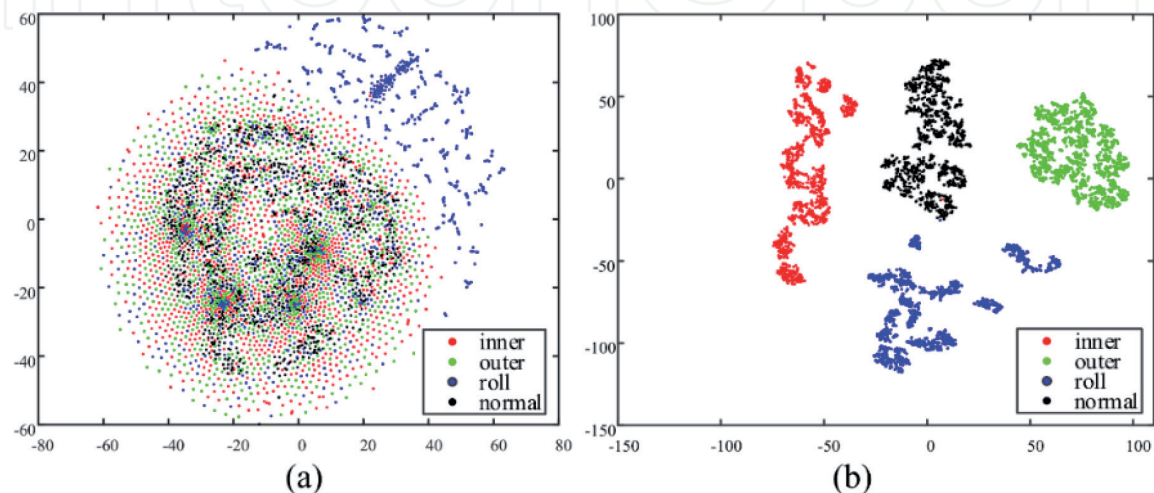


Figure 4.
Resulting two-dimensional by applying: (a) linear technique and (b) DNN architecture.

The task of novelty detection to recognize test data other than the data available during training depends on the method used. The novelty detection process consists of testing the data patterns that were not seen before and comparing them with the normality model, and this may result in a novelty score. The score, which may or may not be probabilistic, is generally compared to a decision threshold, and the test data is considered new if the threshold is exceeded. In applications that use dimensionality reduction to represent the patterns of the data in novelty detection schemes, it is common to find the projections of the data of the normal operation mode delimited by a region or frontier. In these studies, the samples that are outside that delimitation are considered as abnormalities. A representation of a space delimited by two characteristics is shown in **Figure 5**.

Detecting new events is an important need of any data classification scheme. Since we can never train a learning system under all conditions and with all possible objects with the data that the system is likely to find, it is important that it has the ability to differentiate between information from known and unknown events during testing. Many studies have faced in practice the challenging task involved in novelty detection. In this sense, several novelty detection models have been implemented, demonstrating that they work well in different data. Models to novelty detection include both Frequentist and Bayesian methods, information theory, extreme value statistics, support vector methods, other kernel methods, and neural networks.

On the other hand, although the use of DL-based techniques to carry out novelty detection tasks related to the study of the condition of electromechanical systems has not been reported in the literature, in other fields such as automatic driving, it has been proposed to use the reconstruction skills of the AE to carry out this task [19]. For this, the ability of the reconstruction of AE of the input data is used; if the error measurement is low, it is intuited that the input data correspond to known data, whereas if the error loss is high, they are considered unknown data and, therefore, they are data with which the system has not been trained. It is, therefore, believed that DL-based tools can represent a powerful analysis for the study of novelty detection in CBM schemes applied to electromechanical systems.

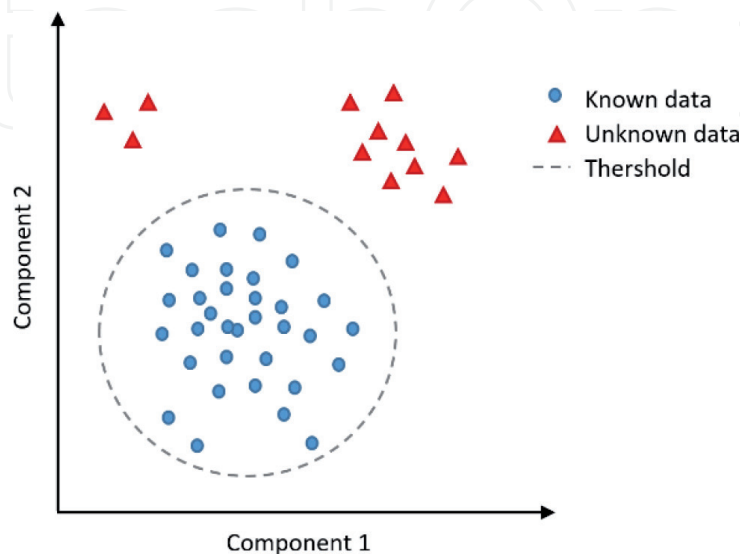


Figure 5.
Delimitation of a boundary by a novelty detection model.

3.4 Transfer learning

Some disadvantages that still prevails in many tasks of classification, regression, and grouping is that the approach that addresses this problem is made under the assumption that all data must be in the same working conditions and have the same distribution of data and space of characteristics to carry out those tasks. However, this assumption in the real world does not happen. This problem occurs because sometimes only a few training data are available for a domain of interest or working condition that is different or similar to that of the planned classification task. For these cases, knowledge transfer would help to improve the performance of the learning process, avoiding strenuous retraining work and the effort of data labeling. In this sense, various applications have begun to explore innovative techniques to address this problem, resulting in schemes based on transfer learning, domain adaptation, and various machine learning techniques.

As seen in the literature, schemes based on deep learning (DL) can learn complex and discriminative relationships from the data. Therefore, it has begun to use structures based on DL with the aim of transferring knowledge from a source task to a target task.

Traditional machine learning algorithms have made great strides in data-based fault diagnosis. They perform the diagnosis on test data using models that are trained on previously collected labeled or unlabeled training data. However, most of them assume that the data must be in the same working conditions and that the distributions of the data for each class considered are the same. The use of transfer learning schemes, in contrast, allows domains (operating conditions), tasks (failure classification), and distributions (number of samples) used in training and testing to be different.

Research on transfer learning has attracted more and more attention; as a result of which, one of the first learning techniques related to knowledge transfer is the multi-task learning framework, which tries to learn several tasks at the same time, even when they are different. In this scheme, transfer learning obtains knowledge of one or several source tasks and applies that knowledge to a target task, being the source task and target task symmetric in many ways. Unlike the learning of multiple tasks, the objective of transfer learning is the target task and not to learn all the source tasks and target tasks at the same time. The roles of the tasks of source and target are no longer the same, but they are similar in the transfer of knowledge.

Figure 6 shows the difference between the learning processes of traditional learning techniques and transfer learning. As we can see, traditional machine learning techniques try to learn each task from scratch, whereas transfer learning techniques try to transfer the knowledge of some previous tasks to a target task when the latter has differences, but also similarities with the source task.

One of the investigations related to transfer learning applied to the diagnosis of faults in industrial systems is the one presented in [20]. In this study, they use the skills of deep learning schemes to extract features with hierarchical representation samples in frequency domain and combine it with a transfer learning process to consider a target task different from the source task. The results obtained show a considerable performance; however, the proposed scheme still considers that the samples of the source domain and the target domain are equal.

Another work related to transfer learning is the one proposed in [21], for the diagnosis of bearing failures. Their proposal analyzed different operating conditions for the source task and the target task. The knowledge transfer process is done through a structure based on neural network, where it first learns the characteristics of a source task, followed, that structure is partially modified to adapt to a new

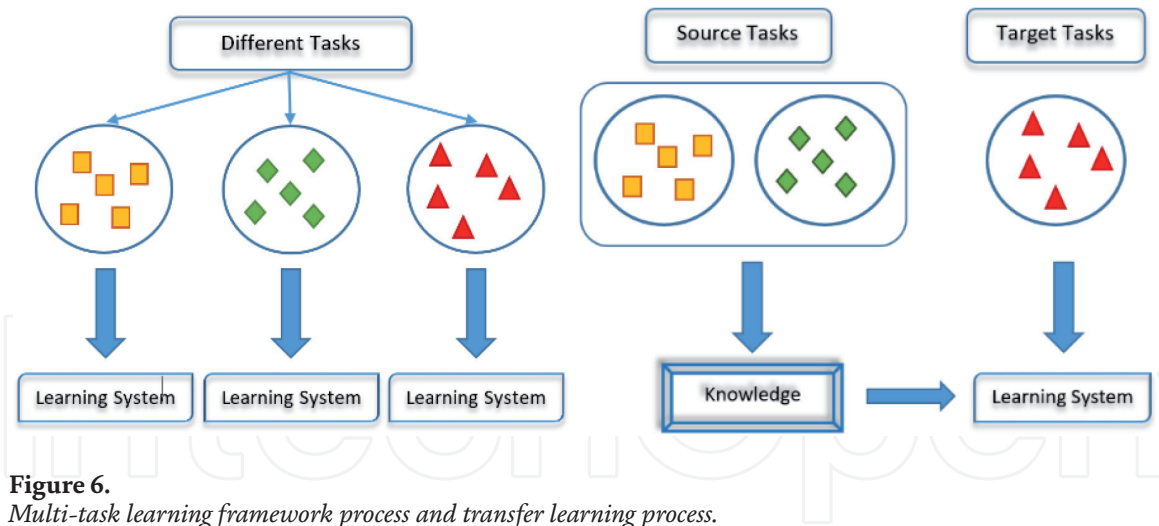


Figure 6. Multi-task learning framework process and transfer learning process.

target task; however, it conserves part of the weights with which the homework network was trained. The obtained results showed that in some occasions, using a method with knowledge transfer improves the diagnostic performance. However, this performance is affected when the differences between the source task and the target task are increased. With the incorporation of schemes based on transfer learning, we can allow us to adapt different structures based on DL to transfer the experience learned in a diagnostic task and improve performance in a similar but different task.

4. Experimental case of deep learning in CBM

As a case study, the comparison of three different approaches to carry out the process of dimensional reduction in a diagnostic analysis of multi-faults in an electromechanical system is presented, by applying two linear techniques: principal component analysis (PCA) and linear discriminant analysis (LDA), and a technique based on deep learning: an auto-encoder.

The proposed case study to evaluate the performance of multiple fault diagnostic detection in an electromechanical system under the three different schemes is presented in **Figure 7**. First, signal conditioning and acquisition is carried out over vibration signals. Second, the estimation of the 15 statistical-time-based features, such as rms, skewness, mean, kurtosis, impulse factor, etc., is done over each signal. Third, the study of three high-dimensional feature reduction methods, that is, principal component analysis and linear discriminant analysis and sparse auto-encoder, is carried out. Finally, fourth, an NN-based classification structure is performed, where the fault diagnosis and corresponding probability value are obtained. The resulting performance of the considered scheme is analyzed in terms of classification in front to different high-dimensional feature reduction schemes. In addition, it is worth mentioning the resulting projections into a two-dimensional space with an accumulated variance of 95 between the two axes, in the case of PCA analysis. While under an AE study, the effectiveness is measured through the calculation of the MSE reconstruction error, which after 1200 epochs for each of the hidden layers is approximately 0.06.

The goal of the proposed approach is to evaluate the information extraction and dimensionality reduction capabilities of a non-linear technique such as auto-encoder. For this, a methodology based on the study of the condition using vibration signals is implemented. For different condition, they have been considered to be evaluated in terms of the induction motor: healthy condition (He), bearing fault

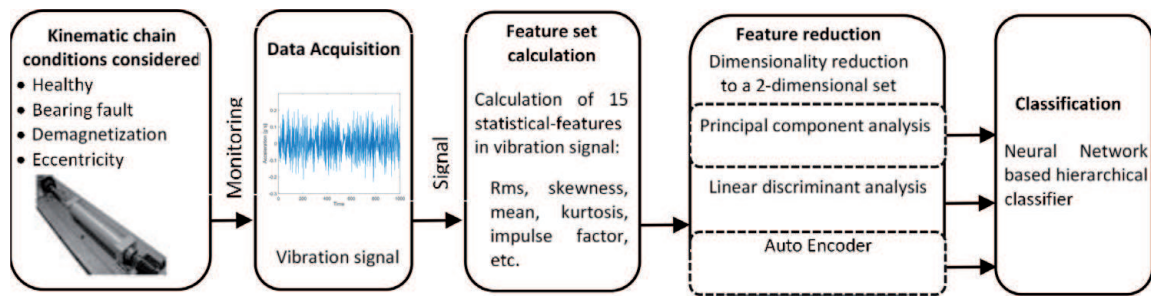


Figure 7.
 Proposed scheme for multiple faults diagnosis.

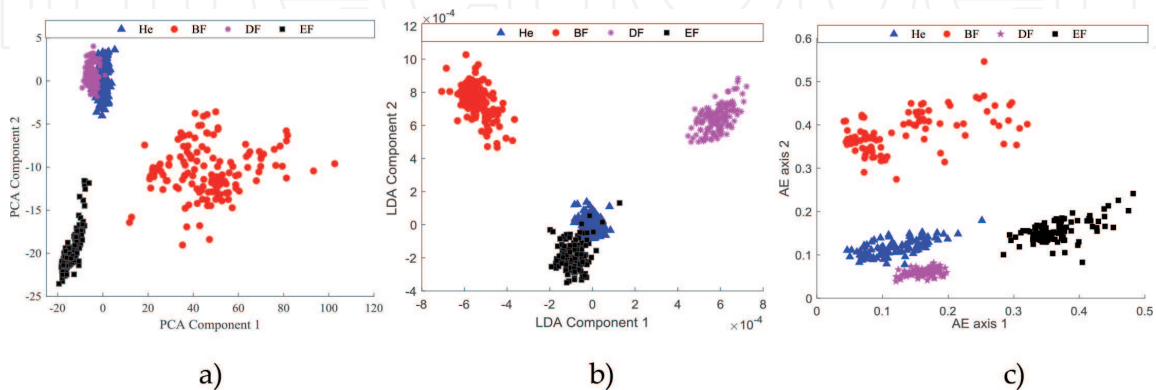


Figure 8.
 Evaluation of resulting projections into a two-dimensional space. (a) PCA. (b) LDA. (c) Auto-encoder.

(BF), demagnetized fault (DF), and eccentricity fault (EF). In order to numerically characterize the acquired physical magnitudes, a 1-s segmentation is proposed. For each segment, a set of statistical-time features is calculated. To verify the effectiveness of a non-linear dimensionality reduction technique, the projections resulting from the process of reduction of the three techniques are shown in **Figure 8**.

Finally, the classification stage with the NN-based classifier has been configured with five neurons in the hidden layer, besides a logistic sigmoid function has been used as output activation function and 100 epochs are considered for training using the backpropagation rule. The classification ratios for the test sets are approximately 95% for PCA, 98% for LDA, and 99% for auto-encoder.

Two important things can be concluded from this study: first, highlight the capabilities of an SAE-based approach to automatic learning of the most significant characteristics (those that provide more discriminative information) and that this translates into an increase in performance. Second, in regard with the dimensionality reduction, the auto-encoder-based approach shows better discriminative capabilities during the visualization of the results than the linear methods PCA and LDA, with it facilitates the task of classification.

5. Conclusion and future challenges

In this chapter, a review of some of the current techniques based on deep learning and some of the functionalities that they may have within the environment of the diagnostic schemes of electromechanical systems is carried out. Having as reference the high complexity that is increasingly being found in the manufacturing processes, and the new challenges to face in the Industry 4.0 paradigm, it is necessary to improve the diagnostic capabilities of traditional schemes, which is why methodologies based on artificial intelligence and deep learning methods have

increasingly called the attention of researchers. However, it remains to be discovered and identified the patterns that these deep neural networks learn, and specifically, within the industry environment, and electromechanical systems, what is the scope and benefits of applying these novel techniques.

IntechOpen


IntechOpen

Author details

Francisco Arellano Espitia* and Lucia Ruiz Soto
Technical University of Catalonia, Barcelona, Spain

*Address all correspondence to: Francisco.arellano@upc.edu

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 

References

- [1] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain, Cornell Aeronautical Laboratory. *Psychological Review*. 1958;**65**(6):386-408
- [2] Stuart G, Spruston N, Sakmann B, Häusser M. Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences*. 1997;**20**(3):125-131. DOI: 10.1016/s0166-2236(96)10075-8
- [3] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Neural Information Processing Systems*. 2012;**25**(2):84-90. DOI: 10.1145/3065386
- [4] Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*. 2016;**89**:171-178
- [5] Sun J, Yan C, Wen J. Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation and Measurement*. 2018;**67**(1):185-195
- [6] Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996;**381**:607-609
- [7] Ma X, Wang X. Convergence analysis of contrastive divergence algorithm based on gradient method with errors. *Mathematical Problems in Engineering*. 2015: Article ID 350102, 9 p
- [8] Liu R, Yang Y, Zhao Z, Zhou J. A novel scheme for fault detection using data-driven gap metric technique. In: *IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*; Enshi. 2018. pp. 1207-1212
- [9] Zheng J, Cao X, Zhang B, Zhen X, Su X. Deep ensemble machine for video classification. *IEEE Transactions on Neural Networks and Learning Systems*. 2019;**30**(2):553-565
- [10] Yuan Y, Mou L, Lu X. Scene recognition by manifold regularized deep learning architecture. *IEEE Transactions on Neural Networks and Learning Systems*. 2015;**26**(10):2222-2233
- [11] Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Transactions on Medical Imaging*. 2016;**35**(5):1313-1321
- [12] Thirukovalluru R, Dixit S, Sevakula RK, Verma NK, Salour A. Generating feature sets for fault diagnosis using denoising stacked auto-encoder. In: *2016 IEEE Inter. Conf. on Progn. & Hlth. Manag. Ottawa, ON: ICPHM*; 2016. pp. 1-7
- [13] Li C, Zhang W, Peng G, Liu S. Bearing fault diagnosis using fully-connected winner-take-all autoencoder. *IEEE Access*. 2018;**6**:6103-6115
- [14] Ding X, He Q. Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*. 2017;**66**(8):1926-1935
- [15] Abolhasanzadeh B. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features. In: *2015 7th Conference on Information and Knowledge Technology*. Urmia: IKT; 2015. pp. 1-5
- [16] Zhang J, Chen Z, Du X, Xu X, Yu M. Application of stack marginalised sparse denoising auto-encoder in fault diagnosis of rolling bearing.

The Journal of Engineering.
2018;**2018**(16):1772-1777

[17] Pimentel MA, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. *Signal Processing*. 2014;**99**:215-249

[18] Carino JA, Delgado-Prieto M, Zurita D, Millan M, Ortega Redondo JA, Romero-Troncoso R. Enhanced industrial machinery condition monitoring methodology based on novelty detection and multi-modal analysis. *IEEE Access*. 2016;**4**:7594-7604

[19] Amini A, Schwarting W, Rosman G, Araki B, Karaman S, Rus D. Variational autoencoder for end-to-end control of autonomous driving with novelty detection and training De-biasing. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid. 2018;**2018**:568-575

[20] Lu W, Liang B, Cheng Y, Meng D, Yang J, Zhang T. Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*. 2017;**64**(3):2296-2305

[21] Wen L, Gao L, Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2019;**49**(1):136-144