UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Enginyeria Electrònica

Deep Learning Based Methodologies applied to Industrial Electromechanical Systems Monitoring

Thesis submitted in partial fulfilment of the requirement for the PhD Degree issued by the Universitat Politècnica de Catalunya, in its Electronic Engineering Program.

Francisco Arellano Espitia

Supervisors:

Dr. Miguel Delgado Prieto

Dr. Roque Alfredo Osornio Ríos

July, 2023

Xijtemiki, ximonekilli, xichiua

Abstract

In the last years, the rotating systems condition based maintenance field has been received focus from both academic and industry sector. This is due to the great relevance of these systems in the industrial sector, especially for production and manufacturing processes. To guarantee the high efficiency, safety and correct performance of the different components that can form a rotating system, such as motors, bearings, gearboxes and shafts, it is essential to have monitoring systems, schemes and methodologies with the ability to diagnose and detect malfunctions on them. However, monitoring schemes present great challenges in dealing with the complexity these of systems, that implies a working environment subject to different operation condition, configurations with multiple components and the presence of faults of different nature, *i.e.*, mechanical, electrical, electromagnetic, under isolated or combined scenarios. Recently, artificial intelligent-based techniques have been widely used in monitoring schemes, especially artificial neural networks. Due to its great capacity for characterization and feature extraction, an effort has been made to study methodologies based on these techniques. In addition, multi-layered architectures, known as deep neural networks, represent a considerable advance in the field of pattern recognition. However, the implementation of these schemes in industrial environments still presents a continuous challenge due to the requirements, such as high reliability, robustness and scalability that must be accomplish, especially in the new Industry 4.0 paradigm.

In this regard, this thesis consists of the investigation and proposal of a series of complementary methodologies based on deep learning that lead to the implementation of predictive maintenance schemes for the industrial sector. Initially, a data-driven monitoring methodology based on deep learning for fault diagnosis is proposed for complex electromechanical system environments. Next, a methodological process to for anomaly identification composed of deep learning strategies and machine learning tools is proposed. Finally, an intelligent diagnostic model is proposed using tools based on adaptation to the adversary domain. This leads to the creation of a scalable model with the ability to diagnose in a general way, fault scenarios under different operating conditions and solve the problem of performance degradation.

The study and validation of the proposed methodologies have been carried out using experimental databases of electromechanical systems produced in the laboratory. Quantifiable performance metrics and qualitative validation procedures are implemented.

Keywords

Condition Monitoring | *Fault Diagnosis* | *Industrial Machines*
Artificial Intelligence | *Deep Learning* | *Anomaly Detection*
Domain Adaptation
Compact Clustering | *Transfer Learning*

Resumen

En los últimos años, el campo de mantenimiento basado en la condición de los sistemas rotativos ha recibido atención tanto del sector académico como de la industria. Esto se debe a la gran relevancia de estos sistemas en el sector industrial, especialmente para los procesos de producción y manufactura. Para garantizar la alta eficiencia, seguridad y correcto desempeño de los diferentes componentes que pueden formar un sistema rotativo, tales como motores, rodamientos, cajas de engranes y ejes, es fundamental contar con sistemas de monitoreo, esquemas y metodologías con la capacidad de diagnosticar y detectar mal funcionamientos de las mismas. Sin embargo, los esquemas de monitoreo presentan grandes desafíos al tratar con la complejidad de estos sistemas, que implica un ambiente de trabajo sujeto a diferentes condiciones de operación, configuraciones con múltiples componentes y la presencia de fallas de diferente naturaleza, es decir, mecánicas, eléctricas, electromagnéticas, bajo condiciones aisladas o escenarios combinados. Recientemente, las técnicas basadas en inteligencia artificial se han utilizado ampliamente en esquemas de monitoreo, especialmente las redes neuronales artificiales. Debido a su gran capacidad de caracterización y de extracción de características, se ha hecho un esfuerzo por estudiar metodologías basadas en estas técnicas. Además, las arquitecturas multicapa, conocidas como redes neuronales profundas, representan un avance considerable en el campo del reconocimiento de patrones. Sin embargo, la implantación de estos esquemas en entornos industriales aún presenta un desafío continúo debido a los requerimientos, como la alta confiabilidad, robustez y escalabilidad que deben cumplirse, especialmente en el nuevo paradigma de Industria 4.0.

En este sentido, esta tesis consiste en la investigación y propuesta de una serie de metodologías complementarias basadas en aprendizaje profundo que conducen a la implementación de esquemas de mantenimiento predictivo para el sector industrial. Inicialmente, se propone una metodología de monitoreo de datos basada en aprendizaje profundo para el diagnóstico de fallas para entornos de sistemas electromecánicos complejos. A continuación, se propone un proceso metodológico para la identificación de anomalías compuesto por estrategias de aprendizaje profundo y herramientas de aprendizaje automático. Finalmente, se propone un modelo de diagnóstico inteligente utilizando herramientas basadas en la adaptación de dominio contradictorio. Esto conduce a la creación de un modelo escalable con la capacidad de diagnosticar de manera general, escenarios de fallas bajo diferentes condiciones de operación y resolver el problema de degradación de rendimiento. El estudio y la validación de las metodologías propuestas se han llevado a cabo mediante bases de datos experimentales de sistemas electromecánicos producidos en laboratorio. Se implementaron métricas de rendimiento cuantificables y también procedimientos de validación cualitativa.

Keywords

Condition Monitoring	*Fault Diagnosis*	*Industrial Machines*
Artificial Intelligence	*Deep Learning*	*Anomaly Detection*
Compact Clustering	*Transfer Learning*	*Domain Adaptation*

Acknowledgments / Agradecimientos

Llegar aquí y haber realizado este trabajo no hubiera sido posible sin la ayuda, el apoyo y el consejo de muchas personas. Mencionar a cada persona que ha estado presente de alguna manera es difícil, pero se los agradezco.

A mis padres, Teresita y Francisco, por su amor incondicional, por enseñarme siempre a ser una mejor persona, por sus invaluables aprendizajes, por todo su apoyo.

A mis hermanas Diana y Arely, por estar para mí siempre que las he necesitado, por han hecho saber que siempre tendré su confianza.

A los directores de mi tesis, Roque, por brindarme la formación necesaria para poder alcanzar este objetivo, a Miguel, por ser una guía constante y aceptar prepararme para afrontar este trabajo, sin su orientación habría sido un camino difícil de superar. Estoy agradecido por sus consejos y productiva dedicación.

A todo el grupo de trabajo de MCIA, al director Luis Romeral por la atención brindada todo este tiempo, por sus consejos sobre ingeniería y aspectos de no-ingeniería. A mis compañeros de despacho, Eva y Víctor, no sólo fueron compañeros, si no también amigos. A Alejandro, Carles, Geovany, Pablo, Erick y Joan por acompañar en esta travesía, por los buenos momentos y hacer agradables los cumpleaños y celebraciones.

A mis colegas y amigos en México, Elizabeth y Betzy por su amistad duradera que atraviesa distancias. A mis compañeros en la UAQ, al doctor René, Juan José, Darién e Israel, por su colaboración y apoyo.

Finalmente agradecer al Consejo Nacional de Ciencia y Tecnología – Conacyt por conceder el financiamiento a través de la beca durante la realización que condujo a esta tesis.

Muchas gracias.

Contents of the document

Index of figures

Index of tables

Acronyms and their definitions

AE	Auto-encoder
AI	Artificial Intelligence
ANN	Artificial Neural Networks
BP	Backpropagation
CBM	Condition Based Maintenance
CNN	Convolutional neural networks
CORAL	Correlation alignment
CPS	Cyber-Physical Systems
DA	Domain adaptation
DAE	Deep auto-encoder
DAN	Domain adaptation network
DANN	Domain Adversarial Neural Network
DBN	Deep belief network
DCC	Deep-compact-clustering
DCNN	Deep convolutional neural network
DDA	Deep domain adaptation
DEC	Deep embedded clustering
DL	Deep-Learning
DNN	Deep neural network
DTCWT	Dual-tree complex wavelet transform
DTL	Deep-transfer learning
EMD	Empirical Mode Decomposition
EMDS	Electrical Motor-Driven Systems
FD	Frequency domain
FE	Feature engineering
FFT	Fast Fourier Transformation
GA	Genetic Algorithms
GAN	Generative Adversarial Networks
GMM	Gaussian mixture model
GPU	Graphics processor unit
GRL	Gradient Reversal Layer
GRU	Gated recurrent units
HHT	Hilbert-Huang transform
HVAC	Heating, ventilation, and air conditioning
IEA	International Energy Agency
IFD	Intelligent fault diagnosis
IIoT	Industrial Internet of Things
IMF	Intrinsic mode function

IT-OT	Information Technology and Operational Technology
KLD	Kullback–Leibler divergence
KWES	Key World Energy Statistics
LDA	Linear discriminant analysis
LSTM	Long-short term memory
MAE	Mean absolute error
MDF	Multiple-Domain-Features
ML	Machine Learning
MMD	Maximum mean discrepancy
MVD	Maximum variance discrepancy
NLP	Natural language processing
NN	Neural Network
OCC	One-class classification
OC-SVM	One-Class Support Vector Machine
OEE	Overall Equipment Effectiveness
PCA	Principal component analysis
PDF	Probability density function
RBM	Restricted Boltzmann machine
ReLU	Rectified linear unit
RMS	Root mean square
RMSE	Root mean square error
RNN	Recurrent neural network
RUL	Remaining useful life
SAE	Stacked auto-encoders
STFT	Short-Time Fourier Transform
SVDD	Support vector data description
SVM	Support Vector Machine
Tanh	Hyperbolic tangent
t-SNE	t-distributed stochastic neighbor embedding
TD	Time domain
TFD	Time-frequency domain
WD	Wasserstein Distances
WPT	Wavelet packet transform

1.

Introduction

This chapter describes the bases on which this thesis research was built. It begins with a brief introduction to the research topic and the current limitations, and continues with the objectives and the statement of the hypotheses. This chapter also includes a description of the organization of the following chapters.

CONTENTS:

1.1 Research topic

In the last decade, manufacturing processes have been characterized by greater integration between the Information Technology and Operational Technology (IT-OT), in response to global market competition, the demand for a high efficiency in the processes and product quality increase [1]. This has given rise to smart manufacturing environments that have achieve to the new revolution in industry, calling the term Industry 4.0. In this regard, as a result of the integration of computer systems, networking and physical processes that involve control and monitoring mechanisms, a new line of intelligent systems has developed, formed from the close integration of cyber and physical systems. Cyber Physical Systems (CPS) is a platform consisting of a computer system which has the function of controlling and monitoring mechanical systems. The physical-mechanical components of the CPS platforms are represented by machinery, smart sensors and actuators, while the software components are represented by computer algorithms and networking devices, all of them closely connected.

However, despite technological advances in manufacturing processes, unplanned stops still represent a crucial issue. A stop-time in industrial plant machinery has a direct impact on production and, therefore, in the economic performance. According to the Overall Equipment Effectiveness (OEE) [2], that corresponds to the metric used to assess how well a manufacturing process is utilized compared to its full potential, the stoppages that cause the greatest losses are due to unplanned breakdowns, followed by configuration procedures for machinery adjustments. In this regard, Electrical Motor-Driven Systems (EMDS) still represent one of the key components in the manufacturing industry, being the main mechanical energy supplier in the industry. In fact, the massive use of electric motors as drivers of industrial processes represents between 43% and 46% of all global electricity consumption [3]. Thereby, EMDS may be found in many industrial applications, such as power generation systems, in the aerospace industry, machining tools, transportation, robotics, and mining [4-8]. The main applications for EMDS fall within the following sectors (**Fig. 1.1.1**): industrial: pumps, fans, compressed air delivery, conveyors, motive power for other machinery, high precision motors for robotics, etc.

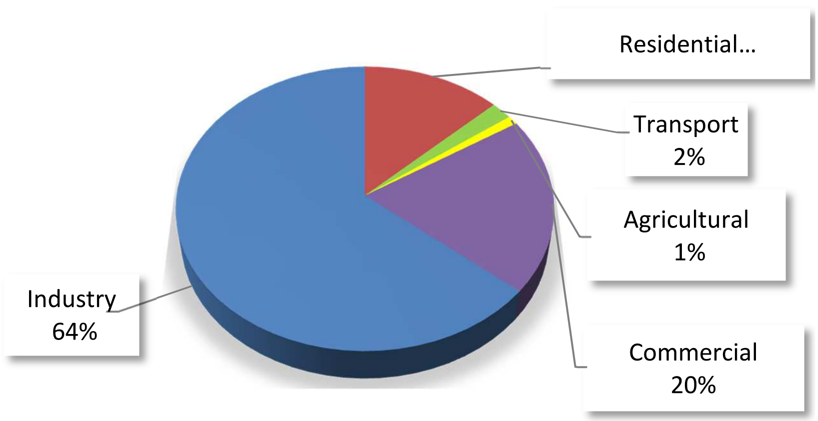


Fig. 1.1.1 Distribution of the use of the main applications for EMDS[1].

[1] International Energy Agency (IEA), IEA Key World Energy Statistics (KWES), statistics 2021 (motors).

Commercial building: pumps, fans, conveyors, lifts, compressors in HVAC systems, etc. Residential: household appliances, air conditioners, IT fans/drives, cooking appliances, extractor fans, garden appliances, pool pumps, and others.

Therefore, with the widespread use of EMDS, systems generally have different configurations that vary depending on the specific task and include various components, such as couplings, bearings, belts, pulleys, shafts and gearboxes among others. In some specific applications there are replicas with the same configuration of elements, such as electric generators or production lines. However, the working conditions of each system are particular, so component wear and breakdowns do not occur with the same severity or under the same conditions. This variability in the configuration makes it difficult to diagnose faults, which are the main reason for stops in manufacturing processes. In this regard, to face the challenges and demands of the manufacturing environment, both the academy and the industrial sector have made efforts to generate strategies for the evaluation and detection of faults in the different elements of the EMDS.

In this regard, many research has emerged in the field of machine health monitoring. These research have leads to the generation of Condition Based Maintenance (CBM) strategies to deal with the loss of effectiveness in industrial systems. Major diagnostic strategies include model-based, signal-based and knowledge-based fault diagnosis approaches, being the knowledge-based fault diagnosis approaches the ones that have gained the most popularity in modern manufacturing systems. This is mainly due to the increase in complexity in industrial environments, more and more the conformation of machinery is composed of a greater number of elements that interact with each other, from electrical components to mechanical elements. At the same time, the behavior of the operating patterns, including faults and degradations, evolves with respect to their use regimes and environmental conditions, which generates a behavior drift with respect to ideal conditions over time. All these considerations make it necessary to generate a monitoring scheme that follows all these changes that occur in the manufacturing equipment throughout its useful life. This is possible through learning the behavior patterns of the machinery through its own data. In addition, for monitoring to accomplish the desired robustness, three fundamental aspects must be satisfying for CBMs: diagnosis, detection and adaptability.

First, a correct **diagnosis**, that is, based on the available information, determine the condition of the monitored system. This challenge is focused on characterizing and analyzing individual and multiple faults in the different elements of the machine. This requires knowledge-based models capable of extracting knowledge from the operating state and accurately classifying the condition, whether healthy or the type of failure that has occurred.

The second aspect is the **detection** capacity. Unlike diagnosis, in which known states are characterized and then an attempt is made to classify them accurately, detection seeks to identify unknown patterns that have not been previously characterized.

The importance of detection is due to the fact that it is not always possible to characterize all the failure states or identify novel operating conditions of the system. Therefore, detect when abnormal behavior occurs, *i.e*. a new fault or a different operating condition represents a challenge owing to these types of approaches are performed in an unsupervised or semi-supervised way. Dealing with unsupervised schemes implies that there is no labeled data, and therefore it is necessary to implement strategies to deal with this lack of information.

Third, have the ability to **adaptation** to varying operating conditions. In real-world applications, most of the EMDS are driven to different work conditions, therefore, the models with which data-driven approaches are initially trained may differ from the conditions with which they are tested. This difference in working conditions causes what is called the domain-shift problem, which means that the trained model will not be able to generalize well on test data with different working conditions or distributions (*e.g.*, working loads and/or speed). In this regard, it is necessary to take into account methods that will be able to generalize well on the data with different distributions. Methods based on transfer learning have been considered as a powerful tool that allows retaining, accumulating and increasing knowledge. The use of these methods will allow the creation of more generalizable and adaptive schemes that deal with the domain-shift problem present in modern industrial machinery.

Indeed, the consideration of these three aspects will result in the generation of more robust schemes. However, although there is research in each of these directions, it is necessary to reframe the existing schemes to face the challenges that arise when dealing with complex industrial systems. In this regard, although different aspects of electromechanical systems are taken into account, the study always focuses on working with replicas of the same system. The variability considered is about multiple operating conditions, the presence of different faults or the change of material of some component (*i.e.*, metallic or ceramic bearings). Although there are investigations that consider the study of different systems, there is still ambiguity about the existence of characteristic fault patterns that are domain invariant (machine, operating conditions, acquisition system).

Additionally, the technological advances of recent years, such as the access and deployment of hardware and software, advances in sensors, the acquisition and processing of large amounts of data, communication tools and algorithms based on Artificial Intelligence (AI) have allowed the emergence of a new generation of CBM data-driven approaches. Multiple studies of CBM based on Machine Learning (ML) as a subfield of AI, have been proposed in recent years. And although they have made significant progress in machine health monitoring, their use is mainly limited to simple scenarios with a few working conditions (*i.e.* speed and torque combinations), over a restricted number of faults and severities and focused in specific components over the electromechanical chain.

Nevertheless, when the monitoring and diagnostic requirements increase (*i.e.* due to the need to consider multiple working conditions, multiple electromechanical components and their possible interactions and/or a higher resolution in the diagnosis - fault severity-), the performances of approaches based on in classical ML techniques decrease.

It has been in recent years where, promoted by the good results in other application sectors such as characterization of clinical images [9], natural language processing [10], and computer vision analysis [11], **Deep-Learning-based** AI techniques are beginning to position themselves as the technological solution to the aforementioned problems. Deep learning is a specialized form of machine learning, characterized by the use of artificial neural network architectures composed of multiple processing stages. As mentioned above, the performance of ML techniques tends to converge, especially under advanced complexity environments, that is, a deadlock is reached at a certain level of performance. On the contrary, one of the advantages of DL-algorithms is that their performance scales with data, that is, performance often continues to improve when more examples and training data are added to the network.

The presence of multiple processing levels allows extracting more complex relationships from the data, and establishing hierarchical knowledge. This way of processing the information of the DL-models is similar to the learning process of the human brain, which enables them to be implemented in various tasks. DL-models have been used as tools in image recognition and natural language processing tasks, thereby improving applications such as autonomous vehicles and language translation services [12-15]. The use of DL has extended in various research fields, meanwhile, its integration into the field of machine health monitoring has been in progress.

However, among the main challenges of implementing DL-models in the field of machine health monitoring is that there are no clear methodological processes to be applied. This serious drawback limits the massive application of such technology, since the interpretability of the resulting DL based methodology is not being considered as a fundamental part of the proposed solutions. In consequence, there is a high risk of overfitting of the resulting models, and the reliability of the outcomes is compromised.

Generally, in applications that involve DL-based methods, there is no consensus on the design of the architecture that should be used, neither on the processing that should be done on the data that is being worked on. In addition, for DL architectures to be efficient and robust, they must comply with a correct application to the study environment. This includes the setting of hyperparameters and tests for the start-up, always trying to maximize the performance of the models.

In this sense, considering the current environment of intelligent manufacturing, taking into account the most innovative tools based on Artificial Intelligence and the necessary requirements around the field of machine health monitoring define the research topic in which this thesis is carried out: **deep-learning based methodologies for electromechanical drives monitoring**.

1.2 Research problem

One of the main problems related with the maintenance in the industrial sector is related to the increase of the corresponding key performance indicators (e.g. OEE), while the complexity of the machinery under monitoring is increasing (i.e. complex electromechanical structure and demanding operating cycles). In the last years, this scenario is leading to a significant research and innovation effort around the integration of new technologies such as IIoT, CPS and IA.

The study of rotating machinery implies considering the presence of faults in different elements and also that they are operating at different working conditions, that is, different speeds and loads (torques). This fact often leads to the generation of highly clustered spaces, so pattern characterization is a crucial issue. In this regard, the selection of the deep-learning algorithm to characterize all the aforementioned variability for an electromechanical system is not clear enough. Although there are different methods to approach the characterization and that have been applied effectively in other research areas such as the computer vision field [16], natural language processing [17], automatic speech recognition [18], the implementation in electromechanical systems is still not clear. Deep learning methods have several variants such as auto-encoders [19], convolutional neural networks [20], deep boltzmann machines [21], recurrent neural networks [22] and deep belief network [23]. It is necessary to face and **evaluate the feasibility of the application of DL-based methods in the field of machine health monitoring of rotary systems**, as well as which method is more convenient to satisfy the requirements of this field.

While deep learning-based algorithms are powerful for learning general features, it is still difficult to learn discriminative features and characteristic patterns for machine diagnosis from complex signals. Since useful diagnostic features are only a small part of the signal information, especially when other information patterns coexist such as environmental noise, harmonics caused by machine rotation, working load, etc. In addition, the availability of data can come from different sources of information, such as vibration signals, stator currents, thermographic images, acoustic emission and it is difficult to discriminate which information provides value to perform the diagnosis. Besides, the proposal of different preprocessing methods of different domains, such as time domain, frequency domain, and time-frequency domain makes it difficult to reach a consensus to build meaningful inputs for deep-learning techniques. Therefore, **adequate strategies are required to consider the different sources of information and to be able to extract features effectively** regardless of the domain in which it is presented.

Classically, data-driven CBM schemes consists of the following three stages: (i) feature extraction, (ii) feature selection/reduction and, (iii) model training. Three important aspects characterize this approach: the first one is related to the manual characterization (feature calculation), that requires a significant amount of expertise from the professional, which is not always available due to the large amount of human labor involved and is also unfeasible due to the complex interactions among machinery components (e.g. gears, wheels, screws, bearings). The second one is related to the structural independency among stages, since although there are ways to include their performance in a global cost function, each stage has its own optimization criteria (e.g. variance maximization, Fischer score minimization or RMSE among others)., which affects negatively the performance of the entire approach. Finally, the third one, is related with the own structure of the algorithms that are, generally, linear based approaches (i.e. PCA and LDA), and/or supported by a limited structure in terms of non-linear relations and non-Riemannian manifolds characterization.

Data-driven approaches based on deep learning are built (in their basic representation) through architectures based on deep neural networks with multiple layers of non-linear transformations, which aim are extract hierarchical representations from input data. Therefore, each layer represents a transformation from input values to output values. Likewise, by establishing stacked structures, they allow learning complex concepts from the data, enabling them as a powerful tool for the characterization of machine health monitoring. However, the application of these tools in the field of machinery monitoring is not simple and is still a challenge. There is no consensus for the criteria to be used during the configuration of hyper-parameters, which leads to overfitted algorithms, not easily replicable and, therefore, limiting the applicability. Therefore, the research on **methodological processes capable of addressing the correct configuration of hyper-parameters using deep-learning algorithms** applied to the field of electromechanical system monitoring is required. Where "correctness" must be assessed in terms of replicability, adaptability, generalization and performance.

As well, CBM data-driven approaches must be guided to an incremental learning environment. Rotating machinery often works with different working conditions, such as different loads and speeds. However, all these working conditions are not always available at the time the models are trained. Considering new operating conditions and incorporating them into already trained models is a research problem that has been considered recently. Training a model from scratch to incorporate new operating conditions has been a way to facing this challenge, however, performed every time there is a change in operating conditions is not feasible and entails exhausting work. Transfer learning (TL), a recently emerged strategy in the field of machine learning, which aims to use knowledge gained from performing one task and apply it to a different but similar task, has been explored as a way to address the issue of change of domain (working conditions).

In this regard, an effective strategy is retraining or fine-tuning the deep learning models to incorporate new working conditions [24]. However, newly collected data under the new working conditions are often unlabeled or have few labeled samples. This usually occurs in the field of monitoring industrial systems, where keeping the machine operational in a fault state can be catastrophic. Although fine-tuning algorithms are comparatively easy to implement, their performance would decrease dramatically if the labeled samples are insufficient or unavailable. Indeed, to **work under an incremental learning framework, it is necessary to consider adaptive or evolving algorithms for fault diagnosis** in order to improve the reliability and robustness of the approaches that are implemented on electromechanical systems.

Summarizing, and facing the needs and requirements still present in CBM schemes for industrial rotary systems, a new series of questions that the literature have not yet addressed:

In regard to the implementation of deep learning algorithms in electromechanical systems

(i) Which algorithms based on deep learning are appropriate to satisfy the requirements of diagnosis, detection and adaptation considering electromechanical systems?

In regard to the information available, the type of signals and signal processing strategies proposed in the literature to characterize an electromechanical system

(ii) What kind of signals and processing are most suitable for using as input for DL-based models in the field of machine health monitoring? Furthermore, how to properly select and discriminate information from different data sources?

In regard to the configuration of hyper-parameters of algorithms based on deep-learning

(iii) Given that there is no common framework for the application of DL-based algorithms, what is the methodological process that should be followed for its implementation in the field of condition monitoring?

In regard to facing changes in operating conditions and the incremental learning framework

(iv) What approaches are most appropriate to address the incremental learning framework? Besides, how to design methodologies to incorporate more information regarding different operating conditions?

In summary, considering all the requirements and needs that are still present in monitoring schemes for industrial rotating systems, the scientific community must continue to make an effort to study and define new contributions in this field, with the aim of proposing methodologies for analysis and action to address monitoring in the smart manufacturing environment. It should be noted that this research topic involves a vanguard research environment that is highly innovative in its application to the field of machine health monitoring. In this regard, this thesis aims to answer all the questions mentioned above.

1.3 Hypotheses

In order to address the presented research problems and considering the state of the art, the following hypotheses have been formulated as a starting point for this research work:

H1 In order to improve the performance of condition monitoring schemes in electromechanical systems, advanced strategies based on deep learning algorithms can be implemented and adapted through methodological procedures that are generalizable and interpretable.

H2 It is possible to generate methodologies that maintain high performance in the task of fault diagnosis in highly complex scenarios of industrial systems by applying automatic and interpretable procedures based on deep-learning algorithms capable of learning and extracting features from the acquired signals.

H3 The problem of detecting anomalies in highly clustered spaces can be addressed through the application of strategies that improve and compact feature spaces in combination with the advantages offered by deep-learning techniques to learn and reconstruct representations, which allows characterize known modes of operation and help detect outlier conditions.

H4 Improving the robustness and generalizability of condition monitoring approaches can be achieved through the adoption of incremental learning tools in combination with transfer learning strategies, in which the convergence of different operating conditions is considered.

In consequence,

The implementation of consistent methodological processes, together with advanced artificial intelligence algorithms, will enable greater performance and reliability in the field of monitoring. The assumptions exposed above represent the basis of the thesis research proposal. The hypotheses are addressed through the research work reflected in this thesis document.

1.4 Aim and objectives

In order to address existing research problems and test the research hypotheses raised, the objective of this thesis consists in the proposal and investigation of novel methodologies for monitoring electromechanical systems capable of assess health status, fault detection and generalization abilities.

Specifically, three main areas of research will be addressed in the investigation of this thesis as follows:

- Health diagnosis: the proposal of a generic data-driven monitoring methodology based on deep learning for fault diagnosis in complex electromechanical systems, considering signal preprocessing variables, a hyperparameter configuration process and the interpretation of the learning process.

- Anomaly detection: the development of a novel a methodological process for the detection of anomalies applied to industrial electromechanical systems, which aims to incorporate the advantages of automatically learned representation by deep neural network to improved performance.

- Incremental learning: the development of a framework to implement an incremental learning scheme, allowing the models to adapt and incorporate new operating conditions with the aim of produce generalizable schemes.

To successfully accomplish these contributions, the following specific objectives shall be achieved:

- The compilation and selection of databases that allow emulating the complexity of industrial systems to obtain relevant information from different electromechanical architectures that allow validating the proposed methodologies.

- Study and evaluation of novel algorithms in the field of artificial intelligence, especially those related to deep neural networks, adapting them to allow their correct application in the field of monitoring electromechanical systems.

- The proposal of a novel data-driven monitoring methodology based on deep learning for fault diagnosis in complex electromechanical systems, which could take into account a feature fusion structure from different sources of information.

- The proposal of new multi-objective approaches considering the functionalities of algorithms based on deep-learning for the detection of new failure nodes, commonly known as anomalies.

- The proposal and validation of a robust and reliable methodology to provide generalization capacity to diagnostic schemes, considering an incremental learning framework.

To allow the research, the perform of the contributions and the validation of the results of the proposed objectives, different electromechanical platforms are selected. Annex 1 provides a description of the different test benches used to evaluate the contributions of this Thesis work.

1.5 Chapters' description

In order to allow the development of the contributions and achieve the established objectives, this thesis project has been divided into different stages, which are reflected in the chapters described below.

A general literature review and summary of the emerging research works of deep learning on machine health monitoring is presented in **Chapter 2**. In this chapter, the different algorithms based on deep learning are analyzed considering their possible application in the field of monitoring, identifying the current state of the field of diagnosis of electromechanical systems and establishing the guidelines behind this research. As a result, this chapter concludes with an exposition of the fundamental aspects must be satisfying for condition-based monitoring schemes. These aspects are analyzed in depth in the following three chapters that cover the proposed objectives.

In **Chapter 3**, a study regarding of the fault diagnosis is performed. A novel methodological process based on the extraction of features from different information sources is presented, leading to the development of a scheme aimed at the implementation of an algorithm based on deep-learning for a diagnostic model for electromechanical systems.

In **Chapter 4**, a scheme for anomaly detection is performed. The capabilities of deep-learning methods to address the anomaly detection issue are investigated, combined with techniques based on one-class classification.

In **Chapter 5**, strategies for an incremental learning framework are addressed. A methodology is proposed to address incremental learning in monitoring schemes through transfer learning strategies. Domain adaptation, as a subfield of transfer learning, is studied and evaluated to improve learning and generalization capabilities.

In **Chapter 6**, the thesis work is analyzed from a global point of view, and the general conclusions and contributions are exposed.

Finally, in **Chapter 7** presents a summary of the publications resulting from the development of the research work, including collaboration.

2.

Deep Learning and its applications to condition monitoring - Literature review

The main purpose of this chapter is to review and summarize the emerging research work of deep learning in the field of machine health monitoring. After the brief introduction of deep learning methods, deep learning applications in rotating machine monitoring systems are reviewed, with special attention to the interpretability of these schemes, to define the state of the art in the research field of thesis.

CONTENTS:

2. Deep Learning and its applications to condition monitoring - Literature review

2.1 Data-driven scheme applied to CBM

Advances in sensors, the rise of advanced computer systems and the integration between information technology and operational technology, have been revolutionizing manufacturing in the industry. In the smart manufacturing environment, a key component is the machine condition monitoring. Compared with traditional model-based and signal-based monitoring systems, data-driven machine condition monitoring systems offer a new paradigm due to the high performance that has been obtained with the application of techniques based on artificial intelligence (AI). On the one hand, industrial manufacturing environments are subject to variable working conditions, complex environments and the presence of constant noise, which makes it difficult to build models that allow a diagnosis or prognosis of the condition [25]. On the other hand, data-driven monitoring systems offer the flexibility to adapt to different sources of information, in addition to allowing diagnostic models to be updated with new measured data [26]. In this regard, techniques based on machine learning (ML) - a field of research dedicated to building and interpreting methods that "learn" - are considered a powerful tool that allows the characterization of patterns from data [27]. Deep learning (DL), as a subfield of machine learning in AI, which consists of methods based on artificial neural networks, has attracted the attention of academics and industry researchers. As a result, numerous intelligent methodologies have been developed to address practical problems in industrial scenarios and have also brought successful advances in the field of machine health monitoring, calling these schemes intelligent fault diagnosis (IFD) [28-31].

Deep learning attempts to model hierarchical representations from data and characterize patterns by stacking multiple layers of information processing modules in hierarchical architectures. Recently, deep learning has been successfully applied in various research fields, such as image recognition, computer vision, natural language processing, automatic vehicle driving and clinical diagnostics [31-36]. The use of DL has spread in various areas and has become a promising tool in many applications of manufacturing industry due to the following aspects:

- Advantages in massive data processing. Data processing is an action to collected and manipulated data, then convert it into meaningful information for decision-making. A general definition of data processing as follows:

  > *Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of statistics which apply to analyzing data* [Judd, Charles; McCleland, Gary (1989). Data Analysis, 37].

Due to great advances in sensing and measuring, large-scale data have inevitably emerged. In this context, the need to manage large amounts of data is a key concern by many organizations, especially for industry. The rapid development of AI-technologies in combination with the advent of graphics processor unit (GPU), which has been reduced the required running time and have allowed the widespread use of DL algorithms, characterized by their great potential to extract hierarchical representations from data, it will have become the answer for managing large-scale data

- Effective pattern recognition and discriminative feature learning: A conceptual definition of pattern recognition is:

  > *The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories* [38].

  Pattern recognition is a problem that exists mainly in data processing applications, among the most common tasks are classification and regression. In this regard, DL-algorithms have shown great capabilities for the automated recognition of patterns and regularities in data. Multi-layer architectures enable DL methods to learn hierarchical features. The learnt features capture both local and inter-relationships for the data as a whole, learned features represent not only global characteristics but also particular patterns, such as characteristic fault patterns.

- Creating mapping relationships between health conditions by intelligent models. Finding relationships between information in multiple systems and defining those relationships is an existing challenge in the field of data science. Thus, data mapping helps to organize the data in such a way that it allows to recognize characteristics that help the execution of various tasks, such as the identification of faults. In the field of machine health monitoring, the characterization of fault patterns from physical magnitudes has been a widely studied topic. However, extracting features manually is hard work. In this regard, DL-based methods provide functionalities such as clustering, in which an automatic restructuring is carried out that allows grouping the information regarding data similarities.

- Industrial data management in an end-to-end way. DL-based diagnostic systems offer the advantage of learning representations automatically from data in an end-to-end way and be able to process accordingly. That is, all model parameters, including the feature learning module and the pattern classification module, can be trained together. In contrast, conventional ML-based condition monitoring systems each of the modules must be optimized separately.

Figure 2.1.1(a) illustrates the concept of multi-modules present in conventional ML-based condition monitoring systems. Conventional Condition Based Maintenance (CBM) systems usually consists of the following key modules: feature extraction, feature selection/reduction and model training. These models are generally assigned the tasks of detection and diagnosis. Additionally, when a new fault or a new operating condition is detected, this new information is sought to be incorporated into the diagnostic scheme. This need to incorporate new information into monitoring schemes implies considering the generation of incremental learning strategies. However, the main disadvantage of the conventional approaches is that the entire diagnostic system cannot be jointly optimized. Besides, the feature mapping capability in the feature extraction and reduction modules is limited by the linear mathematical basis. In addition, the conventional incremental learning strategies only contemplate retraining the models to incorporate new information, discarding the existing models.

In comparison, DL-based CBM systems have shown a high ability to learn complex relationships with the use of neural networks. A DL-based diagnostic framework is also shown in **Fig. 2.1.1**(b). It can be seen that the three characteristic modules of conventional diagnostic frameworks are contained under a single module. This means that the extraction, the mapping and the identification or classification are carried out in a single stage through end-to-end training, which it involves all model parameters are jointly optimized. Also, DL-based methods have the ability to model hierarchical representations from data due to the implementation of multiple stacked layers that have the capability to process information using hierarchical architectures. This modeling is often done using transformations involving non-linear transfer functions, which allows them to be well suited to problems involving dynamic systems, such as rotating industrial machinery.

In summary, there are three fundamental aspects in which DL methods can provide functionality within CBM systems. The first is related to the ability to extract and characterize the signals in conjunction with the performance of health diagnosis. Conventional schemes use specific tools to carry out each of these stages. With the application of DL methods, the aim is to improve diagnostic performance, which can be achieved by mapping the signals in conjunction with the diagnostic stage (fault classification). The second is about the capacity to detect anomalies, these can refer to the presence of new faults or changes in the operating condition.

Due to the anomalous behavior is not previously characterized, the schemes oriented to this type of tasks are unsupervised, that is, there are no labels in the training stage. This is a complex challenge, therefore, algorithms that can face this task are sought, and adapt them to the field of CBM. The third is the ability to transfer knowledge from a previously trained model. This functionality allows pre-existing models to be reused and adapted to incorporate new operating conditions or incorporate new faults to the diagnostic stage. Unlike conventional schemes, in which it is not possible to reuse previous models and therefore new training must be generated to incorporate knowledge, in DL-based schemes it is possible to adjust the weights of the network by reusing previously trained models by performing a fine-tuning. Taking this into account, in this thesis work an exploration of the different DL methods is carried out to take advantage of their functionalities for application in the field of monitoring electromechanical systems.

Deep learning algorithms have several variants such as auto-encoders [39], convolutional neural networks [40], deep belief network [41] and recurrent neural networks [42]. In recent years, several investigations have emerged around the application of machine condition monitoring using these deep learning methods. The current chapter of this research work attempts to provide a broad overview of these latest investigations of DL-based monitoring schemes.

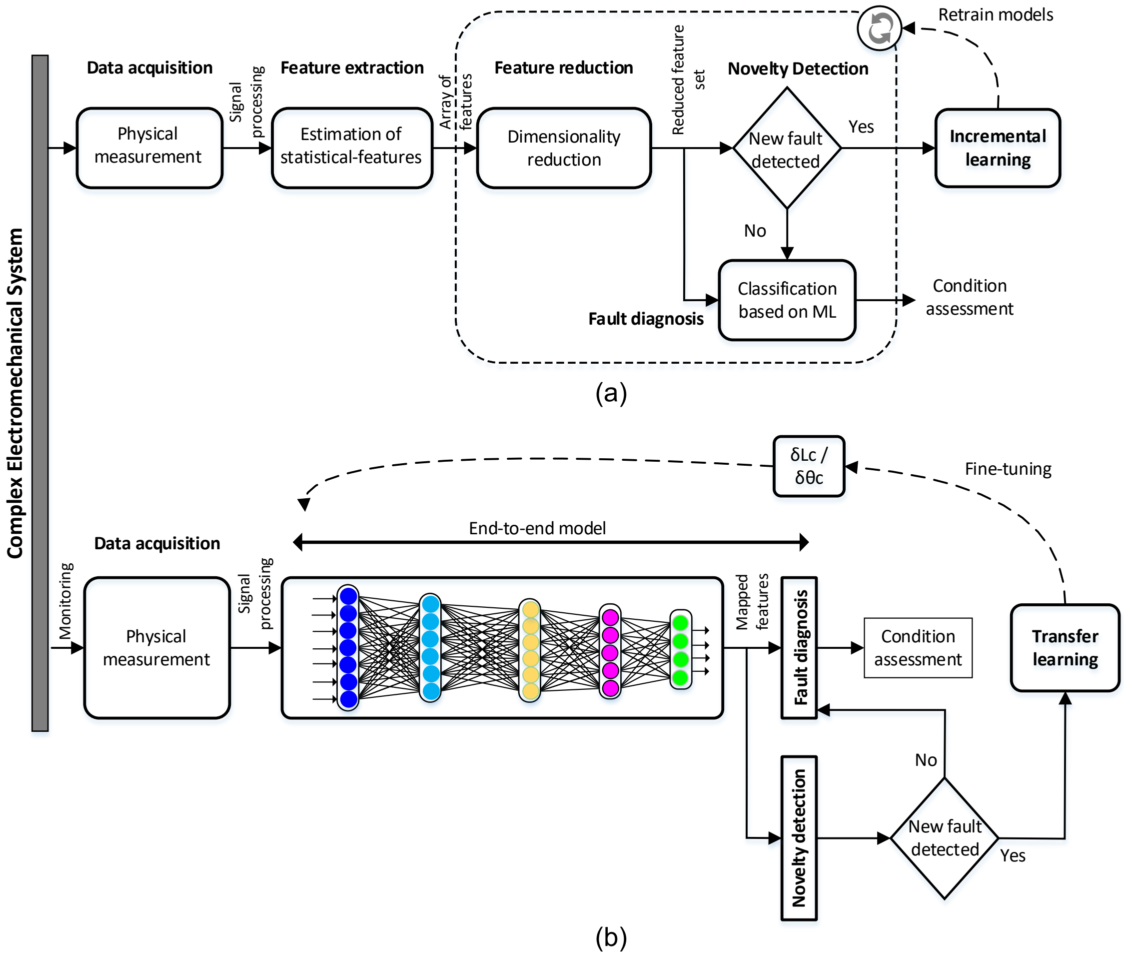


Fig. 2.1.1 Frames showing the two different monitoring systems; (a) conventional ML-based condition monitoring and (b) DL-based condition monitoring.

2.2 Deep Learning based algorithms

As a branch of machine learning, deep learning arises from artificial neural networks which is characterized by performing multiple non-linear processing layers and tries to learn hierarchical representations from data. As a simple description, it can be defined as an application based on artificial neural networks (ANN) for learning tasks that contain more than one hidden layer. The objective is to model high-level patterns in the data to determine features with a high-level of meaning that can be used under supervised, semi-supervised or unsupervised learning. Basically, an ANN with more than two hidden layers, without taking into account input and output, can be classified as a deep architecture. However, DL is not just about of the number of layers, but rather of a series of procedures and construction ideas to properly perform feature learning in an automated way [43-46].

At first deep architectures were created by stacking multiple layers using backpropagation (BP) as the training algorithm. However, BP presents difficulties in training ANN, especially in deeper architectures. The adjustment of weights and the large number of calculations that require a high computational cost are some issues that arise with the use of BP. In addition to suffering from a problem of vanishing gradients [47]. ANN architectures were typically initialized using random numbers and the gradient of the network weights with respect to network error was used. The rapid growth in computing power through GPUs and the advancement in AI technologies have enabled the development of more sophisticated DL methods. There are currently several deep learning methods, and they have been a constantly growing research topic. Among the deep architectures that highlight are auto-encoders, convolutional neural networks, deep belief networks and recurrent neural networks. These methods have been implemented in applications related to computer vision, image processing, speech recognition, natural language processing [48-51]. Taking into account the high potential of DL technology, considering its application in CBM in electromechanical industrial systems has been inevitable. Therefore, in the following subsections of this research work, a detailed explanation of the current state-of-the-art deep learning methods will be made, which will serve as the starting point for this thesis.

2.2.1 Auto-encoder

Auto-encoder (AE) is a type of symmetrical neural network that is trained in a semi-supervised manner. AE is designed to mapping a low dimensional representation from the input data. The AE learning procedure consists in two steps: encoder and decoder. The step encoder takes the x vector of length k containing the set of inputs signals and transforms it to a hidden layer representation h consisting of p sparse-activated neurons via a non-linear mapping as follows:

$$h = f(W_e x + b_{e.}), \tag{1}$$

where f is a non-linear activation function, W_e and $b_{e.}$ are the weights and biases matrices, respectively. The commonly used activation functions include softmax, relu, tanh, sigmoid and others [52-54]. In this regard, the sigmoid function is the most commonly used function.

$$f(z) = 1/(1 + e^{-z}), \tag{2}$$

Then, the encoded hidden layer was transformed to obtain the output representation of the auto-encoder through the decoder transformation in a way as follows:

$$y = f(W_d h + b_d), \tag{3}$$

where W_d and b_d are the weights and biases matrices of the decoder process, respectively, and y is the output of the AE, which has the same dimension as the input x. A typical structure of a single-layer auto-encoder is show in **Fig. 2.2.1**. It shows the main parts: input layer, hidden layer and output layer. The encoder process involves the input layer and the hidden layer (blue box), while the decoder process involves the hidden layer and the output layer (orange box).

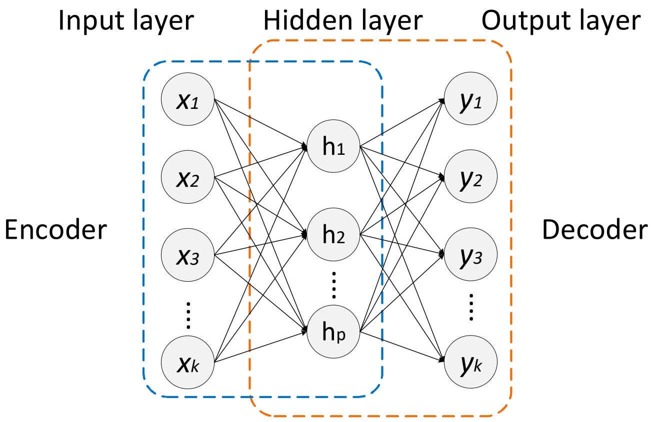


Fig. 2.2.1 Schematic structure of an auto-encoder

The auto-encoder training process consists of encoding the input to a compressed dimensional representation, then, from that compressed representation, replicating the input in the output. This process is based on the optimization of $\theta = [\boldsymbol{W}_e, \boldsymbol{b}_e, \boldsymbol{W}_d, \boldsymbol{b}_d]$ to reduce the reconstruction error between the input x and the output y by measuring of a cost function. This can be achieved by one commonly adopted measure for the average reconstruction error over a collection of N data samples: the mean squared error, and can be written as follows:

$$\Omega_{Mse} = \frac{1}{N}\sum_{k=1}^{N}(x_k - y_k)^2, \tag{4}$$

By adding a regularization term on the weights to the cost function, it can prevent an overfitting in the network. A weight-decay regularization term is added to prevent large weights from appearing. This term is denominate the L2 regularization term and is defined by:

$$\Omega_{weights} = \frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} w_i(j,k)^2, \tag{5}$$

where a is the number of weight parameters, b is the number of rows and c is the number of columns in each weight matrix.

2.2.1.1 Sparse Auto-Encoder

In order to generate more specialized network models that can learn the structure information of the input data more effectively, a sparsity regularizer is introduced. This regularizer is a function of the average output activation value of a neuron. The average output activation measure of a neuron i is defined as:

$$\widehat{\rho_i} = \frac{1}{n}\sum_{j=1}^{n} h_i(x_j) = \frac{1}{n}\sum_{j=1}^{n} f(w_i^T x_j + b_{i.}), \tag{6}$$

where n is the total number of training samples. x_j is the jth training sample, w_i^T is the ith row of the weight matrix $\boldsymbol{W}_e$, and $b_{i.}$ is the ith entry of the bias vector, $\boldsymbol{b_e}$. Therefore, a sparsity function aims to constrain the values of $\widehat{\rho_i}$ to be low, causing the auto-encoder to learn a representation, such that each neuron in the hidden layer fires to a small number of samples. This means that, each neuron function as specialized feature detectors by responding to some feature that is only present in a small subset of the training samples. So that neurons can into becoming sparse, is add a term to the cost function that restricts the values of $\widehat{\rho_i}$. One such sparsity regularization term is the Kullback-Leibler divergence, which is a standard measure of how a probability distribution diverges from the expected distribution, where ρ is the desired sparsity parameter and $\widehat{\rho_i}$ is the effective sparsity of a given hidden neuron. This term is defined by:

$$\Omega_{sparsity} = \sum_{i=1}^{n} KL(\rho||\widehat{\rho_i}) = \sum_{i=1}^{n} \rho \log\left(\frac{\rho}{\widehat{\rho_i}}\right) + (1-\rho)\log\left(\frac{1-\rho}{1-\widehat{\rho_i}}\right), \tag{7}$$

Finally, the cost function $FunCost$, for training a sparse auto-encoder to be minimized is defined as the sum of the error term plus the regularization penalty terms, so the parameter-tuning problem can be stated as an optimization problem where the networks parameters are adjusted in order to minimize the resulting cost function.

$$FunCost = \Omega_{Mse} + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity} \tag{8}$$

where λ is the coefficient for the $L2$ regularization that controls the weight decay and β is the coefficient for the sparsity regularization term. The cost function consists of three parts: reconstruction error, weight decay term and sparse restriction term.

2.2.1.2 Stacked Auto-encoders

One of the current problems that paralyzes the implementation of deep models based on ANN is the difficulty of training models with several layers at the same time. One of the solutions seems to be to apply different shallow networks and later stack them in a more robust model. Deep-Neural-Network (DNN) are composed of several hidden layers, and each hidden layer can perform a nonlinear transformation from the previous layer to the next one. A DNN can be built by stacking several AEs and performing a layer-by-layer training. Various training methods have been proposed to stack AEs, of which the one proposed by Hinton stands out [55]. Besides, since the auto-encoders can be trained in an unsupervised way, being that it auto-fit by replicating its input on the output, provide a solution for implementing DNN to train the model. **Fig. 2.2.2** illustrates the architecture of a stacked auto-encoder. The architecture is made up of two AEs and results in five layers with four stages: two encoders, followed by two decoders.

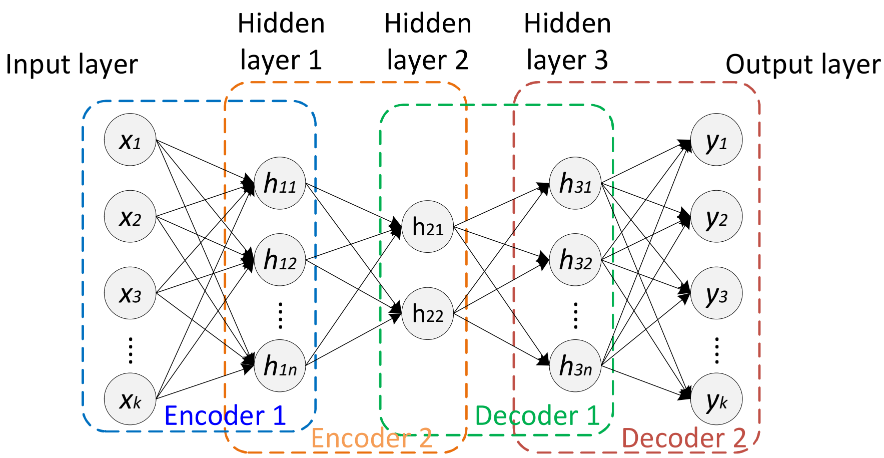


Fig. 2.2.2 Schematic structure of the construction of a deep-autoencoder.

The training protocol based on stacked auto-encoders (SAE) can make DNN models more effective in characterizing complex patterns from data. In addition, through hierarchical architectures such as DNNs, it is possible to learn representations with different levels of abstraction, starting from simple to more complex characteristics. Therefore, in the encoding stage, the input data are constantly compacted by several hidden layers to a more compressed representation, as follows:

$$\boldsymbol{E_1} = f(\boldsymbol{W_{e1}x_k} + b_{e1}) \tag{9}$$

$$\boldsymbol{E_2} = f(\boldsymbol{W_{e2}E_1} + b_{e2}) \tag{10}$$

$$\vdots$$

$$\boldsymbol{E_n} = f(\boldsymbol{W_{en}E_{n-1}} + b_{en}) \tag{11}$$

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

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

$$\boldsymbol{D_1} = f(\boldsymbol{W_{d1}E_n} + b_{d1}) \tag{12}$$

$$\boldsymbol{D_2} = f(\boldsymbol{W_{d2}D_1} + b_{d2}) \tag{13}$$

$$\vdots$$

$$\boldsymbol{Y_k} = f(\boldsymbol{W_{dn}D_{n-1}} + b_{dn}) \tag{14}$$

where D_j represents the corresponding the decoding vector for each of the encoding layers. W_{di} and b_{di}, are weight parameters and the corresponding biases of each decoding layer respectively. Y_k is the final output vector of the DNN, which is also the reconstruction vector of x_k.

As a result, the deep auto-encoder is able to reconstruct the input representation at the output of the network, and during this process the hidden layer network learns various representations of the data, as well as multiple relationships between the input signals. Thus, the SAE can be adapted for classification or regression applications by extending and applying an additional supervised stage, focusing the low-dimensional mapping of the encoder stage to a specific target. To accomplish this, the SAE network adapts by discarding the decoder stages and adding a classification or regression stage to the last encoder each. In **Fig. 2.2.3** a SAE with a softmax layer configured for the classification task is shown. Considering the capabilities and potential of SAEs as DNN, makes them ideal candidates to explore for condition monitoring applications for rotating systems.

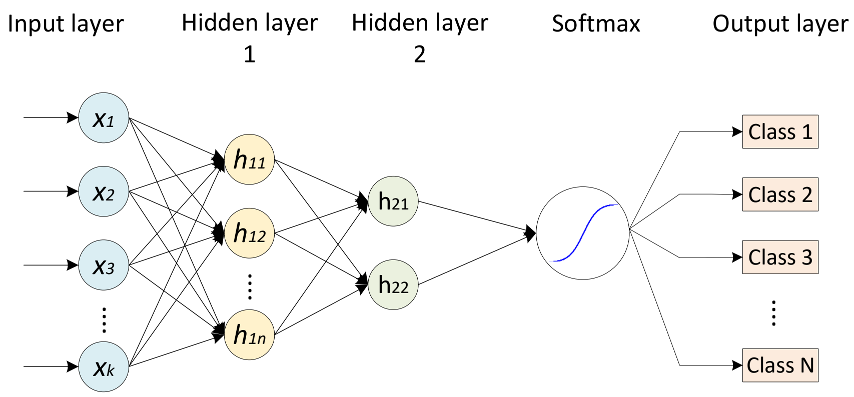


Fig. 2.2.3 Structure of an SAE with a softmax classifier.

2.2.2 Convolutional Neural Network

Convolutional neural networks (CNNs) are a type of artificial neural network proposed by analyze visual images. CNNs were initially proposed by LeCun [56] to process images. The field of applications include image and video recognition, image classification, image segmentation and medical image analysis. Recently CNN has also been introduced to address sequential data including natural language processing (NLP), speech recognition, financial time series, sensor measurements, and others.

CNN has the functionality to learn abstract features by toggling and stacking convolutional layers and layer pooling. Generally, CNN is featured by three kinds of layers: convolution layers, pooling layers, and full-connected layers. Convolutional layers use local filters to produce specific features from the input data, while pooling layers extract fixed-size features with sliding windows following average or maximum pooling rules.

2.2.2.1 Convolution

While shallow neural networks establish fully connected connections, CNN implement the concept of local connection features, in which only a limited part of the input data is connected to each node in a convolution layer. The dot product defines the convolution operation that can be presented as follows:

$$a(y) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} w_{i,j,k}^{y} \times x_{i,j,k} + b \qquad y=1, 2, \ldots, q, \tag{15}$$

where y is i^{th} convolutional kernel; $a(y)$ is feature map; x is the input signal, b is the bias; i, j, k represent the dimension of the input signal. As mentioned, the use of CNN for sequential data has recently become widespread, resulting in the adaptation of CNN to a one-dimensional format. So for a one-dimensional signal, (15) simplifies as follows:

$$a(y) = \sum_{i=1}^{m} w_{i}^{y} \times x_{i} + b \qquad y=1, 2, \ldots, q, \tag{16}$$

Following the convolution operation, an activation function adds $A(\cdot)$ giving nonlinearities to the convolutional layers. Rectified linear unit (Relu), sigmoid, linear, Gaussian and hyperbolic tangent (tanh), are the most common activation functions for this stage.

2.2.2.2 Pooling

After the convolution process, a pooling layer is followed. Pooling performs down sampling operation, which aims at data dimension reduction as well as better feature extraction. The most common the pooling operations are the max pooling function or the average pooling one. In the following, max pooling is explained in details.

$$p_{max}^{l(i,j)} = \max_{(j-1)\omega<t<j\omega} \{x^{l(i,t)}\} \qquad j=1, 2, \ldots, q, \tag{20}$$

where ω is the width of convolutional kernerl; $x^{l(i,t)}$ represents the neural node of the i^{th} feature representation in l^{th} layer; *j* is the number j^{th} pooling kernel.

2.2.2.3 Full-connected layers

Finally, after several convolution and pooling processes, fully connected layers are added to perform classification or regression to make predictions [57-58]. And the softmax function is generally adopted on the top classification layer [59], its provides the corresponding probability as part of the activation function values, where the $\lambda_j(z)$ represents the probability of j^{th} class. The calculation of $\lambda_j(z)$ is defined following:

$$\lambda_j(z) = \exp(z_j)/\sum \exp(z_j) \qquad j=1, 2, \ldots, C, \tag{21}$$

$$z_j = W_j M + b_j \tag{22}$$

The softmax function takes as input a vector z_j with C real numbers. To give a clear illustration, a framework corresponding to one-layer CNN has been displayed in **Fig. 2.2.4**.

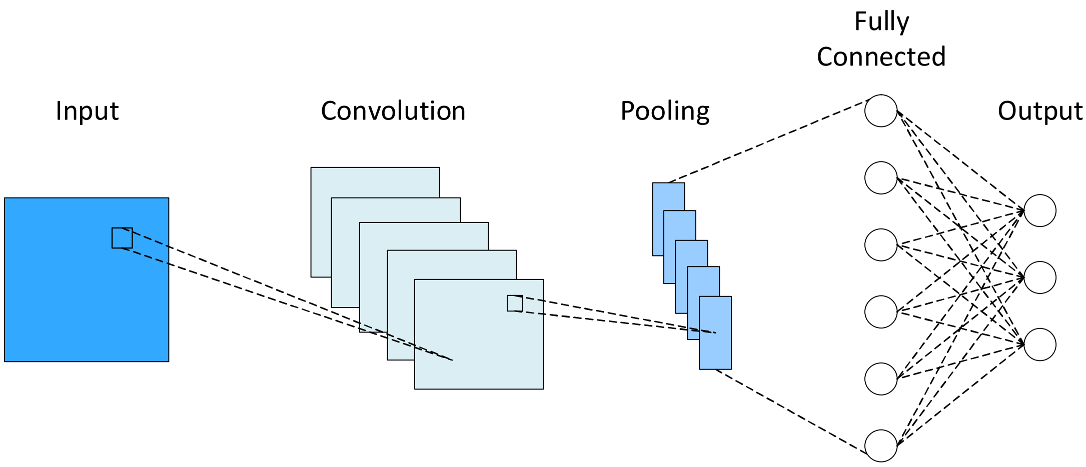


Fig. 2.2.4 Schematic diagram for one-layer convolutional neural network.

2.2.3 Restricted Boltzmann Machine

Restricted Boltzmann machine (RBM) is a two-layer neural network forming a generative graphical model that consists of two groups of units including visible units **v** and hidden units **h**. An RBM consists of a two-layer network of fully connected nodes with both forward and backwards connections in a kind of cycle. The restriction takes place by sharing the weights of the forward and backward connections.

Therefore, according to the parameters $\theta = [\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{a}]$, the energy function is defined as follows:

$$E(\boldsymbol{v}, \boldsymbol{h}; \theta) = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} v_i h_i - \sum_{i=1}^{I} b_i v_i - \sum_{j=1}^{J} a_j h_j \tag{23}$$

where w_{ij} is the weight of the connection between visible unit v_i, where the total number is I and the hidden unit h_i where the total number is J. b_i and a_j corresponds the bias for visible units and hidden units, respectively.

The joint distribution for all units is determined from the energy function $E(\boldsymbol{v},\boldsymbol{h};\theta)$ through the following function:

$$p(\boldsymbol{v},\boldsymbol{h};\theta) = \exp(-E(\boldsymbol{v},\boldsymbol{h};\theta))/\boldsymbol{Z} \quad (24)$$

where $\boldsymbol{Z} = \sum_{\boldsymbol{h},\boldsymbol{v}} exp(-E(\mathbf{v},\ \mathbf{h};\ \theta))$ corresponds to the normalization factor. The conditional probabilities for hidden and visible units **h** and **v** is defined by:

$$p(h_j = 1|v;\ \theta) = \delta\left(\sum_{i=1}^{I} w_{ij} v_i + a_j\right) \quad (25)$$

$$p(v_j = 1|v;\ \theta) = \delta\left(\sum_{j=1}^{J} w_{ij} h_j + b_i\right) \quad (26)$$

where δ corresponds to the logistic function $\delta(x) = 1/1 + exp(x)$.

2.2.3.1 Deep Belief Network

The deep belief network (DBN) is a class of deep neural network composed of several layers, also called hidden units, connected between each layer but not between units within each layer. DBN are considered an evolution of the RBM proposed by Geoffrey Hinton [60]. A DBN is composed of multiple RBM models, which performs a non-linear transformation on its input vectors and produces as output vectors, which will be the inputs of the next RBM model.

The DBN learning process aims to learn to probabilistically reconstruct its inputs. Its layers act as feature detectors, where the hidden layer of each sub-network serves as the visible layer for the next. Once the learning process has been completed, the DBN can be trained in a supervised manner to perform classification. After pre-training, the deep architecture parameters can be fine-tuned to target classification or regression tasks by adding a softmax layer as the top layer, as shown in **Fig. 2.2.5**.

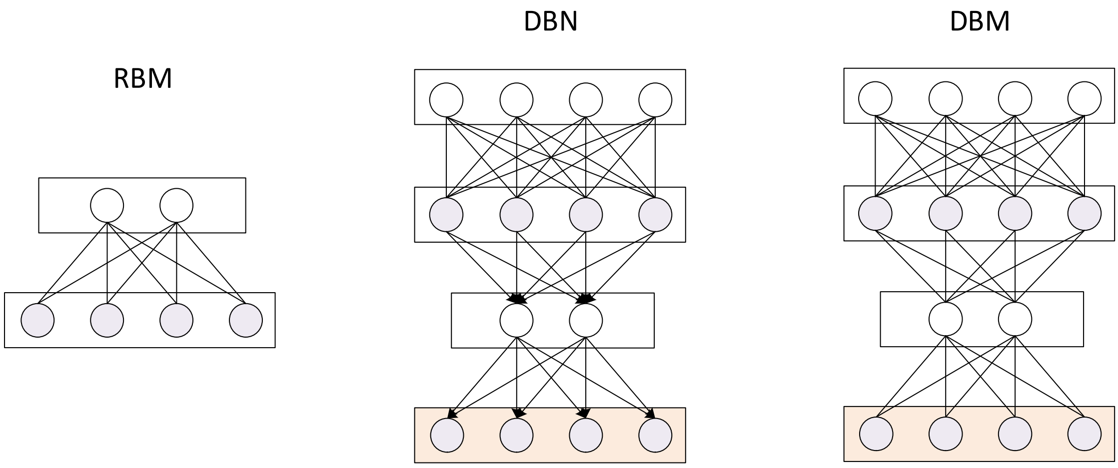


Fig. 2.2.5 Representation of the different architectures of RBM, DBN and DBM.

2.2.3.2 Deep Boltzmann Machine

Deep Boltzmann machine (DBM) is another type of deep neural network based on deep structure RBM [61]. In DBMs the hidden units are grouped hierarchically and not in a single layer. DBMs follow the established restrictions of RBMs, in which there is only connectivity between subsequent layers, maintaining the restriction of connections within layers and between non-neighboring layers. Unlike DBN, deep Boltzman machine is an undirected graphical model. While the DBNs are directed or mixed models, in addition to the fact that the training is layered. DBM models are trained in joint way, which makes them computationally expensive to implement.

2.2.4 Recurrent Neural Network

Recurrent neural networks (RNNs) are kind of artificial neural networks, considered the deepest networks because they can generate and store memory of sequences of input features [62]. This temporal dynamic behavior allows them to process variable length input streams. This makes them applicable for tasks with continuous or segmented time signals, such as speech recognition. The main difference from traditional neural networks is that the RNN makes connections from the entire history of previous inputs to the target values, allowing a memory of previous inputs to be generated in an internal state of the network.

The RNN training process can include backpropagation over time focused on supervised tasks with sequential input data. Through their internal memory, the RNNs address the sequential data. The transition function for a defined time step t starts by taking the current time information $\boldsymbol{x}_t$ and the previous hidden output $\boldsymbol{h}_{t-1}$, updating the new hidden output as follows:

$$\boldsymbol{h}_t = H(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}) \tag{27}$$

where H defines a transformation function, Usually a hyperbolic tangent function is used (tanh). $\boldsymbol{h}_t$ corresponds to the hidden output at the last time step which is the learned representation of the sequential data after processing the entire input sequence, whose length is T. This memory block can be represented by the cell structure of **Fig. 2.2.6**.

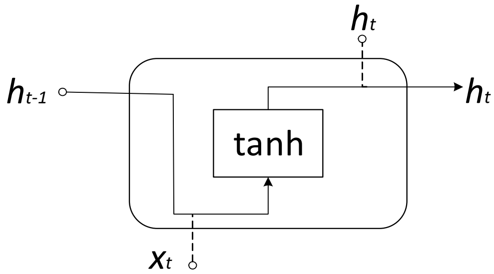


Fig. 2.2.6 Cell structure of a simple recurrent neural network.

For an RNN to learn long range temporal dependencies from the dataset, more complex structures are used. An example of these RNNs are Long-short term memory (LSTM) and gated recurrent units (GRU). The LSTM cell consists of memory blocks that allow write, read, learn or forget information through gates that open and close, with the corresponding weights, as shown in **Fig. 2.2.7**. The LSTM characterization process follows several steps from when the information enters the cell until the output is generated. This process is explained as follows:

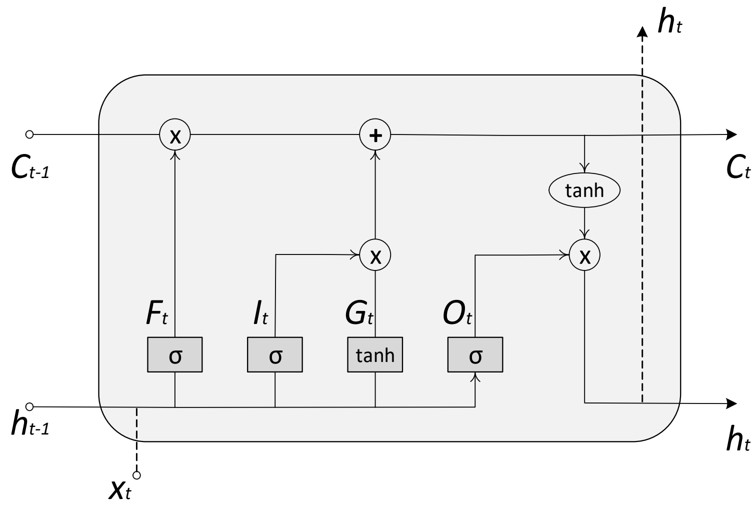


Fig. 2.2.7 Cell structure of long short term memory.

1. Input and recurrent values pass through the forget gate, here an activation function, usually a sigmoid function, determines the output value, as shown in **Fig. 2.2.8**. The output value is between a range of 0 and 1, when the value is zero, the cell forgets the data. The transfer function is defined by:

$$F_t = \sigma(W_f x_t + R_f h_{t-1} + b_f) \qquad (28)$$

where W are the weights of the respective gate, R the recurrent weights and b the bias values.

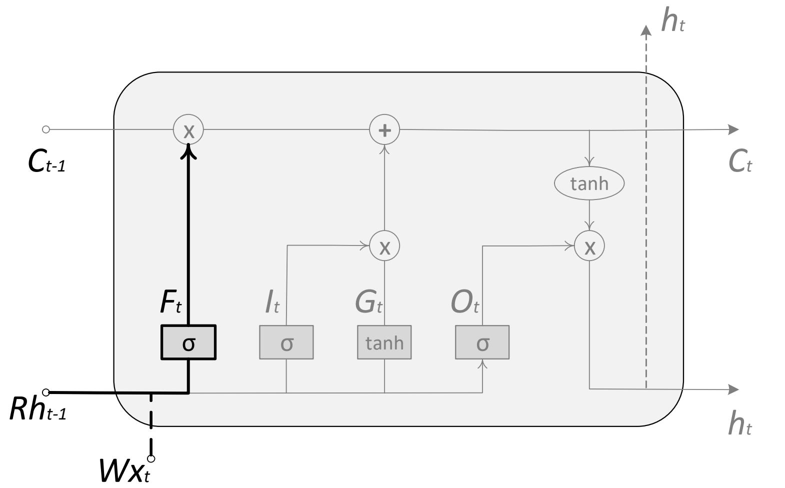


Fig. 2.2.8 Step of the forget operation in the LSTM cell.

2. The next step is update. The data is sent to the input gate where the activation function creates new values to add, as shown in **Fig. 2.2.9**. Both values are multiplied and update the new state. The activation functions are defined as:

$$I_t = \sigma(W_i x_t + R_i h_{t-1} + b_i) \tag{29}$$

$$G_i = tanh(W_g x_t + R_g h_{t-1} + b_g) \tag{30}$$

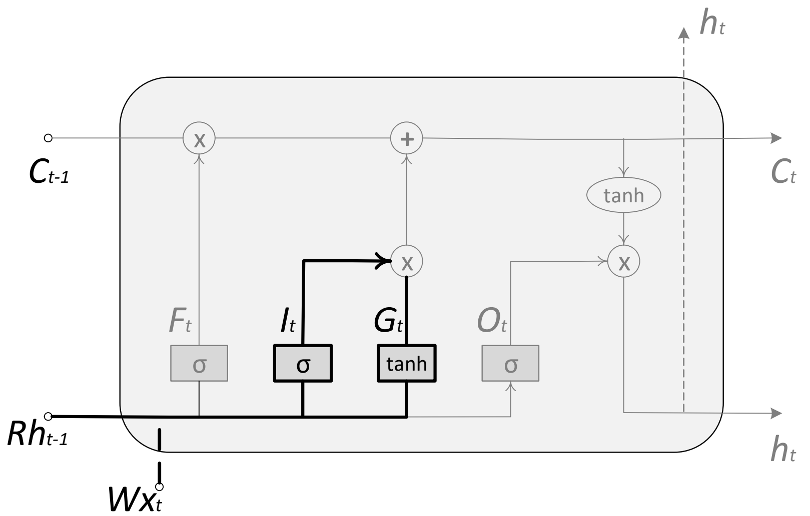


Fig. 2.2.9 Input gate operation and update in LSTM cell.

3. The old cell is then updated to a new value. The multiplication operation is carried out between the old state and the forger gate, adding the new value from the previous step, as shown in **Fig. 2.2.10**. The transfer function is defined as:

$$c_t = F_t \cdot c_{t-1} + I_t \cdot G_t \tag{31}$$

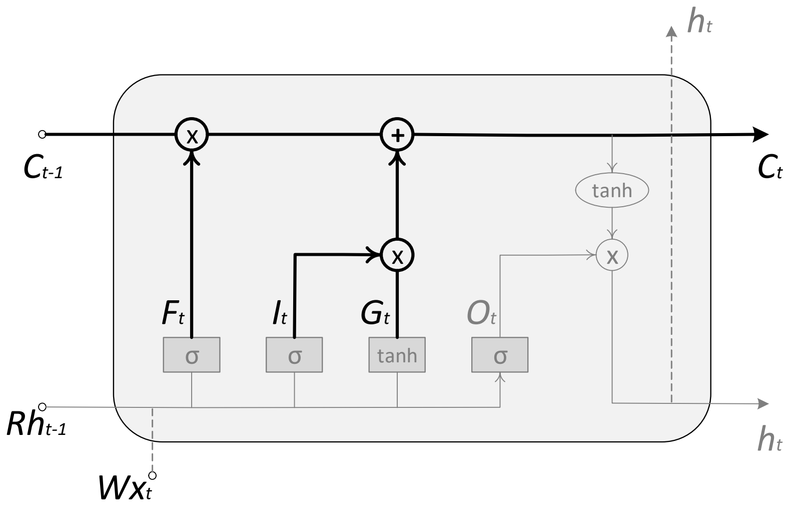


Fig. 2.2.10 Multiplication gate operation and update in LSTM cell.

4. Finally, the output is obtained. In this state, the input and recurrent values pass through the sigmoid activation function, the current state passes through a tanh function. The output is defined by the multiplication of the two values, this step is shown in **Fig. 2.2.11**. The transfer function is:

$$O_t = \sigma(W_o x_t + R_o h_{t-1} + b_o)\, F_t \cdot c_{t-1} + I_t \cdot G_t \tag{32}$$

$$h_t = O_t \cdot tan(c_t) \tag{33}$$

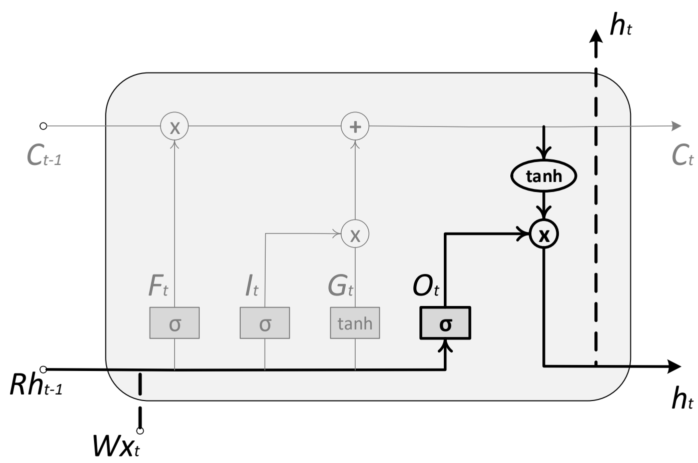


Fig. 2.2.11 Output gate operation in LSTM cell.

With the updated hidden states, the predicted output is defined as:

$$y_t = W_{fc} h_t + b_{fc} \tag{34}$$

where W_{fc} and b_{fc} are the weights and bias of the fully-connected layer, respectively

2.3 Applications of Deep Learning in Condition Monitoring

Artificial Intelligence-based tools have been widely used in the field of machine health monitoring in recent years. In this regard, the multifunctionalities offered by techniques based on deep-learning in various research fields have caused them to be considered for the application of machine monitoring systems. The layer-by-layer training process of DNNs based on autoencoders and RBM facilitates their implementation since they reduce the computational cost of traditional NN models, in addition to preventing the presence of the vanishing gradient problem [63] and improving the pattern recognition capabilities [64]. Convolutional neural networks and RNN allow to lead information both in images and temporal data, with structured learning mechanisms that allow to adequately characterize highly complex environments. These techniques allow to adapt both to tasks of classification, detection and prediction, functionalities necessary in the field monitoring in industrial systems for fault diagnosis, anomaly detection and calculation of the remaining useful life (RUL). In this regard, several publications related to DL-based in the field monitoring have emerged in recent years. In the following, some of these investigations related to the study of the condition in electromechanical systems based on deep-learning are presented, about the four aforementioned methods: SAE, CNN, RBM and RNN, highlighting its corresponding functionalities in its application in condition monitoring.

2.3.1 Auto-encoders for machine health monitoring

Autoencoder-based diagnostic schemes, especially on stacked autoencoder (SAE) architectures, can learn complex representations from data acquired from machinery, such as vibration, stator current, or sound. On the one hand, the proposed investigations have emerged around the functionalities offered by SAEs in the field of condition monitoring in electromechanical systems. Xia *et al* [65] proposed a monitoring scheme applying the denoising autoencoder to learn representative features. The proposed work uses the automatic feature extraction capabilities of autoencoders to characterize the different fault states of bearings in a motor-driven system. Similarly, Haidong *et al* [66], propose a feature learning method, in which, by a deep-autoencoder, extracts feature in a hierarchical way to learn the working conditions of a rotating machinery. Qu *et al* [67], proposes a new method for diagnosing gearbox faults. It proposes the use of a dictionary learning using a deep sparse autoencoder to learn features from vibration data. The goal of dictionary learning is to learn a basis for representation of the original input data. In addition to being a tool for feature extraction, AE's are used for their ability to learn and econstruct input signals. In this sense, Principi *et al* [68], proposes an unsupervised detection method based on the reconstruction error to establish whether or not there is a fault in an electric motor. Another method focused on using a reconstruction model is the one proposed by Memarzadeh *et al* [69], in this scheme, it is proposed to mapping the features optimizing by the reconstruction error. Thus, a low error indicates that the system state is normal, while a high reconstruction error indicates that it is abnormal.

A present problem in the field of diagnosis of complex electromechanical systems is the one referring to the high computational cost to process the high dimension of the physical magnitudes acquired for monitoring. In this context, Shao *et al* [70], propose a method applied to analyze experimental signals collected from bearings in an electromechanical system, using an auto-encoder to compress data focused on obtaining a low-dimensional representation.

On the other hand, some research has focused on addressing different issues around the application of autoencoders models in the field of machine monitoring. Therefore, aspects such as the configuration of hyperparameters, the type of input information, and strategies to prevent overfitting are some of the problems that have been addressed in recent years. A detailed empirical study by stacked denoising-AE with three hidden layers for fault diagnosis for rolling bearing fault diagnosis is presented by Lu *et al*. [71]. This work allowed to compare the advantages of a DNN compared to traditional diagnostic algorithms. Highlighting the ability to extract salient features adaptively to effectively identify health status in bearings. Sun *et al*. [72] performed a diagnosis scheme implementing a compression technique to extract low-dimensional features from raw vibration signals to feed a SAE-based deep neural network (DNN) model. Specifically, in their experiments they were able to highlight the importance of network hyperparameters to achieve better diagnostic performance. Due to the variety in the information that can be used as input to the DNN, a study based on characteristic frequency-related intrinsic mode function components and the raw signal proposed by Dai *et al* [73], has made it possible to demonstrate that DNN methods such as SAEs can be fed by different sources of information. The study uses a stacked sparse denoising auto-encoder which combines sparse auto-encoder (SAE) and denoising auto-encoder (DAE) as a learning method. Faced with the challenge of working with limited size of data, Wenjun Sun et al [74] proposed an SAE-based approach focused on preventing overfitting. It implements a regularization method called ‘‘dropout”, in order to overcome the deficiency of overfitting.

Accordingly, due to the functionalities offered by SAEs and some of the issues present in their implementation in the CBM field, new methodologies may arise to deal with existing problems, both for the diagnosis of the condition and the application of SAE-based algorithms.

2.3.2 CNN for machine health monitoring

In deep-learning, convolutional neural networks (CNNs) are a type of ANN-based method most focused on image analysis and processing. In the field of health monitoring in electromechanical systems, CNNs have been applied for their ability to learn features by their main processes: convolution and pooling. In this sense, numerous efforts have been carried out to adapt CNN algorithms to diagnostic methodologies for rotating machinery conditions recognition. In addition, strategies have recently been emerging to implement CNN with temporal signals, which enables them as a tool for the study of signals acquired from monitoring. Lu *et al* [75] carry out an investigation based on convolutional neural network for diagnosis bearing health using vibration. In this work they explore the advantages of image recognition and visual perception to fault diagnosis. They performed an image transformation method to map the original monitoring information to a series of feature maps as illustrated in the **Fig. 2.3.1**. The acquired monitoring data with time scale and amplitude scale normalized are transformed into an image matrix in order to establish them as input for CNN.

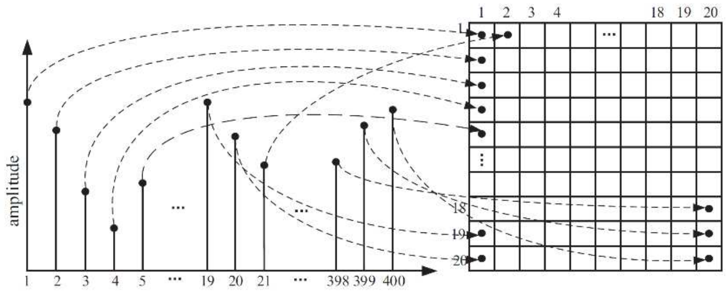


Fig. 2.3.1 Schematic diagram of image transformation from the original vibration signal.

Another proposed research that makes use of feature maps is carried out by Weifang Sun *et al* [76]. A method for fault recognition for gear using a dual-tree complex wavelet transform (DTCWT) is proposed. The wavelet sub-bands of DTCWT are displayed in columns to generate a two-dimensional image matrix that serves as input to the CNN model. The use of temporal signal to image conversion methods has also been studied by Long Wen *et al* [77]. It is proposed to process the raw vibration signals and transform them into a 2-D representation of features. The result is a gray pixel image as illustrated in **Fig. 2.3.2**, which is capable of being processed as input by the LeNet-5 CNN model proposed in this research.

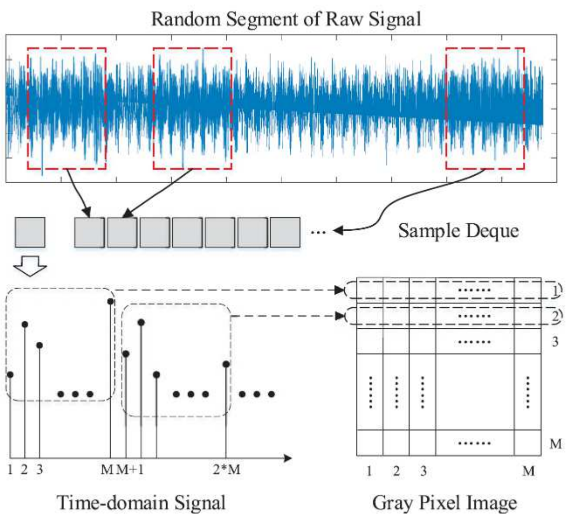


Fig. 2.3.2 Time-domain raw signals conversion into images.

Although these investigations have been remarkably successful in implementing and adapting CNN-based deep-learning models to machine monitoring schemes, the interpretation of signal-to-image conversion is unclear. A study regarding the analysis of time-frequency representations to be used as input to CNN models is performed by Verstraete *et al* [78]. Two linear time-frequency transformations and one nonlinear nonparametric time-frequency transformation are examined to be used as image input: short-time Fourier transform spectrogram, wavelet transform and Hilbert-Huang transformation, respectively. The proposal was validated for the fault diagnosis of rolling element bearings. In this study, it is concluded that conventional schemes based on manual process of feature extraction and methods of feature selection can be substituted with a deep learning frameworks, since the results were more robust and more effective.

As is known, the quality of the extracted features is essential for diagnostic schemes can perform pattern recognition more efficiently and make a correct analysis of the condition of the machinery. In this regard, Shaobo Li *et al* [79] proposed a study based on extracting RMS values from the Fast Fourier Transformation (FFT) features of the vibration signals to generate 2-dimensional maps that serve as input to train a CNN model. A notable contribution is the fusion of data from two sensors as inputs. This functionality allows a more robust study to be carried out since it is possible to integrate various physical magnitudes into the diagnosis. Using multiple sensors to perform condition monitoring is a promising tool to deal with complicated damage detection problems of mechanical systems. In this sense, the work presented by Luyang Jing *et al* [80] proposed a multi-sensor data fusion method based on deep convolutional neural networks (DCNN) for fault diagnosis. In such work, the physical magnitudes of different sensors are used: accelerometer, microphone, current sensor and optical encoder; to perform the diagnosis of gear faults. Different layers of feature extraction using CNNs models are used to fuse the information coming from the different sensors. This approach suffers mainly from the selection of a suitable fusion level, an existing challenge in monitoring schemes, especially when robustness is sought when combining data from different information sources.

Although monitoring schemes that using 2-dimensional representations have prospered in recent years, some researchers have made an effort to perform time signal analysis. Due to most of the data is extracted through temporal sensor measurements, and because most of the information extracted through the various processing is presented in a 1-dimensional way, it is more efficient to adapt deep-learning models to use temporary data as input. Furthermore, the signal-to-image conversion processes and the subsequent analysis result in a higher computational cost, a drawback that is sought to be avoided. Ince *et al* [81] suggested a monitoring method for early fault-detection using 1-D convolutional neural networks that has a design to fuse the feature extraction and classification steps into a single scheme. The proposed approach is applied directly on raw data, thereby, eliminates the need for a separate feature extraction algorithm.

Jing *et al* [82] suggested a study for automatic feature extraction through a convolutional neural network. The study was based to perform a comparison for different data processing: raw time-domain data, frequency data, time-frequency data, time-domain features, frequency-domain features and wavelet-domain features. The investigation concluded that for the analysis of fault detection in gearbox the most adequate processing is to use frequency spectrum of vibration data. The study was further validated against two gearbox databases. It was also concluded that the automatic extraction of features by deep-learning methods offers a better performance, in comparison with classical manual feature extraction. Zhang *el at* [83] propounded a model based on deep-learning for fault diagnosis in bearings. The proposed method works in an end-to-end way by taking raw temporal signals as input. The model was tested under a noisy environment and working under different load conditions, obtaining high accuracy. The study also included a sensitivity analysis of strategies that help to better converge the CNN model used, such as batch normalization, kernel dropout and ensemble learning. Another investigation focused on the study of noisy environments is the proposed by Zhang *et al* [84]. The advantages of feature mapping through multilayer nonlinear are used to improve diagnostic accuracy. In addition, with the use of wide kernels in the first convolutional layer it is possible to suppress high frequency noise, commonly present in most industrial environments. Another significant contribution is the use of time signals in a 1-dimensional format. This allows the use of raw vibration signals as input, avoiding the hard work of manual feature extraction. Sun Wenjun *et al* [85] developed a convolutional discriminative feature learning method for induction motor fault diagnosis. The method utilizes back-propagation based neural network to learn local features that capturing discriminative information to perform a diagnosis. Experiments performed on a machine fault simulator that include different elements such as: an induction motor, bearing, bevel gearbox, belt, shaft, load disc. This implies that the proposed approach is useful for the diagnosis of different faults. The fault conditions studied include: stator winding defect, unbalanced rotor, defective bearing, broken bar and bowed rotor. All these research have allowed the advancement of the implementation of algorithms based on artificial intelligence in machinery diagnostic schemes. However, there are still some pending points to consider, such as the interpretability of the learning processes, the adaptation of the techniques and the different architectures to the case studies, the consideration of new faults or new operating conditions of the system, etc.

2.3.3 DBN for machine health monitoring

DBN is a model based on probability of energy generation, which comprises multiple layers of restricted Boltzmann machines (RBM) under a function of a backpropagation neural network (BPNN) [86]. The essence of DBN is the ability to automatically extract features through successive learning stages.

In the field of condition monitoring, this ability to automatically extract features is fundamentally useful, it allows characterizing complex environments that include multiple components, and which are also susceptible to the presence of different faults. In this regard, Jie Tao *et al* [87] developed a diagnosis method using multi-vibration signals and a deep belief network (DBN) model. The proposed scheme can adaptively fuse multi-feature data and identify various faults in bearings. The DBN-based model showed that it can learn both features from multiple vibration signals and individual features of each sensor. In addition to fuse vibration signals at different points in the system, a comparative study was carried out with state-of-the-art methods. The results show that the DBN-based model has higher diagnostic accuracy compared to schemes based on machine learning techniques. Chuan Li *et al* [88] propounded a method of diagnosing faults in gearboxes using acoustic and vibratory measurements. The signals of both physical magnitudes are fuse by extracting the statistical parameters of the wavelet packet transform (WPT) through two deep Boltzmann machines. The study of multiple sources of information has been a crucial issue in the field of condition monitoring, it allows integrating more information to generate more robust and reliable schemes, in addition to reducing biases caused by the use of a single physical quantity. In addition, signal processing generates considerable attention because a great part of the performance of fault detection and diagnosis schemes depends on it. For this purpose, Shao *et al* [89] suggested an adaptive method based on deep belief network with dual-tree complex wavelet packet. Under this processing, the proposed method is capable of diagnosing the fault severities and compound faults of rolling bearing. One of the challenges that had to be faced was during the training process of the RBMs. For this, an adaptive learning rule was adopted, which modifies both the learning rate and the momentum, adjusted at each epoch. The selection of a suitable architecture of the deep learning models, and also the configuration of the hyperparameters is a great challenge. Currently there is no clear strategy to select the optimal structures, so procedures and methodologies that help in this sense would contribute to the development of better fault diagnosis schemes. In this sense, Si-Yu Shao *et al* [90] presented a study of the structure of the DBN model. The analysis included an investigation of the scale and depth of the DBN architecture, parameters that directly affect the classification performance of the model. In addition, in this study the feature extraction procedure is combined with the classification task to achieve an automated fault diagnosis. The DBN model composed of several stacked RBM units is fed at the input by frequency data obtained from vibration signals. This frequency processing is practical since it allows the DBN-based model to learn features from frequency distribution of vibration signals, essentially useful information for the recognition of characteristic fault patterns in rotating systems. Another study that addresses the processing of raw vibration data is the proposed by Hyunseok Oh *et al* [91]. In this research, a method based on unsupervised feature engineering that uses transformation of vibration signals into images for fault diagnosis of rotor systems is suggested.

In addition, the work highlights the functionalities of using a deep-learning-based approach due to the high-level feature extraction abilities that allow to recognize and identify fault patterns through vibration-imaging. Furthermore, schemes based on deep-learning have not only shown the ability to diagnose and classify types of faults, studies such as the presented by Deutsch *et al* [92] developed an approach that using deep-learning to predict remaining useful life in rotating systems. This research uses a DBN-feedforward neural network (DBN-FNN) for RUL prediction using vibration sensors. The approach uses the capabilities of the DBN for the automatic extraction of features from vibration data to perform the predict. The presented method overcomes the performance of the prediction algorithms of the state of the art. The presented approach was validated with two different test-benches, one for the prediction of gear fault and another for bearing fault, demonstrating the adaptability for the study of different rotating equipment. Another research that has addressed the challenge of configuring architectures and hyperparameters of the deep learning models is the suggested by Zhiqiang Chen *et al* [93]. In the study, a comparison was carried out between three different deep-learning algorithms: deep Boltzman machines, deep belief network and stacked auto-encoders. The number of hidden layers, the number of neurons and four different processing schemes that include raw vibration signals, spectrum in time-frequency, statistical features in time and frequency domain were studied for each one. The investigation was performed to diagnose seven fault patterns in rolling bearings, which element is of great importance for the healthy condition of rotating machinery equipment.

2.3.4 RNN for machine health monitoring

Most of the existing monitoring schemes make use of sensor measurements to characterize and perform both diagnostic and detection tasks as well as the prediction of the health of the components. For this reason, the use of temporary signals has extended significantly in the field of monitoring electromechanical machinery. However, extracting useful information from the collected time series is an existing challenge. Avoiding the loss of local information, processing large amounts of data and effectively managing the information are some of the challenges present when working with temporary signals. In this sense, Recurrent Neural Networks (RNN) is a type of neural network capable of dealing with multiple time sequence data. The ability to capture relevant information from temporal signals by RNN has been demonstrated in various applications [94-97]. Additionally, tasks such as condition prediction or useful life forecast, which are necessary functions to be carried out in the condition monitoring field, can be performed by RNN, since they are very useful tools for managing of signals collected by sensors. In this regard, most of the investigations have focused on the diagnosis of faults after the occurrence, however, the prediction of the wear of machine elements or even the occurrence of faults before they occur is a vital effort in many sectors of the industry.

Thus, Yuan *et al* [98] investigated four RNN models, which include a standard RNN, a gate recurrent units (GRU), a modified LSTM, and a standard LSTM for remaining useful life estimation of aero engine. The investigation concluded that, to carry out the prognosis of the remaining useful life of aircraft turbofan engines, the standard LSTM model is the most suitable and with the best performance. Furthermore, the proposed prediction model can be adapted for the diagnosis of four types of faults under different operating conditions such as high pressure with single or six operation condition. Furthermore, the implementation of data-based monitoring schemes continues to be a challenge due to the difficulties present both, the complex manufacturing environment and the issues present during monitoring. The consideration of noise, the variable length and the irregular sampling in the data, are some of the difficulties to be faced within the condition monitoring schemes. In this sense, Rui Zhao *et al* [99] propose a scheme that combines local feature extraction using a CNN and temporal encode through a bi-directional LSTM model from vibration signals. The study focused on predicting tool wear from a CNC machine. The proposed scheme was compared against nine state-of-the-art methodologies, which include statistical features extracted from raw signal, shallow models based on back-propagation, machine learning models such as support vector machine and standard LSTM models. The results showed a superiority of prediction of the proposed approach. However, the study was limited to the use of a single processing, lacking the analysis of more effective transformations of the data for the study of machine monitoring condition. Then, Zhao *et al* [100] developed a hybrid approach that combines handcrafted feature design with automatic feature learning for the prognosis of the occurrence of failures in machines. The representation of the features is learned by a proposed bidirectional gated recurrent unit (GRU) model. Three different monitoring tasks are analyzed under this approach, which include: tool wear prediction, gearbox fault classification and incipient bearing fault detection. Although the suggested approach is practical for fault diagnosis, the approach has not been tested for prognostic tasks, such as RUL estimation. Continuing with the diagnostic tasks, in [101] Han Liu *et al* proposed a scheme for the diagnosis of faults in metallic bearings, which combines denoising autoencoders with RNN model. Under this proposed approach, the robustness of time series data reconstruction is improved, since it is possible to diagnose the type of fault based on the reconstruction error. However, the study is limited to the characterization of only four conditions that include normal condition and three types of bearing faults. The characterization of new faults or faults in different components of the machine are not contemplated. Although the main challenge for the prediction task continues to be obtaining data that allows the study of the remaining useful life of machine components, RNNs have been contributing to the generation of effective schemes associated with this task.

2.4 Summary and Conclusions

In this chapter, first of all, the background of data-driven schemes applied to condition-based monitoring for industrial systems is provided. Special attention is paid to the objective environment, where the application of monitoring schemes is not trivial due to the complexity of industrial systems, which involves the integration of various components such as motors, gearbox, wheels, screws and bearings, the consideration of different working conditions, and the presence of multiple faults under various severities.

Given the recent developments in hardware and software technologies and their rapid growth, they have allowed the emergence of advanced algorithms based on artificial intelligence (AI) capable of learning complex relationships from the data provided. Deep learning, as a subfield of machine learning, offers various AI-algorithms that have served as a learning tool in numerous applications. The main algorithms based on deep-learning can be summarized in four categories: stack-autoencoders, convolutional neural networks, deep belief network and recurrent neural network. Each of the algorithms is studied and its application to the field of condition monitoring is considered. Some of the first applications that have emerged in this regard are presented and some of the issues and challenges that are still present are identified.

Although the latest diagnostic methods around monitoring with the use of intelligent algorithms have achieved successful advances in many industrial applications, there are still many questions to be resolved before these schemes are widely adopted in practical industry systems. In addition, for intelligent diagnostic schemes to be adopted in manufacturing processes, they must meet highly demanding requirements, such as standardization, stability, repeatability, and precision.

Faced with this, some problems and challenges present in the diagnostic schemes have been detected. The first is that most diagnostic schemes are not very generalizable, which makes them insufficient in uncertain circumstances. This loss of generalization makes them unreliable in meeting the demands of complex modern manufacturing environments. A second issue that hinders its adoption in industrial systems is the lack of interpretability of artificial intelligence models. Deep-learning models are often referred to as "black box" techniques as they do not provide clear insight into how they make predictions. The third is regarding a lack of standardization of the configuration of the hyperparameters of the models. There is still no clear process on how to build architectures and set the hyperparameters of the models, which make them application-particular schemas. Hyperparameters are generally selected according to the application and the information available, which makes their application difficult in a general way.

3.

Diagnosis for Electromechanical Drives Monitoring

Fault diagnosis in electromechanical systems represents one of the most critical challenges dealing with condition-based monitoring in the recent era of smart manufacturing. In this regard, this chapter presents the contributions to improve the reliability and robustness of CBM data-driven approaches through methodologies based on deep-learning.

CONTENTS:

3. Diagnosis for Electromechanical Drives Monitoring

3.1 Introduction

This chapter describes the importance of fault diagnosis in the field of machine monitoring, a special emphasis is performed on the study of monitoring methodologies based on deep-learning algorithms. As a result of this study, the proposal and development of a novel methodology for the implementation of DL models to deal with the issues present in the monitoring of smart manufacturing processes is carried out. The background and motivation are presented as follows, followed by reviews of the state of the art related to machine health monitoring and describes the contribution proposal of this work.

3.1.1 Background and motivation

As mentioned in the previous chapters, the advances in digital technologies and their integration with industrial machinery have allowed the existence of more advanced manufacturing environments, and consequently with more complexity. Chains cinematics or powertrains are increasingly composed of a greater number of elements, ranging from induction motors or permanent magnet synchronous motors, couplings, gearboxes, bearings and screws. Additionally, operating environments are characterized by variability in working conditions, which implies changes in the speed of rotating components, increased load and torque. Monitoring methodologies for fault diagnosis are an important concern for predictive maintenance experts. In this regard, the development of reliable and robust condition-based maintenance programs plays a key role in this complex industrial environment. With the implementation of condition-based maintenance schemes, it is sought that the machinery is operational for as long as possible, reduce stoppages caused by unforeseen failures and reduce economic losses due to production stoppages [102].

The progress of algorithms based on Artificial Intelligence and their successful application in various research areas make them ideal tools to deal with the existing complexities in industrial environments. Specifically, these techniques have been applied to characterize the health condition of machines through the measurement of physical parameters, such as vibrations, stator currents, temperature, magnetic flux, or others [103-106]. Therefore, various schemes have been proposed, in which the conventional workflow establishes the use of multi-modules. Different strategies participate in each of these stages or modules, particularly algorithms based on machine learning, a subfield of AI focused on "learn" characteristics and patterns from data to perform a specific task. The modules can vary depending on each application but usually involve three parts: feature extraction, feature selection/reduction and model training.

Although these methodologies have achieved notable advances in the field of condition-based maintenance, they present difficulties when facing the complex scenarios of the industry, mentioned above. The main problem is that each of the modules is trained and optimized in a particular way and cannot be jointly optimized which may harm the final performance of the whole scheme. Therefore, as the complexity of the system increases, the performance of ML-based methodologies is expected to decrease, and consequently the robustness of the maintenance scheme will be affected.

Under this framework two main issues are presented. The first, the methodological proposal of a diagnostic scheme to deal with the increase in complexity of industrial rotary systems. Second, the methodological proposal must take into account the requirements involved in carrying out a diagnostic scheme, especially the type of physical magnitude that will be used as input to the model, signal processing, the architecture of the artificial intelligence model, as well as the establishment of its hyperparameters.

3.1.2 State of art in diagnosis based on DL

Since its inception in 2006, deep-learning (DL) has become the fastest growing research tool, directing and redefining various research in a wide range of areas, such as pattern recognition, image identification, natural language processing and machine translation [107-110]. The application of the DL has been extended to reach the implementation in industrial environments. Its implementation has also been motivated by the need to face the increasing complexity of production processes, which have been characterized by increased demand, higher product quality and greater manufacturing precision [111]. Furthermore, avoiding production stoppages due to breakdowns and that these cause losses has been one of the main motivations for the implementation of AI-based models. In this regard, numerous studies based on DL have been proposed around the monitoring of the condition of industrial systems with the aim of contributing to the development of more robust and reliable prediction schemes [112]. However, has not been proposed a study framework that includes a methodological process for the implementation of a diagnostic model, taking into account both the acquired signals, the processing stage, the configuration and adequacy of the model based on deep-learning and the correct characterization of different fault conditions.

Regarding signals measured on motors, studies such as the one presented in [113], confirm the importance of using a multisensor approach to carry out condition monitoring. Specifically, the vibration and current signals are the ones that provide the most information to characterize the state of the machine and consequently carry out the fault diagnosis. A similar comparative investigation [114] concludes that, depending on the nature of the fault, whether it is a mechanical failure or an electrical failure, it is more effective to use only one type of signal, such as vibration or current. However, since it is not possible to predict what type of fault will occur, it is most appropriate to use both signals together.

Therefore, the investigation of effective strategies for the integration of multiple signals becomes another challenge for the field of condition monitoring. Indeed, most of the current methodologies use only one physical magnitude to carry out monitoring, which implies that there may be unseen operational points of view and the lack of a comprehensive vision of the condition of the machinery. Existing studies have not been able to resolve how to integrate different sources of information, either how to feed DL-based feature learning models. Furthermore, the performance of DL-based algorithms is not only affected by the type of signal used as input, but also by the architecture and hyperparameters of deep model. Therefore, building, selecting, and fitting hyperparameters during the construction of a diagnostic model is a crucial step. However, there is no consensus on the correct way to select and fitting hyperparameters. Generally, in the applications that involve the use of deep models, the hyperparameters are established through manual configuration and exhaustive search by grid search technique, which implies a great effort and time-consuming to obtain optimal results [115]. One way to solve this issue could be through the implementation of automatic search and selection processes. Thus, by combining deep models with search algorithms, high-quality solutions can be generated to optimize diagnostic schemes, adapting to the variability presented in each work environment.

Another subject that requires attention when implementing DL models in condition monitoring environments is that these methods have been perceived as “black box” tools and lack interpretability. Although DL methods have made significant achievements in the field of condition monitoring and other research fields, the aspect of lack of interpretability has been a recognized limitation for implementation in real industrial environments. Applications inside industrial environments have strict requirements regarding security, accuracy and reliability, therefore the interpretability and credibility of the applications cannot be in doubt, but must be convincing in all aspects. In this regard, a limited number of publications have focused on addressing this issue. Fortunately, in recent years, interpretability in machine learning processes has increasingly captured the attention of academic researchers [116], making an effort for monitoring applications based on artificial intelligence models to comply the demanding requirements of the industry.

3.1.3 Methodological proposal

In the development of this chapter, a methodological process is proposed for the implementation of a deep model to perform fault diagnosis in electromechanical systems. The methodology aims to provide a step-by-step process considering existing issues within the field of condition monitoring in the state of the art. The proposal takes into account different physical magnitudes that can be obtained from the monitoring process, signal processing to highlight fault indicators or patterns, as well as the construction of the deep model and the optimal configuration of hyperparameters.

The proposed methodology aims to develop a novel monitoring and fault diagnosis scheme based on deep-learning with the ability to extract features from different available signals and from different signal processing, such as time domain, frequency domain. and time-frequency domain. With this objective, the characterization of the different fault states and the different operating conditions of the electromechanical system is carried out. Then, the built model based on DL is optimized by tuning the hyperparameters with the support of an automatic search and optimization algorithm. While the hyperparameters are adjusted, it is sought that different metrics, such as the classification accuracy or the reconstruction of the signals, are the most appropriate. Finally, a sensitivity study is carried out to confirm that the learning process is appropriate. Through this study, it wants to confirm that the patterns learned from the signals correspond to behaviors related to the condition of the machinery and not to external factors. Thus, is generated an interpretable diagnostic methodology that helps to understand the learning behavior of the deep models.

The main contribution of this study lies in the proposal of a new data-driven monitoring methodology based on deep learning for fault diagnosis in complex electromechanical drives, and in the verification of the learning process of the deep-learning model used to characterize the various conditions presented in the system.

Aligned with the current research challenges in the field of condition monitoring and in the security and reliability demands requested in the industrial environment, the proposed methodology takes advantage of the data acquired from the different physical magnitudes through monitoring to have a comprehensive vision of the condition of the system. Furthermore, due to the difficulty in predict the nature of the type of failure, whether it is mechanical or electrical, a method of integration of multiple signals is proposed, which can include vibrations and current, among others, and from different domains, resulting in a more robust scheme capable of performing the diagnosis of mechanical, electrical and combined faults.

The novelty of this work includes the proposal of a procedure that includes the implementation of various tools that together offer important advantages over traditional studies focused on fault diagnosis. In particular, in the diagnostic approach, signal processing techniques, learning algorithms based on artificial intelligence, search and optimization techniques, and feature fusion methods collaborate, allowing its adaptability to operating conditions and the different faults presented, improving the resulting precision through a global diagnostic scheme.

Therefore, the result of the TP values is given by the anomaly score obtained through OC-SVM, while the TN values are given by the combination of OC-SVM and the resulting samples above the reconstruction model threshold.

3.2 Theoretical approach and proposed contribution

The procedure of the proposed method is shown in **Fig. 3.2.1** step-by-step, which is divided into five main stages: signal processing, where transformations are performed on the acquired signals, feature learning, where the behavior of the machine is characterized by means of an artificial intelligence model, multisource fusion, where the features extracted from the different domains are combined and classification, where predictions are obtained and the model is evaluated through the classification accuracy.

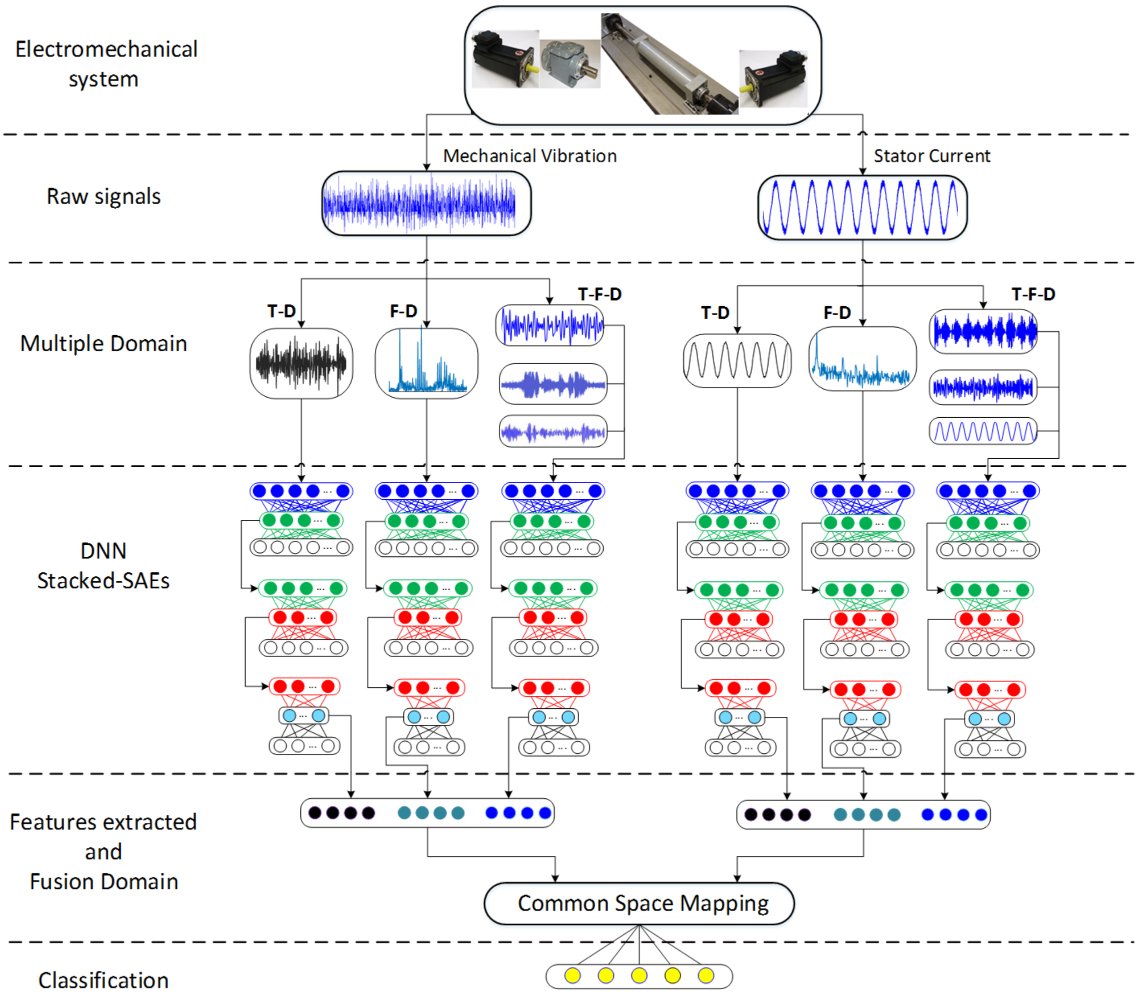


Fig. 3.2.1 Steps of the proposed fault diagnosis methodology based on hybrid feature learning and feature fusion integrated by deep-autoencoders.

The developed methodology is fundamentally a multiclass fault classifier. For each sample collected from the acquisition, the proposed approach generates a probability distribution for each system condition. This means that under this approach it is possible to determine if a certain type of fault occurs or not, comparing the probability between the different fault conditions and the healthy condition. Thus, a tool for decision making for maintenance is provided.

First, in the signal processing stage, the data is segmented and filtered to remove noise and atypical frequency components. In this regard, consideration of digital filters for signals greatly helps to remove spurious harmonics, in this work a digital FIR-based low pass filter is implemented to remove low frequency components commonly associated with noise. The signals are then converted into different domains, that is, frequency domain and time-frequency domain, in addition to considering the time domain (raw data). The key idea of this step is to incorporate as much information as possible prior to the analysis of the deep neural network.

Afterward, an architecture based on stacked auto-encoders is used to learn the features of each of the domains of each physical magnitude (vibration or current), separately. By performing the learning stage by SAE over signals, the features are more robust due to the representation-based learning process. That is, the SAE tries to reconstruct the input signal at the output, which implies that through the layer-by-layer process more effective features are learned, something especially useful when dealing with the randomness of the operation of rotary systems. The features extracted from each vibration and current domain are concatenated to group the information obtained from each signal, then a dimensionality reduction stage is performed under the linear discriminant analysis (LDA) algorithm, thus, the fusion of the features is obtained in a reduced dimension.

Finally, an ANN is used to generate the membership probability of each condition of the electromechanical system. The following subsections present the main stages of the methodology in detail.

3.2.1 Signal processing

It is believed that artificial intelligence (AI)-based methods can provide functionalities within the field of condition monitoring due to their extraordinary ability to learn complex relationships from data, and thus generate more reliable and efficient monitoring schemes. However, something that is often overlooked is the quality of the signals that are introduced to these learning algorithms, since for these methods to have high efficiency, the quality of the data that is introduced must be as descriptive as possible and provide representative information on the condition of the machine. Therefore, to get the best performance from AI-based models, it is desirable to perform signal processing beforehand, such as segmenting the signals considering several rotation cycles and performing filtering procedures. Removing noise and unwanted harmonics from signals is an effective way to process information before introduce it into learning models. During this proposed methodology, a digital FIR-based is employed as a filtering tool under the window method. The digital low pass filter by the windowing method operates under principle of the ideal “brick wall” filter, using a cut-off frequency of ω_0 described as follows:

$$h(n) = \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{j\omega t}\, d\omega \tag{35}$$

where, the low pass filter generates a magnitude equal to one for all frequencies that have frequency values less than cut-off frequency ω_0, and generates a magnitude equal to 0 for those frequency values between ω_0 and π. Then, its impulse response sequence is represented by $h(n)$. With the digital FIR-based low pass filter, can effectively remove spurious components and ensured that the fault-related features clear enough to be to be characterized.

In addition to signal cleaning, considering different domains ensures that as much information as possible from the acquired signals is being taken into account during the learning process. That is, consider learning from the space of features in the time domain (TD), the frequency domain (FD) and in the time-frequency domain (TFD). Although most applications of fault diagnosis in rotating machines consider only the feature space of a single domain, considering multiple domains allows taking into account valuable information from the signals, in addition to eliminating blind spots in the analysis. First, for the TD the segmented and filtered raw signals taken since the acquisition are considered. Second, for the FD the Fast Fourier Transform (FFT) is used and applied to each segment of the signal to obtain the corresponding frequency amplitude. Third, TFD is obtained through signal decomposition. The Hilbert-Huang transform (HHT) is a self-adaptive processing algorithm for nonstationary and nonlinear signals, the Empirical Mode Decomposition (EMD) is a signal demodulation method based on HHT that recursively scans local minima or maxima and calculates the lower or upper envelopes by spline interpolation, that can be cubic, polynomial and/or smoothing [117]. The Hilbert-Huang transform of a time-signal x(t) is described as follows:

$$\tilde{x}(t) = H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} x(\tau)\frac{1}{t-\tau}\, d\tau \tag{36}$$

The EMD generates a frequency bands of complete and quasi orthogonal intrinsic mode functions (IMFs) using the HHT. Thus, fault-patterns could be extracted through this process. Therefore, when considering these three domains, a more informative monitoring process is obtained that will later be characterized by methods based on deep neural networks.

3.2.2 Feature learning

The characterization or feature learning stage consists of creating a deep neural network (DNN) based on stacked-autoencoder (SAE), as shown in **Fig 3.2.2**. For each of the domains obtained from the previous process, several auto-encoders are stacked to build the DNN. The SAE-based DNN aims to extract intrinsic relationships and learning the mapping of complex features between the sampled signals and fault conditions of the system under supervision.

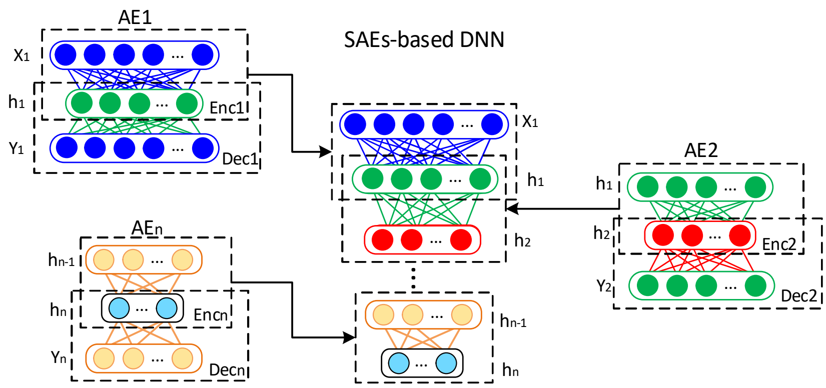


Fig. 3.2.2 Scheme for the construction and training of a deep neural network based on SAE.

The DNN architecture is composed of an input layer, an output layer and several hidden layers, each of which can learn a non-linear transformation from the previous layer. The training process follows a layer-by-layer method proposed by Hinton [55]. First, the dimension of the input of each network is established, therefore the number of input neurons corresponds to the dimension of the data for each of the domains. An unlabeled X_{TD}, X_{FD}, and X_{TFD} training data set is then used, corresponding to the time, frequency, and time-frequency domains. X_1 is the input to the first AE1, the hidden layer is h_1, corresponding to the encoding, and Y_1 is the reconstructed output. X_1 and h_1 also represent the input of the DNN and its first hidden layer, correspondingly. Then, h_1 becomes the input of the second auto-encoder AE2, which is trained to obtain the encoding vector h_2, which corresponds to the second hidden layer of the DNN. The DNN training process follows the steps described in subsection 2.2.1.2 stacked auto-encoder. Finally, each SAE is trained and the DNN is initialized according to the previous steps up to the *n*-th hidden layer. The last output layer of DNN corresponds to the space of extracted low-dimensional features that will be served to obtain the representation of the extracted features.

In addition to the creation of the DNN structure, its configuration and the selection of the hyperparameters turn out to be a complex task. To deal with this optimization problem, in which the network parameters must be adjusted to minimize the cost function, a process of optimization of the key parameters for the construction of DNN is carried out. In this work, an optimization of the hyper-parameters is performed by a genetic algorithm (GA). This optimization is carried out, individually, for each domain space considered as follows: a logical vector containing the key parameters of optimization to minimize the cost function (8) of subsection 2.2.1.1, and the unit numbers of the hidden layers. The DNN parameters are as follows: the coefficient for the L2 regularization term, the coefficient for the sparsity regularization term and the parameter for sparsity proportion. The unit numbers of the hidden layers are also obtained through the GA optimization process. The number of neurons in the output layer is established in two to generate a two-dimensional space that allows the mapping of the data to be represented visually. Therefore, an initial population is randomly generated to be evaluated. Some restriction criteria are established for each of the parameters contained in the logical vector with the aim that the GA converges on the optimal values.

With the initial population generated, the fitness function is evaluated, which corresponds to minimizing the reconstruction error through Equation (8). Then, a new population is generated by the roulette wheel selection process and applying a mutation in the GA, based on a Gaussian distribution. Then, the process is repeated iteratively until finding the best set of hyper-parameters that perform a better reconstruction of the data for each domain. The GA stops until the minimum reconstruction error is reached or a maximum number of generations is reached.

3.2.3 Multisource fusion

After building the learning model and performing feature extraction for each of the domains, the feature set is subjected to a fusion process. In this regard, linear discriminant analysis (LDA) is used to perform this fusion process, in addition LDA allows to generate a new mapping of the data of a low dimensionality. Therefore, the new feature mapping represents the combination of the set of features obtained from the different domains through the DNN architectures. The fusion space is mapping in a two-dimensional space, allowing a visual interpretation of the conditions considered.

Linear discriminant analysis is one of the best-known supervised feature fusion and dimensionality reduction algorithms for multiclass problems. The aims of LDA is to find a mapping of features in a low-dimensional representation to maximize the linear separation between the most discriminating information belonging to different classes. The criteria used by the LDA algorithm is to first evaluate the compactness within each class by the calculation of the within-class scatter matrix, S_W, while also calculating the separability of the different classes by the scatter matrix, S_b. LDA aims to find a linear transformation matrix, $M \in R^{Cxd}$, mapping the original C-dimensional space into a reduced d-dimensional feature space with $d < C$, for which the between-class scatter matrix is maximized, whereas the within-class scatter matrix is minimized. Considering M the total data set, each point m_i belongs to a class $z_i = \{1, 2, \dots, z\}$. Where y_i is the number of data points in the *i*-th class and y is the number of data points in all classes. The between-class scatter matrix, S_b, and the within-class scatter matrix, S_W, are defined as follows:

$$S_b = \sum_{k=1}^{z} y_k (\mu_k - \mu)(\mu_k - \mu)^T \tag{37}$$

$$S_w = \sum_{k=1}^{z} \sum_{m_k \in z_k} y_k (m_1 - \mu_k)(m_k - \mu_k)^T \tag{38}$$

where $\mu_{k=} {}^{1}/_{y_k} \sum_{m_k \in z_k} m_k$ is the mean of the data points in the *i*-th class, and $\mu = {}^{1}/_{y_k} \sum_{k=1}^{z} m_k$ is the mean of the data points in all classes. With common space mapping, all the learned representation of different domains obtained from the DNNs can be merged. Compared with the representation learned under a single domain, the multisource representation obtained gives a more complete representation of the conditions of the electromechanical system.

3.2.4 Condition assessment

Finally, the condition assessment of the rotating system can be predicted using the multisource features space from the trained models obtained following the previous steps. The processing of the acquired signals provides a global vision of the state of the machinery. Then, the learning model based on deep neural networks characterizes the different states generating a set of features from multiple sources of information. The fusion and dimensionality reduction step allows the development of a simple configuration for the classification task due to the data is embedded in a two-dimensional space.

In this regard, a simple structure of an ANN-based classifier is used to obtain the diagnosis assessment of the considered conditions. Indeed, the ANN-based classifier has a classical three-layer structure. The input layer is composed of two neurons corresponding to the 2-D feature vectors resulting from the multisource fusion stage. The hidden layer is established in eight neurons following recommendations [118]. The output layer is established according to the number of conditions considered. This simple ANN structure has been successfully implemented in different condition monitoring schemes [89]. In addition to the condition assessment, the ANN-based classifier also provides the corresponding probability value due to the sigmoid function is used as the activation function in the output layer, providing a diagnostic probability. The ANN training process uses the backpropagation method to compute the gradient and the scaled conjugate gradient as a minimization technique.

3.3 Experimental implementation and analysis of results

This section shows the implementation of the proposed methodology for fault diagnosis, and discusses the results obtained using the test bench in Annex 1. The experimental test bench consists of two identical motors connected by means of a screw and a gearbox. One of the motors drives the system directly connected to the gearbox and the other works as a load. The output axle of the gearbox runs the screw, which, at the same time, displaces a movable part. From this experimental test bench, vibration signals and stator currents are acquired for monitoring. Therefore, these two signals are used to carry out the methodological proposal. On the one hand, the vibration signals of the axis perpendicular to the rotation of the motor are used, according to the investigations related to the diagnosis of faults, the most relevant information on the condition of the machine is present in these axes perpendicular to the rotation. On the other hand, electrical faults present in motors cause phase modulation in the current and affect the magnetic field generated by multiple frequency components. Therefore, the implementation of this methodology is carried out around the measurements of mechanical vibrations of the plane perpendicular to the axis of rotation of the motor, while for the study of the current one of the motor supply phases is taken.

In order to perform a more complete characterization of the system, a parallel processing of the vibration and current signals is carried out. Under this processing, the proposed methodology has the capacity to generalize the occurrence of faults of different nature, even if it is unknown if the type of fault is mechanical or electrical.

The set of data acquired through the test bench allows extracting a considerable number of samples to carry out the analysis, since multiple acquisitions of the signals were made. In this regard, and following the proposed methodology, the vibration and current signals are segmented into samples containing 4096 data points. Each one of these samples also represents two complete turns of the motor, so that in each one of the samples the rotational behavior of the system is represented. After the segmentation, the expected processings are applied, that is, the FFT and EMD transforms are applied to represent the frequency domain and the time-frequency domain, respectively. Therefore, the samples in TD (raw data), FD and TFD contain 4096, 2048 and 4096 data points, respectively. A representative sample of each vibration and current signal in TD is shown in **Fig. 3.3.1**. Then, the representation of the resulting spectra for a vibration and current signal for frequency domain are shown in **Fig. 3.3.2**. Finally, three main intrinsic mode components obtained to represent the time-frequency domain are shown in **Fig. 3.3.3**.

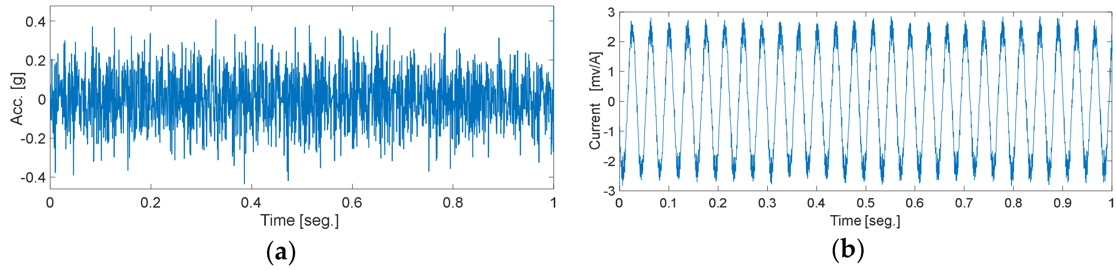


Fig. 3.3.1 Signal samples in time-domain from the same time step. a) Vibration signal, b) current signal.

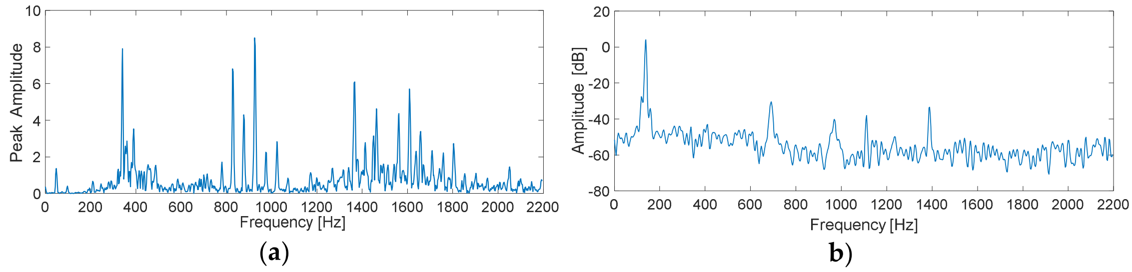


Fig. 3.3.2 Signal samples in time-domain from the same time step. a) Vibration signal, b) current signal.

The consideration of multiple physical magnitudes and multiple domains of the signals allow considering a study under Multiple-Domain-Features (MDF), therefore a more robust analysis is performed, which guarantees greater reliability and reliability in the diagnosis. The MDF is the most effective way to carry out a study of the condition, because the state of the system is characterized in the broadest way. In addition, in this way blind spots are prevented during monitoring, since large of the information related to fault conditions is not present in a single signal or is not perceptible in a single domain. Thus, the set of MDFs is then analyzed through a hybrid learning and feature reduction model proposal based on deep learning to characterize the condition of machinery and diagnose the presence of faults.

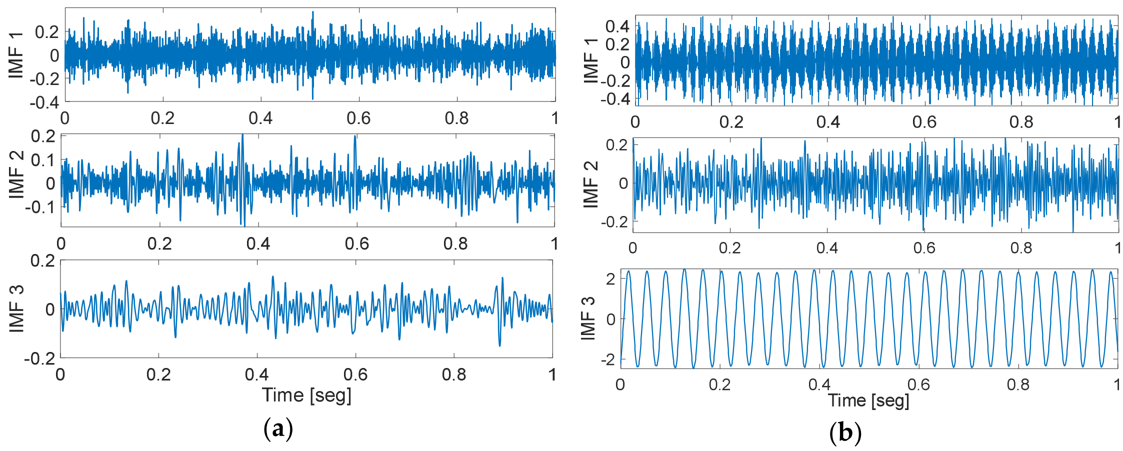


Fig. 3.3.3 Signal samples in time-domain from the same time step. a) Vibration signal, b) current signal.

Next, the characterization model based on deep learning is built, which both the structure and the hyperparameters must be configured. First, the architecture of the DNN based on stacked-autoendors is configured using a genetic algorithm (GA).

The parameters to be optimized through the GA are those described in the previous subsection, while the GA implementation is defined as follows: an initial population of five is established for the number of individuals, previously defined in the methodology, the maximum number of generations is established empirically at 50, and a roulette selection scheme is set together with a mutation algorithm based on a function of selection with a Gaussian distribution. The optimization process is applied individually to each of the considered MDF domains. Thus, the objective of GA is to find the optimal hyperparameters with which the DNN models based on staked-autoencoders estimate a minimum value of MDF reconstruction error.

The results of the optimization process by GA conclude with the configuration of hyperparameters that are established for each of the available physical magnitudes, vibration and current. **Table 3.3.1** and **Table 3.3.2** show the results obtained for vibration and current, respectively, and for each of the domains.

Table. 3.3.1 Hyperparameters established for the DNN stacked-autoencoder-based for vibration signals.

Feature Domain	Hyper-parameters				
	L2 regularization	Sparsity regularization	Sparsity proportion	First hidden layer	Second hidden layer
Time-Domain	5.0×10^{-5}	5.0×10^{-5}	0.4	670	208
Frequency-Domain	5.0×10^{-5}	5.0×10^{-3}	0.5	500	110
Time-Frequency-Domain	1.0×10^{-5}	5.0×10^{-5}	0.4	635	145

Table. 3.3.2 Hyperparameters established for the DNN stacked-autoencoder-based for current signals.

Feature Domain	Hyper-parameters				
	L2 regularization	Sparsity regularization	Sparsity proportion	First hidden layer	Second hidden layer
Time-Domain	5.0×10^{-6}	5.0×10^{-6}	0.09	700	220
Frequency-Domain	5.0×10^{-6}	5.0×10^{-3}	0.05	600	150
Time-Frequency-Domain	5.0×10^{-5}	5.0×10^{-5}	0.10	750	160

Once the DNN architecture and hyperparameters have been configured, each of the models for the MDFs is trained to learn the intrinsic modes of the system and extract features for each domain. Then, the space of features extracted from each domain is submitted into a features fusion stage by LDA, in which all the MDFs are subjected to the compression procedure. As a result, a compact feature mapping is obtained in which the different acquisition signals and the different transformation processing in different domains are combined. This mapping of features allows, on the one hand, to effectively integrate the characterization of various sources of information in such a manner that the classification of the condition of each of the samples acquired can be carried out, on the other hand, to provide a visual representation of features in a two-dimensional space.

Finally, the classification stage is carried out by applying a multilayer ANN-based classifier. The compact feature mapping is used as input to the ANN-based classifier, which has a 2-D dimension.

The ANN classifier has a hidden layer configured with 10 neurons, while the output layer corresponds to the number of faults conditions considered. The back-propagation algorithm is used for training and the probabilistic sigmoid function is set as the activation function. To perform the condition assessment, the average diagnostic accuracy is introduced as performance metric. Additionally, to demonstrate the generalization capabilities of the proposed methodology, a cross-validation strategy was followed. The cross-validation implementation removes a subset of data from the full dataset, the model is trained using the remaining data, and validation is performed with the removed subset. A five-fold cross-validation scheme is considered, and the diagnostic accuracy results are presented as the mean value of the cross-validation.

The training and validation experiments consist of nine test scenarios. Each test scenario consists of the combination of a power supply and a load condition. The nine scenarios considered for the experiments and the combinations of the operating conditions of the electromechanical system are described in **Table 3.3.3**.

Table. 3.3.3 Performance comparison between different acquisition signals and multiple feature domains.

Index		Power supply	Load condition
1	Ps1-Lc1	30 Hz	40 %
2	Ps1-Lc2	30 Hz	70 %
3	Ps2-Lc1	60 Hz	40 %
4	Ps2-Lc2	60 Hz	70 %
5	2Ps-Lc1	30 Hz & 60 Hz	40 %
6	2Ps-Lc2	30 Hz & 60 Hz	70 %
7	Ps1-2Lc	30 Hz	40 % & 70 %
8	Ps2-2Lc	60 Hz	40 % & 70 %
9	2Ps-2Lc	30 Hz & 60 Hz	40 % & 70 %

The test motor was powered by two power frequencies (30 and 60 Hz) and loaded by two load conditions (40% and 70% of rated load). Therefore, the test scenarios considered in **Table 3.2.3** are presented in a way that the complexity of the operating conditions are increasing. First, indices 1-4 present simple conditions by considering a single power supply and a single load condition. Then, indices 5-8 present combined operating conditions, that is, they present two power supply conditions or two load conditions. Finally, index nine represents the highest level of complexity when presenting multiple operating conditions.

In order to quantify the performance provided by the experiments carried out of the proposed methodology, the average results obtained through the cross-validation scheme are shown in **Table 3.3.4**. In this regard, the results of the nine test scenarios provided in **Table 3.3.3** are presented for each of the physical magnitudes considered, that is, vibration, current, and the proposed scheme that contemplates the integration of both. As well as the different signal processing, that is, time domain, frequency domain and time-frequency domain, and the fusion of all of them. First, the results obtained will be discussed under the proposed approach of information fusion and then without fusion. In this regard, the average of the accuracy of the proposed method is 92.03%, while the average of the accuracy for current fusion and vibration fusion is 35.01% and 88.6%, respectively.

Clearly the accuracy results obtained for the current measurements are inferior to those of vibration and the proposed approach. These results are in alignment with previous research, and it is mainly due to the nature of the faults considered, that is, they are mostly mechanical faults and not faults of an electrical nature. Regarding the results without fusion, the best accuracy obtained corresponds to the frequency domain for vibration measurements with 92.2%, similar to this accuracy, the approach proposed in the frequency domain obtains an accuracy of 88.90%.

Table. 3.3.4 Performance comparison between different acquisition signals and multiple feature domains.

Index		Current				Vibration				Proposed Approach			
		TD	FD	TFD	Fusion	TD	FD	TFD	Fusion	TD	FD	TFD	Fusion
1	Ps1-Lc1	25,4	32,2	39,3	42,5	70,3	99,8	63,8	95,3	65,1	96,0	54,9	**95,3**
2	Ps1-Lc2	26,6	36,9	35,0	42,0	66,2	83,2	54,8	83,3	43,8	84,8	58,5	**86,3**
3	Ps2-Lc1	27,0	38,3	23,0	41,2	68,5	99,8	74,6	98,1	64,2	97,3	63,0	**92,3**
4	Ps2-Lc2	21,9	40,4	23,4	22,6	66,0	96,7	86,0	99,5	53,0	99,5	87,3	**99,0**
5	2Ps-Lc1	26,3	35,6	36,2	39,8	65,2	92,0	58,9	89,6	59,8	87,6	51,2	**93,5**
6	2Ps-Lc2	23,2	33,6	33,6	38,7	64,8	91,5	52,1	80,2	44,3	87,2	55,8	**92,6**
7	Ps1-2Lc	25,5	38,1	24,8	37,9	66,3	92,0	71,0	87,9	57,8	88,1	59,8	**92,8**
8	Ps2- 2Lc	22,8	39,3	22,4	27,8	63,5	90,5	83,5	86,3	45,6	81,5	63,0	**91,2**
9	2Ps- 2Lc	21,3	32,1	21,0	22,6	63,5	85,1	63,0	77,8	43,1	78,4	50,1	**85,3**
Average		24,4	36,2	28,7	35,01	66,03	92,2	67,5	88,6	52,9	88,9	60,4	**92,03**

Although for this particular case a better diagnostic performance is obtained, it should be noted that the proposed approach obtains better performance in most test scenarios. In general, for each of the physical magnitudes considered, the results obtained under the fusion of domains are higher than those of the individual domains. In this regard, it should be noted that the knowledge of the nature of the failure is not always known and the presence of both mechanical and electrical failures can appear indiscriminately, therefore the main motivation of the proposed methodology is that it is capable of adapting to the different types signals and the different domains to carry out the diagnosis effectively.

In addition to the quantitative diagnostic results, the proposed methodology is validated in a qualitative way. In this respect, in order to interpret the distribution of the considered conditions, a 2-dimensional space is projected where it is possible to obtain a visual representation of the extracted features by DNN model. **Fig. 3.3.4** shows the mapping of the set of extracted features resulting from the characterization process of the proposed approach by extracting and fusing the MDF for the operating conditions of index 5 in Table 3.3.3. From the projection map, it can be seen that the features associated with the different health conditions of the electromechanical system are generally clearly separated from each other. The conditions that present a type of overlap, only a few test samples are grouped incorrectly. Specifically, in the mapping corresponding to the classification of **Fig. 3.3.4** (b), some data points of the healthy condition and gear fault condition are overlap with each other. Moreover, a small number of data points of the healthy condition and demagnetization fault are overlap. While the bearing fault condition and eccentricity fault are clearly separated.

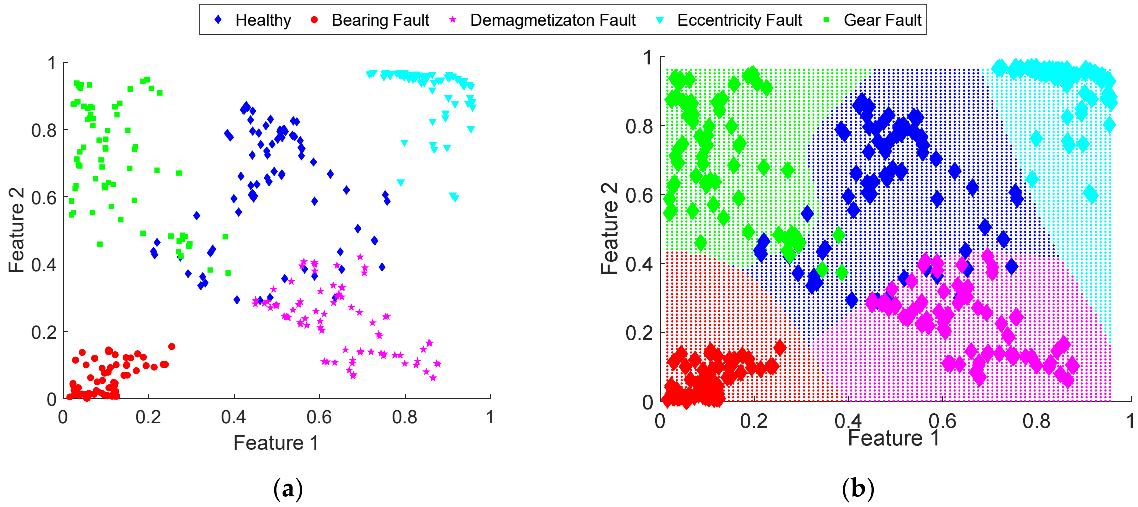


Fig. 3.3.4 Mapping of the set of extracted features obtained through the proposed approach: (a) 2-dimensional mapping representing each of the conditions of the electromechanical system; (b) mapping resulting from the ANN-based classification algorithm.

In this way, the high learning capacity and extraction of features of the DNN-based model can be confirmed. Since the SAE-based model is capable of correctly characterizing each of the conditions of the electromechanical system, the classification task becomes simple. In addition, in the classification stage, an artificial neural network is applied with a sigmoid activation function, which allows the probability for each of the classes, this also allows generating the decision regions for each condition of the system, as shown in **Fig. 3.3.4** (b).

3.3.1 Comparison with other approaches

In order to verify the effectiveness and performance of the proposed diagnostic methodology, the results obtained are compared with different existing diagnostic schemes. Classic machine learning approaches based on the extraction of engineering features (FE) and a technique based on deep-learning, widely studied and validated methods in the field of monitoring of electromechanical systems are used as a comparison. Seven different schemes are used for this purpose and they are listed below:

1) Linear discriminant analysis + ANN-classifier
2) Linear discriminant analysis + SVM
3) Principal component analysis + ANN-classifier
4) Principal component analysis + SVM
5) Isomap + SVM
6) Laplacian + ANN-classifier
7) LSTM

For the ML approaches (1-6) the classical procedure of feature extraction, dimensionality reduction and classification is performed. A total of 15 statistical features are calculated to carry out the extraction stage, after which the corresponding dimensionality reduction technique is applied.

Regarding the dimensionality reduction techniques, linear discriminant analysis (LDA) and principal component analysis (PCA) are linear reduction tools, while Isomap and Laplacian-eigenmaps are non-linear reduction techniques. Regarding Lstm, it is a technique based on deep-neural networks described in Subsection 2.2 **Deep Learning**.

Table 3.3.5 summarizes the average accuracies of the classical fault diagnosis approaches in comparison with the proposed method under the same operating conditions and the same fault states. The average accuracy for all the case studies of the proposed method is 92.03%, which represents the best performance of all the approaches presented. In contrast, the resulting accuracy for the two LDA schemes with ANN and SVM are 85.61% and 86.26%, respectively. Similar to these, the two PCA approaches with ANN and SVM obtain average accuracies of 81.86% and 83.18%, respectively. Although good results can be considered, the proposed method is still superior in a range of 5% and 10%. Non-linear dimensionality reduction techniques do not obtain better results, for example, Isomap has an average accuracy of 83.11% and Laplacian an average accuracy of 81.5%, around 10% lower than the performance of the proposed method. Finally, the performance of Lstm is presented, a state-of-the-art technique used for diagnosis and other areas of pattern recognition, obtaining an average accuracy of 88.63%, which is clearly higher than that of classical schemes but does not overcome the performance of Proposed method.

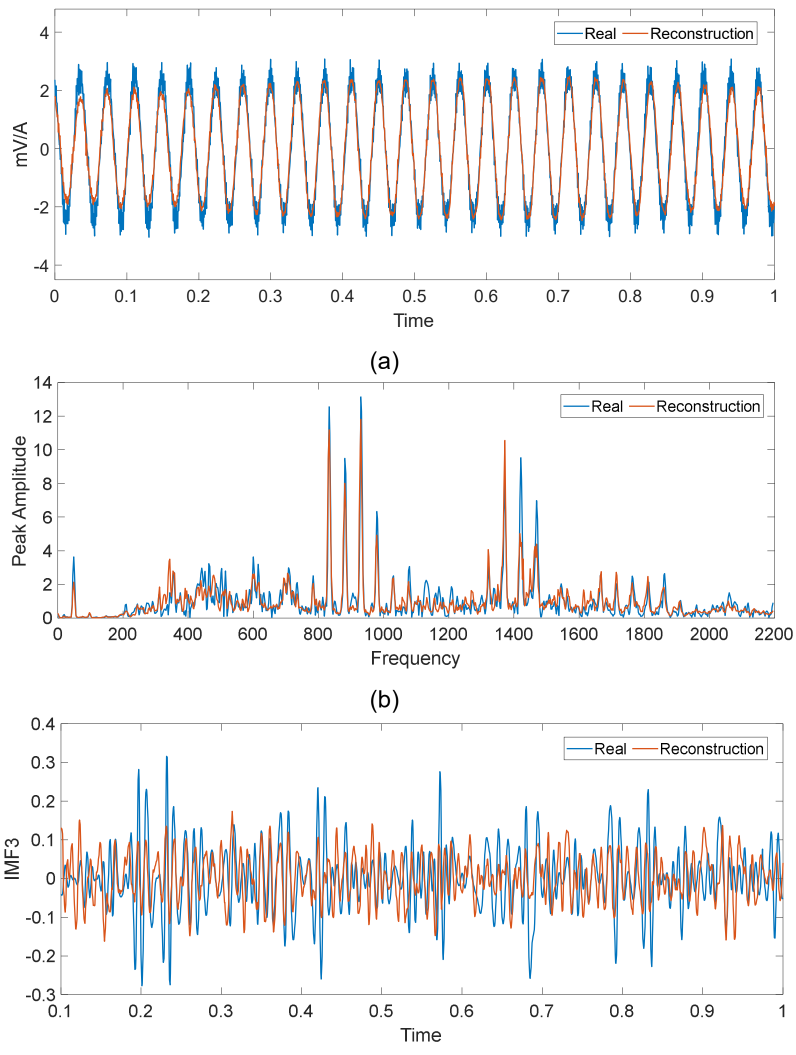
Table. 3.3.5 Performance comparison between different acquisition signals and multiple feature domains.

Index		Diagnostic methods							
		LDA +ANN	LDA +SVM	PCA +ANN	PCA +SVM	Isomap +SVM	Laplacian +ANN	LSTM	Proposed method
1	Ps1-Lc1	92.5	93.1	90.5	91.3	98.0	93.0	95.0	**95,3**
2	Ps1-Lc2	88.6	81.9	76.7	77.8	86.3	85.6	85.1	**86,3**
3	Ps2-Lc1	91.3	91.1	88.1	89.3	84.6	86.0	89.2	**92,3**
4	Ps2-Lc2	88.5	95.9	90.5	89.6	87.9	85.0	93.2	**99,0**
5	2Ps-Lc1	83.5	86.0	81.2	82.6	78.6	75.6	90.5	**93,5**
6	2Ps-Lc2	76.5	84.3	77.0	81.1	79.6	78.3	92.0	**92,6**
7	Ps1-2Lc	85.5	85.0	80.3	82.6	78.5	79.6	90.0	**92,8**
8	Ps2- 2Lc	88.4	89.8	81.5	81.6	77.5	78.9	88.0	**91,2**
9	2Ps- 2Lc	75.7	79.3	71.0	73.1	77.0	71.5	80.1	**85,3**
Average		85.61	86.26	81.86	83.18	83.11	81.5	88.63	**92,03**

Through this comparison, it can be concluded that: (1) the performance of the classical approaches depends to a large extent on the optimization of each of the stages (extraction, reduction and classification), which cannot be optimized at the same time. (2) The proposed method shows superior performance due to two reasons, first is that feature learning, extraction and reduction are produced simultaneously, second, it is capable of adaptively learning the essential features of data from different sources of information.

3.3.2 Model performance discussion

AI-based models are often referred to as "black box" models, especially when referring to deep-learning-based models. This is because it is often difficult to understand how models make predictions. There is a concern in research about the use of DL-based models, as there is a risk that the high performance obtained through these models could be the result of inadequate learning. In this regard, in the proposed methodology an intuitive study is carried out with the objective of analyzing that the learning of the model is aligned with characteristic patterns of the condition of the rotating machinery and not with characteristics external to the environment, or that are outsider to the state of machine operation. This analysis attend as feedback for leads to the generation of more reliable diagnostic models and that these are not only focused on obtaining good performance. This contributes to diagnostic models being more reliable, generalizable and the risk of overfitting is eliminated. The intuitive study consists of verifying that the signals that feed the deep-learning models are correctly characterized. In addition, confirming that the model learns intrinsic patterns of the signals (important for diagnosing the machine) and that it does not learn outsider patterns or irrelevant to the diagnosis. As mentioned above, a deep-learning model based on stacked-autoencoders is used, which allows validating that the input data is being effectively reconstructed at the output through the measurement of the reconstruction error. In this regard, and to give an idea of the high characterization effectiveness of the proposed methodology, **Fig 3.3.5** presents the characterizations of some signals through the SAE model.



(c)

Fig. 3.3.5 Representation of real signals and the respective reconstruction obtained through the SAE-based for healthy condition model. (a) current signal in TD; (b) vibration signal in FD; (c) one IMF of vibration signal in TFD.

The signals presented correspond to the training result of the model for the conditions of index 2 described in **Table 3.3.3**, in **Fig 3.3.5** (a) presents a TD current signal and its corresponding characterization. **Fig 3.3.5** (b) shows a vibration signal in FD and finally in **Fig 3.3.5** (c) an IMD corresponding to a vibration signal in TFD is shown. All the signals shown refer to healthy states and each one shows its corresponding reconstruction. It can be noted that the real signals and the reconstructed signals are closely similar. In quantifiable terms, the average reconstruction error that presents in these signals is considerably low. For the entire data set, the resulting mean square error is 0.014 in the training stage and 0.019 for the validation stage, with the minimum error being obtained from 0.001, and the maximum error obtained is 0.026. The correct characterization of the signals must be followed to achieve a correct diagnosis of the health condition of the system. To accomplish, verifying the learning process must be a stage to be included in the monitoring schemes. In addition, with a correct study of the learned parameters, more interpretable and robust schemes are generated. Therefore, the proposed methodology includes an interpretability testing stage to verify that the relevant parameters in the measurements are learned and the outsider parameters are not included in the learning. The validation and interpretability test stage that is included in the monitoring scheme is shown in **Fig. 3.3.6**.

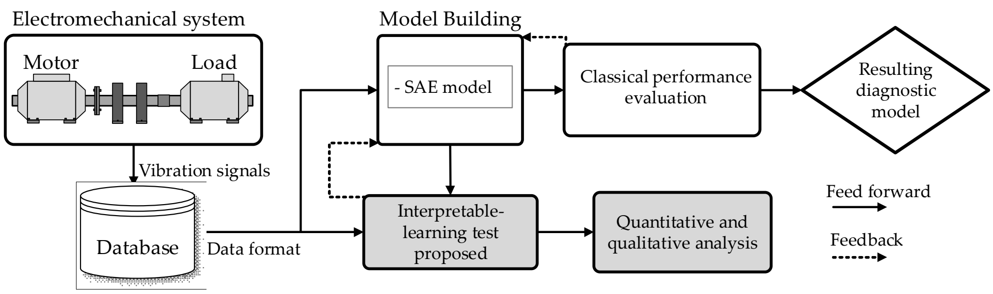


Fig. 3.3.6 Procedure of the proposed diagnostic methodology based on DL models with the addition of the interpretable-learning analysis step.

This stage of interpretable-learning consists, first, a parameter of relevance is analysed, that is, the root mean square (RMS), which the model is expected to learn and characterize appropriately. Second, a parameter that is not relevant and therefore it is not desirable to be characterized, that is, the noise. A common time domain fault detection technique for basic motor systems is to calculate the RMS value of the acquisition signals. In this technique, when the RMS residual value exceeds a predefined threshold, a fault indication signal is generated. Therefore, it has been shown in various investigations that the RMS value is an intrinsic parameter associated with the condition of the machine. Besides, there are parameters present in the system that are undesirable to be characterized since they do not provide descriptive information for the diagnosis. In this regard, noise is one of these unwanted parameters, and it is desirable that a diagnostic approach does not characterize the noise variations. The experiment consists of characterizing the range of RMS values for the health and fault conditions of the signals and then inducing variations.

For the current signals, the maximum RMS value is 1.83 and the minimum value is 1.43. The mean RMS value for the healthy condition is 1.71, the average value for the bearing fault is 1.46, for the demagnetization fault is 1.77, for eccentricity fault is 1.48, and the for gearbox fault is 1.69. For vibration signals, the RMS range is between 0.28 and 0.13 as maximum and minimum, respectively; with 0.14 being the average value for healthy condition, for bearing fault is 0.25, for demagnetization fault is 0.23, for eccentricity fault is 0.15 and for gearbox fault is 0.15. In contrast, white Gaussian noise was added to the time signal vector in SNR of 10 to 100 db. In **Fig 3.3.7** shows a case of a signal without variations and the same signal with induced variations of both RMS and Gaussian noise, each with the respective reconstruction obtained from the DNN model.

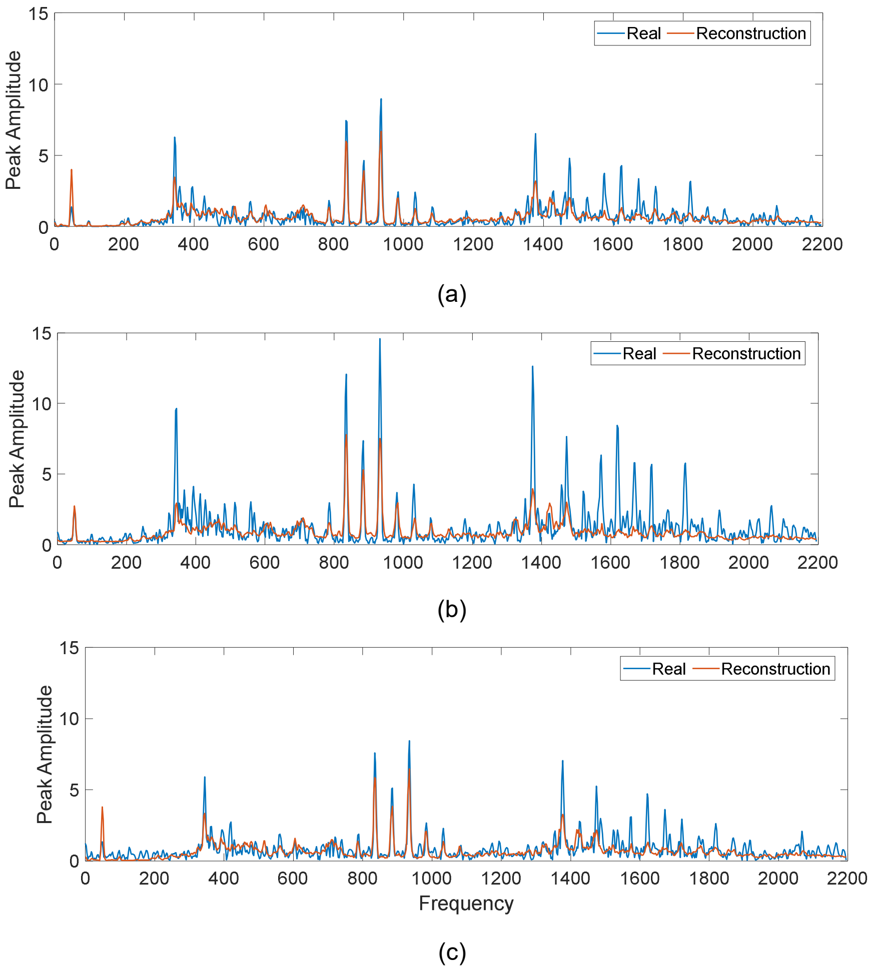


Fig. 3.3.7 Characterization of vibration signals in FD: (a) healthy condition, (b) healthy condition with RMS added (c) healthy condition with white Gaussian noise added.

In **Fig 3.3.7** (a) presents a sample vibration signal in frequency domain without variations with its respective reconstruction. In **Fig. 3.3.7** (b) the same vibration signal is presented with the modified RMS value, the RMS of the signal is increased, this variation causes an increase of the harmonics for the frequency spectrum, as can be seen in the figure. The vibration signal with noise variations is shown **Fig 3.3.7** (c), the noise of the original signal was increased which causes an increase in amplitude at low frequency values.

Regarding the reconstructions, for the healthy signal without variations an MSE reconstruction error of 0.0097 is obtained, for the reconstruction of the healthy condition with the modified RMS a MSE of 0.0120 and for the reconstruction of the healthy condition with noise added a MSE of 0.0110. These MSE values are aligned within the range of values obtained during the training process of the DNN model.

In order to quantify the results of the interpretability stage, the results are presented in terms of accuracy. Therefore, for the healthy samples with RMS variation, only 15% accuracy of the samples are classified as healthy, the rest of the samples are classified as a type of fault. For healthy samples with noise, the samples are classified as 97% healthy samples. This confirms that under the proposed deep-learning approach, the relevant patterns are correctly learned and characterized, such is the case of RMS. In other words, the model shows to be sensitive to RMS changes, a parameter that in other applications allows to characterize the condition of the system. On the contrary, the model presents little sensitivity to the increase of noise in the signals, maintaining a correct diagnosis.

In the case of the healthy condition with RMS variation (He+RMS), the RMS was modified to an average value of 0.16 in the time signal, while, for the healthy condition with noise added (He+Noise), white Gaussian noise was added with 50 dB SNR. **Fig 3.3.8** (a) shows the projection of all the conditions of the electromechanical system including the He+RMS samples. And in **Fig 3.3.8** (b) shows the projections of all the conditions, including the He+noise samples. When generating the feature mapping, the samples corresponding to He+RMS extracted from the learning model have a different dispersion than the healthy condition and is more similar to the condition whose RMS value is similar. On the contrary, in the mapping of the He+Noise samples, they are kept in the same cluster as those of the healthy condition.

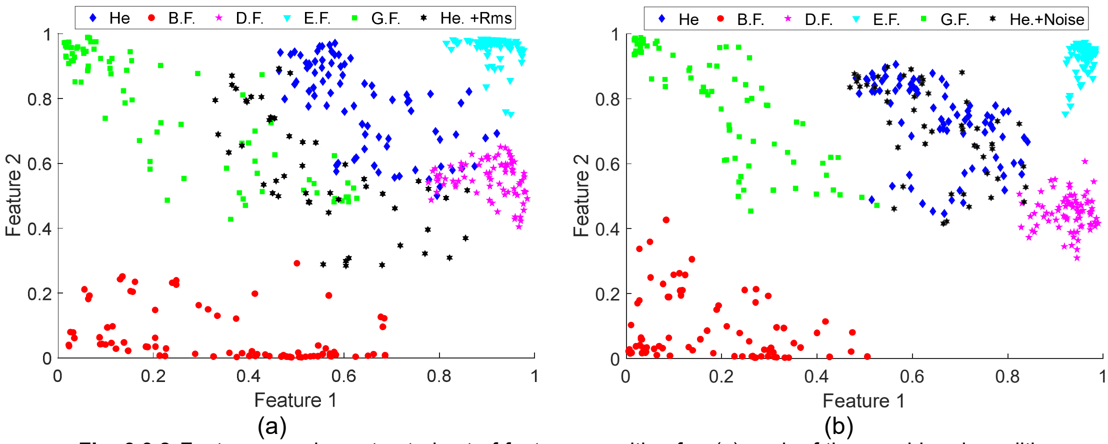


Fig. 3.3.8 Feature mapping extracted set of features resulting for: (a) each of the considered conditions and the healthy condition with a modified RMS (He+RMS), (b) each of the considered conditions and the healthy condition with added noise (He+Noise).

3.4 Discussion and conclusions

A methodological process for the implementation of a diagnostic model for electromechanical systems based on learning features by deep-learning approach is presented in this chapter. The proposed method consists of three main aspects. The first, the learning of multi-sensor information features. The functionality of using multiple sources of information allows to better characterize the industrial system and different fault conditions and to better capture the variability of system operation. The second, the generation and optimization of a model based on deep learning for fault diagnosis. An optimizer based on genetic algorithm is implemented to obtain hyper-parameters of a stacked-autoencoder architecture in order to build a model capable of obtaining the best possible characterization for considered signals. The third, the learning interpretability of the diagnostic model. Adding a learning interpretability review stage to the diagnostic model not only implies increasing diagnostic accuracy but also generating more sophisticated, generalizable, and reliable schemes.

Condition monitoring schemes must face the difficulty of adapting to the increase in variations in the operating condition of industrial systems, as well as the presence of multiple failures in the different components. With the proposed methodology has been shown to facilitate the implementation of diagnostic schemes based on artificial neural networks, improving performance and enhancing the ability to identify faults even in highly complex environments. Regarding this, algorithms based on deep-learning, such as autoencoders, have shown a significant capacity to correctly characterize multi-pattern environments, learning the intrinsic modes of the signals, which facilitates subsequent classification, becoming a powerful tool for condition monitoring implementations.

Finally, it should be notice that, dealing with industrial applicability requirements, the condition monitoring schemes must fully comply with the requirements that are presented. In this regard, presenting a correct diagnosis of the condition of the system is the most important aspect, however, dealing with such an important environment, such as supplying mechanical energy to industrial systems, maintaining its proper functioning also requires schemes that present security and reliability. For this reason, the construction of monitoring schemes should not only include the adjustment of hyperparameters to achieve the best diagnostic performance, but also to introspect that the diagnosis is being carried out properly, and that the inferences made by the intelligence model artificial are based on learning patterns that describe the functioning of the system condition. In this way, the feature learning models are prevented from falling into overfitting or characterizing non-relevant modes for diagnosis.

4.

Anomaly Detection for Electromechanical Industrial Systems

Automatic detection of unknown events in industrial systems represent a challenge, especially in complex manufacturing environments. Anomaly detection, as this task has been commonly called in monitoring applications, represents e the way to meet the demands within the smart manufacturing environment. This chapter introduces a new methodology to improved anomaly detection performance based on a deep learning approach.

CONTENTS:

4. Anomaly Detection for Electromechanical Industrial Systems

4.1 Introduction

In this chapter the functionalities that algorithms based on deep-learning can provide in the field of anomaly detection in rotating systems are reviewed. As a result of this analysis, a novel methodological process is proposed for the implementation of a detection scheme using the benefits provided by deep-learning tools. The background and motivation are presented as follows, followed by reviews of the state of the art around anomaly detection for machine monitoring and describes the contribution methodology of this work.

4.1.1 Background and motivation

The increase in complexity in industrial rotating systems has caused that predictive maintenance strategies have to be constantly modernized. The complexity and dynamic operating conditions of the machinery, such as changes in the rotation speed or increase in load, produce changes in the patterns of the monitoring signals, which has repercussions when making a diagnosis or detecting a fault in the system. Deep learning (DL) techniques have recently been applied in monitoring schemes, especially for the excellent pattern management capability. Note that the term pattern management is considered as the ability to deal with a large number of features and the ability to react to the changes present in them. By the implementation of DL, it is possible to extract the complex relationships from the monitoring data through the use of neural networks with multiple layers and non-linear transformations [119]. For this reason, a great number of research have emerged around the condition monitoring with diagnosis schemes that make use of deep-learning. Despite the fact that most DL-based research are successful in pattern management and diagnosis, that is, fault classification, there is an important issue that has not been clearly solved in regard to how to consider the detection of unseen patterns. Therefore, the detection of unseen patterns is a topic that attracts the attention of researchers in the field of pattern recognition, and especially in the field of health condition prognosis of industrial systems. It is a real problem, that when not taken into account during the condition assessment, it has a negative impact, especially to early decision-making and an incorrect diagnosis that can lead to considerable economic losses and catastrophic faults.

Due to in a real industrial environment only the normal (healthy) class is available as initial knowledge, it is necessary to adapt approaches that consider the use of one-class classifiers. One-class classification (OCC) problems are sometimes referred to as anomaly detectors, novelty, or outliers [120-122]. These are trained with known patterns that are organized as one or more groups (in feature space) of normal classes.

The model is then used to identify unknown patterns such as novelties or anomalies [123], which are somewhat different from those present in the original training data set. However, OCC-based anomaly detection approaches still struggle to achieve optimal performance. Solving a bound optimization problem is often required, however, this optimization depends largely on the representation of the data. Therefore, the implementation of methods that improve the representation capacity is required to improve the performance of anomaly detection methodologies.

4.1.2 State of art in anomaly detection

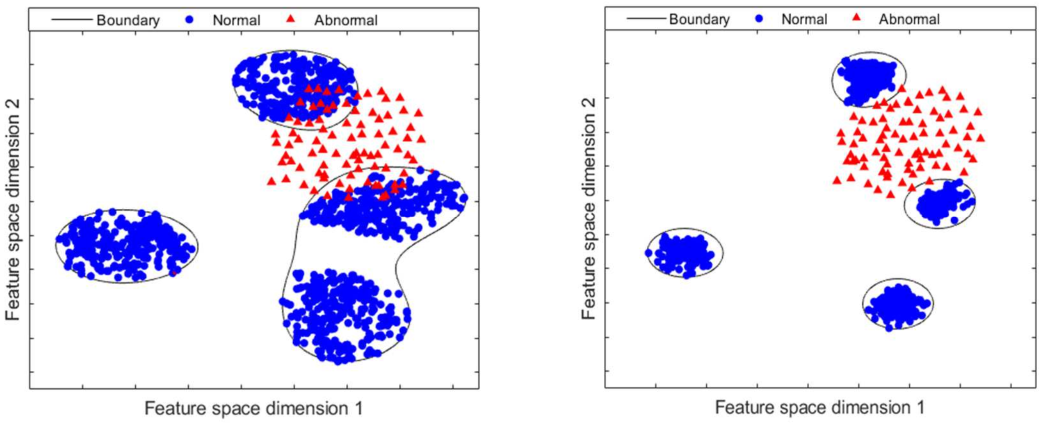
Anomaly detection in manufacturing processes is one of the main concerns in the new era of the Industry 4.0 framework. The detection of uncharacterized events represents a major challenge within the operation monitoring of electrical rotatory machinery. The most appropriate way to deal with the detection of unknown events is through approaches that are based on one-class classification (OCC). One of the main difficulties around OCC problems is that detection schemes must learn complex representations of the data. In addition, the detection schemes must face the challenge of training the models under an unsupervised or semi-supervised environment, since only the labeled data of the normal condition are available. A fact that it is a challenging task to separate the known and unknown conditions classes in an unsupervised manner.

There are different methods to address OCC issues. The first is based on estimating the generative probability density function (PDF) of the data, such as the Gaussian mixture model (GMM) [124]. Second, distance-based methods, which are based on grouping all those data that are close as known information, for example, the nearest neighbors [125]. Third, the methods are based on a reconstruction model. autoencoders (AE) are examples of this type of approach [126]. Finally, domain-based methods impose a boundary that will be created based on the structure of the training data set. In this case, a bound optimization problem is solved to represent the data. The class membership of the failing data is then determined by its location relative to the boundary. One-class support vector machine (OC-SVM) and support vector data description (SVDD) are the most popular methods [127].

OC-SVM is a method that has been widely applied to address detection problems in different research areas, due to the simplicity of configuration and the high performance presented against a diversity of data. However, the performance of OC-SVM, and the majority of methods that are used around anomaly detection problems, is that they are highly dependent on the representation of the data. Especially for real-world problems, regardless of the approach, the distribution and behavior of the characterized patterns represent a challenge for the execution of anomaly detection schemes. Generally, when carrying out the representation of the characterization of patterns, it can be noticed that in industrial operation environments, the behavior of this patterns tends to have a sparse representation, that is, the distribution of the data is dispersed and more than one single cluster can be presented.

In this regard, different strategies have emerged that contribute to improving the representation of the data, essentially seeking to generate more compact representations. Lately, deep-learning algorithms, such as the convolutional neural networks (CNN) and deep-autoencoders (DAE) that have better representation abilities have been widely applied in many research to improve clustering tasks, that is, to achieve more compact representations of data [128-129]. These types of clustering approaches based on deep learning have been referred to as deep clustering. In [130] a feature transformation model using a DNN to generate a data representation is proposed. In [131], a model for clustering based on deep-learning called graph clustering is presented. In [132] is proposed deep embedded clustering (DEC), a clustering algorithm that aims at finding learn representations and cluster assignments using deep neural networks. DEC makes use of deep-learning tools to reinforce the compactness of the data representation and increasing clusters separability.

A graphical way of how a compact feature mapping can improve the detection of anomalous samples is shown in **Fig. 4.1.1**. In **Fig. 4.1.1** (a) shows the mapping of features with scattered clusters, when mapping the abnormal samples, they overlap with some of the normal samples. In addition, the boundary that encloses the normal samples is wider. In contrast in **Fig. 4.1.1** (b), a compact feature space is shown, the boundary that encloses the normal samples is also tighter and therefore there is a higher probability that the anomalous samples are outside the space characterized as known.



(**a**) (**b**)

Fig. 4.1.1 Differences between a feature mapping in an anomaly detection environment. (a) mapper with sparse clusters; (b) mapping with compact clusters

4.1.3 Methodological proposal

In the development of this chapter, a novel methodology for anomaly detection for electromechanical industrial systems is presented. The proposed methodology provides a step-by-step process considering the application of state-of-the-art tools within the field of artificial intelligence and condition monitoring. The detection methodology includes a hybrid scheme based on stacked-autoencoders with which two anomaly scores are obtained that are combined to generate a more robust and reliable assessment.

The main contributions of this research work include:

- The proposal of a novel methodological process to carry out the detection of anomalies applied to industrial electromechanical systems.
- The proposal consists of a hybrid scheme that combines the ability to learn features of a deep-learning method based on data reconstruction and an OCC scheme applied to a compact data representation mapping.
- A deep embedding clustering algorithm is extended and adapted to the OCC context. Through the encoder part of a deep-encoder, a compact representation is obtained which is used for the OC-SVM application, with which the data provided under multiple working conditions of an electromechanical system can be managed correctly to carry out the detection of new faults.

It must be noted that this work represents an advance to deal with monitoring environments in which fault modes have not been characterized, therefore, the methodological process follows a non-supervised manner. Moreover, the proposed method is validated in two case studies on electromechanical environments and the results are compared with classical anomaly detection schemes.

4.2 Theoretical approach and proposed contribution

A step-by-step diagram of the proposed anomaly detection methodology is shown in **Fig 4.2.1**, which is divided into two main stages: the estimate of anomaly membership obtained by a clustering-based analysis and the estimate of anomaly membership obtained through a reconstruction model.

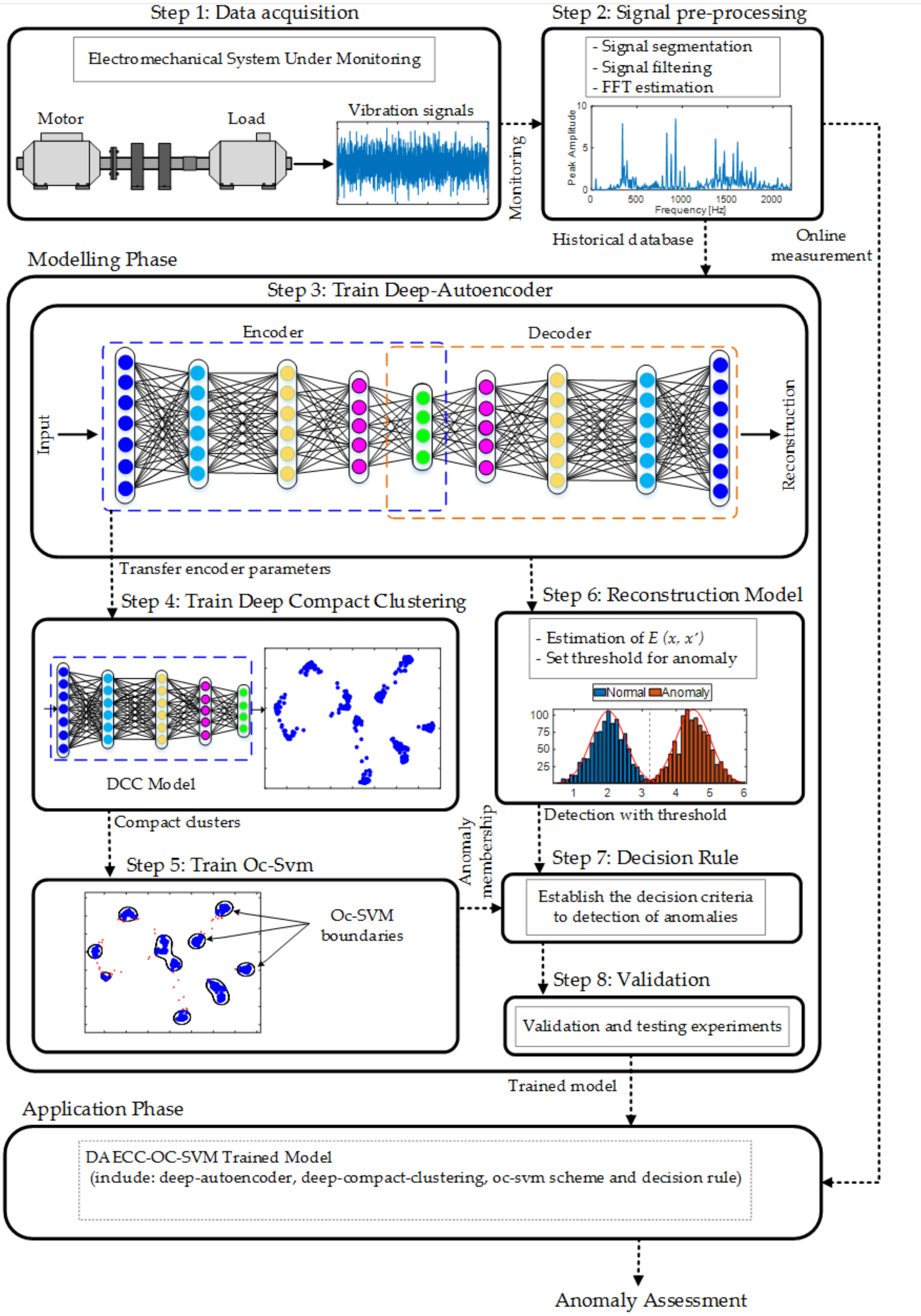


Fig. 4.2.1 Framework of the proposed anomaly detection methodology. This hybrid scheme consists of a compact clustering strategy and a reconstruction model.

The proposed anomaly detection framework can be divided into the following two phases: modeling phase, that is, the offline procedure and, application phase that refers to the online procedure. The modeling phase extracts available monitoring data (i.e. historical, ad-hoc acquisitions, etc.), to perform data-driven training in which the optimization of the models' hyperparameters that compose the developed methodology is carried out in an off-line mode. Once the detection methodology has been trained and validated, it is equipped to be integrated into online operation to perform monitoring of electromechanical systems. The proposed detection methodology is considered for continuous monitoring, therefore, continuously, for each sample collected, the method generates an evaluation of whether the new measurement corresponds to a known condition. Continuous evaluation provides criteria for making a maintenance decision. That is, when the new sample is known, the complementary diagnostic systems (referring to chapter 3 of this thesis) can be executed reliably. If the sample is unknown, the monitored system is operating in conditions different from those characterized and, therefore, no more information can be provided for maintenance decision-making than the operating anomaly itself. The proposed detection methodology consists of seven key stages, of which data acquisition and signal pre-processing correspond to the two phases. Train deep-autoencoder, train deep-compact-clustering, train oc-svm scheme, reconstruction model, decision rule and validation are part of the modeling phase. Once these steps are completed, the online process is equipped to execute and to evaluate the new measurements of the system under monitoring. All these steps are detailed below.

4.2.1 Step 1: Data acquisition

The first stage of the proposed detection methodology is related to the collection of signals associated to the condition of the electromechanical system. For this aim, three considerations need to be addressed. First, the acquisition of physical magnitudes, second, the details of the acquisitions and finally, the conditions of the electromechanical system. Regarding the acquisition of physical magnitudes, the proposed methodology focuses on the acquisition of vibration signals from multiple points of the electromechanical system. However, the proposed methodology allows adapting to the use of other sources of information, such as stator current, acoustic emission, speed, among others. In the proposed methodology, one-second signal segments are used to perform the characterization. In this regard, considering static behaviors in this period of time, this acquisition time ensures sufficient statistical consistency in most practical applications (i.e., rotation speed greater than 500 rpm). However, the time period of each acquisition sample can be increased to achieve higher resolution or to avoid information loss for slow speed applications. Regarding the conditions of the electromechanical system, this methodology is designed to begin the analysis from the data corresponding to the healthy operating condition, including the available torque and speed variants.

Then, as long as the acquisition details are the same, new acquisitions corresponding to different operational states (*i.e.* faults and degradation levels) are added to the analysis. As a result of this stage, there are acquisitions of raw vibration signals for different conditions.

4.2.2 Step 2: Signal pre-processing

Next, the vibration signal is segmented by time windows of equal length in order to generate the database of consecutive samples. The signal segmentation process by windows is expressed as follows:

$$\mathbf{X}_{vib} = \left[\mathbf{X}_{vib}^{1:L}, X_{vib}^{L+1:2L}, \ldots, \mathbf{X}_{vib}^{(n/h-1)h+1:n}\right] \tag{39}$$

where L denotes to the length of the time window used for segmentation and n denotes the sampling number. With the segmentation process, the vibration signal is divided into n/L segments. In addition, since the field of application is an industrial environment, the generation of significant noise levels occurs due to the inherent operation of rotary systems; as a consequence, the nature of the acquired signal is affected, requiring the use of filtering stages. Consequently, this proposal considers the implementation of a pre-processing stage that aims to filter a specific range of frequencies of interest. Specifically, in the study of electromechanical systems, harmonics occur at low frequencies, which can cause them to overlap with fault-related frequency components. To avoid this, consider the application of a digital FIR-based low pass filter that is implemented in software, and it is designed under the window method [133]. The digital low pass filter based on the windowing method works the under basic principle of the ideal “brick wall” filter with a cut-off frequency of ω_0 (rad/sec) following:

$$h(n) = \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{j\omega n}\, d\omega \tag{40}$$

where, the low pass filter produces a magnitude equal to one for all frequencies that have frequency values less than ω_0, and produces a magnitude equal to 0 for those frequency values between ω_0 and π. Thus, its impulse response sequence is depicted by $h(n)$. The main objective of this filtering process is to obtain the vibration signal in a suitable form for further processing. In this regard, each segmented part of the signals $\boldsymbol{X}_{vib}$ have been individually subjected to the low pass filtering stage with a cut-off frequency equal to 1500 Hz. After the filtering process, a transformation of the raw signal is performed to obtain the spectra of the signals, using the fast Fourier transform (FFT). The frequency coefficients are then scaled and used as the final inputs to the model. Since the frequency amplitude is too small to cause the change of network weights, the samples are multiplied by a coefficient following the recommendations of [134]. Once these processes have been completed, the database is configured.

The dataset, it has been divided in three different parts: the first one composed of 60% of the available samples for training purposes, the second one composed of 20% of the samples for validation purposes, finally, the third one composed of 20% of the samples for test purposes. In addition, a five-fold cross-validation approach is applied over the dataset in order to corroborate that the results are statistically significant. In this step, the result is a database with segments of vibration signals filtered and transformed into frequency components.

4.2.3 Step 3: Train deep-autoencoder

The architecture used for deep-autoencoder trained layer by layer is as described in Subsection **2.2.1.2**. The hyperparameters of the deep-autoencoder, such as the coefficient for the L2 regularization term, the coefficient for the dispersion regularization term, and the parameter for the dispersion ratio, as well as the number of neurons in each hidden layer, are established from the search for the optimal configuration using a genetic algorithm (GA). The optimization procedure by GA is performed as follows:

1. Population initialization: the chromosomes of the GA are initially defined with a logical vector containing five elements: each one of the three hyperparameters and the number of neurons in the two hidden layers. Afterwards, a randomly initialization of the population is performed by assigning a specific value to each particular parameter. The assigned values are in a range of established values that have been used in other applications around the monitoring of rotating systems. Once the initialization of the population is achieved the procedure continues in Step 2.
2. Population assessment: in this step is evaluated the fitness function which is based on the minimization of the reconstruction error between the input and the output features. Specifically, Equation (8) described in Subsection **2.2.1.1** is the optimization function of the GA. Therefore, the objective function of the GA is to achieve a minimum reconstruction error. Consequently, the optimization problem to be solved by the GA involves the search for those specific parameter values that lead to a high performance feature mapping. Then, once the whole population is evaluated under a wide range of values, the condition of best parameter values is analyzed, and the procedure continues in Step 4.
3. Mutation operation: GA mutation refers to a new population value by roulette wheel selection, the new generated population takes into account the choice of the best fitness value achieved by the previous tested population. In addition, a mutation operation is applied, which is based on the Gaussian distribution during the generation of the new population. Afterwards, the procedure continues in Step 2.

4. Stop criteria. There are two GA stopping criteria: (i) obtaining a reconstruction error value lower than a predefined threshold, 5%, and/or, (ii) reaching the maximum number of iterations, established at 1000 epochs. In the first case, the procedure is repeated iteratively until finding the optimal values of the parameters until the GA evolves, then the procedure continues in Step 3.

Following the configuration of the deep-autoencoder architecture, the input layer corresponds to the length of the FFT data obtained from the processing, which is used for training the network. The output layer is set to two to generate a two-dimensional feature space. The Adam optimization algorithm that is an extension to stochastic gradient descent, is used to optimize the loss function [135]. The weights are initialized using the Glorot uniform initializer, also called Xavier uniform initializer, which automatically determines the scale of initialization based on the number of input and output neurons [136]. The activation function that is established is a sigmoid function. The training process of a DAE is unsupervised, so it does not require label information from the input data. Due to this peculiarity, the final DAE structure can effectively reconstruct the FFT signals used during the training process, and at the same time, the bottleneck coding layer can generate a feature space that can be mapped into a two-dimensional space. The output of the encoder is established in two neurons so that the mapping is two-dimensional in order to generate a space of features that any user can interpret. These two DAE functionalities can be used in the anomaly detection framework. On the one hand, the ability to reconstruct the signals learned during training allows the reconstruction error to be used as a measure of anomaly as long as the signal differs from the learned ones. On the other hand, the generation of a feature space mapping, on which limits can be built and anomalous memberships can be established. Anomaly memberships depend on the implementation of OCC-based methods. By establishing two-dimensional projections, the anomaly membership can be generated as a boundary enclosing the known data, while the anomalies lie outside the membership or boundary. The result of this step is a model based on trained deep-autoencoders, capable of generating both a feature mapping and an effective reconstruction of the input signals.

4.2.4 Step 4: Train deep-compact clustering

Models based on deep-autoencoders have the ability to generate a feature space mapping at the output of the bottleneck encoding layer. However, the DAE encoder is often inefficient when applied directly to OCC problems, due to the bottleneck feature space mapping is sparse, *i.e.* not guaranteeing a compact mapping of the data in the bottleneck, which is an essential issue in OCC problems to obtain an optimal result [137].

Therefore, to improve the representation, a strategy is applied that improves the compactness of the clusters in the feature space. The deep-compact-clustering (DCC) introduced in this work follows the idea presented in [132]. DCC refers to a technique that seeks to increase the compactness of the clusters produced by the encoder bottleneck. The procedure to carry out is followed below.

First, the concept of clustering a set of n $\{x_i \in X\}_{i=1}^{n}$ points of a feature space into k clusters, each represented by an $\mu_j, j = 1, \dots, k$. centroid. Before clustering the data directly in the X space, a transformation of the data is performed to a non-linear mapping $f_\theta : X \to Z$ where θ are the learning parameters and Z is the new latent feature space. The f_θ parameters are initialized using the data mapped by the bottleneck encoder layers of a deep-autoencoder, discarding the decoder layers. DCC seeks to learn simultaneously a set of k cluster centers $\{\mu_i \in Z\}_{j=1}^{k}$ in the feature mapper Z and the hyperparameters $\theta = \{\boldsymbol{W}_e, b_e\}$ of the set of encoder layers. Starting from the mapping generated by the deep-autonecoder encoder and having the initial k cluster centers, the main idea is to iteratively perform the following two main steps: (1) measure a soft assignment between the embedded points and the cluster centroids and (2) calculate an auxiliary target distribution, based on learning the current high-confidence assignments, then, update the deep mapping θ and improve the cluster centroids. This is accomplished through an optimization process minimizing the Kullback-Leibler divergence loss between soft assignments q_{ij} and the auxiliary target distribution p_{ij}:

$$L = KL(P|Q) = \sum_i \sum_j p_{ij} \, log \frac{p_{ij}}{q_{ij}} \tag{41}$$

The soft assignment is defined as the probability of assigning sample i to cluster j, using the Student's t-distribution as a kernel to measure such similarity, according to [138].

$$q_{ij} = \frac{\left(1+||z_i-\mu_j||^2/\alpha\right)^{-\frac{\alpha+1}{\alpha}}}{\sum_{j'}\left(1+||z_i-\mu_{j'}||^2/\alpha\right)^{-\frac{\alpha+1}{\alpha}}} \tag{42}$$

where $z_i = f_\theta(x_i) \in Z$ correspond to $x_i \in X$ after embedding, α are the degrees of freedom of Student's t-distribution. The auxiliary target distribution is calculated using soft assignments as follows:

$$p_{ij} = \frac{q_{ij}^2/f_j}{\sum_{j'} q_{ij'}^2/f_{j'}} \tag{43}$$

where $f_j = \sum_{j'} q_{ij}$ are soft cluster frequencies. The selection of the target distributions P is the most relevant step to achieve compactness for each of the clusters. According to [132], this distribution must satisfy the following concepts: generate good predictions to improve cluster purity, give higher importance to data points assigned with high confidence, and normalize the loss contribution of each centroid to avoid that large clusters distort the embedded feature space.

The cluster centers μ_j and the deep-autoencoder parameters θ are jointly optimized using the Stochastic Gradient Descent method with momentum and standard backpropagation to calculate the parameter gradients, as follows:

$$\frac{\partial L}{\partial z_i} = \frac{\alpha+1}{\alpha} \sum_j \left(1 + \frac{||z_i - \mu_j||^2}{\alpha}\right)^{-1} \times \left(p_{ij} - q_{ij}\right)\left(z_i - \mu_j\right) \tag{44}$$

$$\frac{\partial L}{\partial \mu_j} = -\frac{\alpha+1}{\alpha} \sum_i \left(1 + \frac{||z_i - \mu_j||^2}{\alpha}\right)^{-1} \times \left(p_{ij} - q_{ij}\right)\left(z_i - \mu_j\right) \tag{45}$$

The optimization continues iteratively until the stopping criterion is reached.

For the application of DCC in the monitoring environment of electromechanical systems for the detection of anomalies, the following strategy is followed. First, the centers are initialized using the fuzzy C-means algorithm, k is a parameter to be defined by the user. Due to the different working conditions of rotary systems, the feature mapping of these produces multiple clusters. To deal with this, it is established that the number of centroids k corresponds to each of the known operating conditions (i.e., pairs of torque and speed setpoints considered during the acquisition step) of electromechanical system under study. For the training stage of the DCC algorithm, the Adam boost optimizer is used with a standard backpropagation procedure. A learning rate of 0.001 is established and the batch size to 200 is set. The degrees of freedom α of Student's t-distribution are set in one. Finally, the result of this stage is the generation of a feature space with compact clusters.

4.2.5 Step 5: Train Oc-Svm

The next step is to take the representation of compact clusters obtained in the previous step by the DCC algorithm and apply a classifier of a class to characterize the known states. For this purpose, the one-class support vector machine (oc-svm) algorithm is used. Oc-svm was proposed by Schölkopf [139], it is focused to estimate the support of a high-dimensional distribution. The oc-svm classification goal is focused on separating a class of samples, called the target class, from the rest of all other samples in the feature space.

For the implementation of oc-svm the following points are followed. To establish the hyperparameters of the oc-svm model, a five-fold cross-validation is implemented to set the best results. The RBF kernel is used in each of the experiments. After optimizing and defining the oc-svm hyperparameters, the model is trained with the electromechanical system database. The model is first trained on the healthy data set, subsequently fault conditions are added to the known data set. As a result, the oc-svm model finds a boundary that encloses the samples belonging to the known states, called class membership. A positive value of this membership indicates that the sample is within the limit and therefore it is considered as known, on the contrary, a negative value indicates that the sample is outside the hyperplane and consequently it is considered an anomaly. The result of this stage is the boundary on the compact cluster and the respective anomaly membership.

4.2.6 Step 6: Reconstruction model

The idea of the reconstruction model is to use the ability to learn and rebuild signals from the deep-autoencoder. The deep-autoencoder reconstruction error metric is used as an anomaly metric. In this regard, as described in Subsection 2.2.1, deep-autoencoder models are trained to learn optimal mapping of the input data, generate a feature space through the encoder bottleneck, and successfully reconstruct its input on output in the decoder layer. Learned data produces low reconstruction error *i.e.* Ω_{Mse}. Conversely, data never seen by the model generates a high error rate. In this way, the reconstruction error can be used as a metric to identify outliers or anomalies. Thus, given a test sample x_{Test}, this is detected as unknown when the magnitude of its reconstruction error is larger than a certain threshold δ:

$$\Omega_{Mse} = \|x_{Test} - y_{Test}\|^2 \ > \delta \tag{47}$$

For the anomaly score, a simple threshold δ95 corresponding to the 95% percentile of the training data distribution is established. It should be noted that the quality of the data, the different operating conditions of the system, and the level of fit to the training data need careful consideration, since deep-autoencoder-based models can also fit the unknown data if it is similar to the training data. If this occurs, the anomaly data reconstruction error metric may be as low as the nominal data error, which is an undesirable result. The result of this step is the second anomaly score based on the reconstruction error measurement.

4.2.7 Step 7: Decision rule

At first, the anomaly membership obtained from OC-SVM on the feature space mapping as well as the anomaly score obtained from the reconstruction process could be used as the classification result itself. However, both methods can fail to perform anomaly detection. On the one hand, for anomalous data, the mapping to the feature space is not optimized, despite the cluster compression process, data overlapping can occur on known samples. On the other hand, the measurement of the reconstruction error can fail to detect abnormal samples, as the data considered as anomalies can be fit to the reconstruction values of the nominal data and only those samples with a high error value are perceived outside the threshold and therefore as anomalies. In this regard, in the proposed methodology, anomaly detection is carried out by combining the two anomaly scores obtained, on the one hand using the feature rendering capability of the deep-autoencoder, then improving the quality of the clusters through the process of compacting the mapping space through DCC and finally the application of OC-SVM. On the other hand, detection is improved using the anomaly score obtained from the deep-auto-encoder reconstruction process. Thus, for a given sample X, it is classified as known if the OC-SVM anomaly membership is positive, meaning that the sample is in bounds. Instead, sample X is classified as abnormal if:

$$\Omega_{Mse} > \delta \ \ or \ AM_{oc-svm} < 0 \tag{48}$$

where AM_{oc-svm} is the anomaly membership from OC-SVM. Hence, the result of the detection of known samples is given by DCC + OC-SVM, while the identification of anomalies is given by the combination of DCC + OC-SVM and the measurement of Ω_{Mse} through deep-autoencoder.

4.2.8 Step 8: Validation

In order to evaluate the anomaly detection performance of the proposed methodology. the assessment is carried out with the validation and test databases. It has been decided to use true positive (*TP*) to represent the number of correctly classified known samples, false negative (*FN*) to represent the number of known samples misclassified as anomaly, true negative (*TN*) to represent correctly classified anomaly samples and false positive (*FP*) to represent the number of anomaly samples misclassified as known. The anomaly detection performance is evaluated considering the true positive rate (*TPR*), true negative rate (*TNR*) and the balanced accuracy that refer to recall-based metrics. In addition, precision-based metrics are taken into account, such as, positive predicted value (*PPV*) and negative predictive value (*NPV)*. Are defined below:

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP + FN} \tag{49}$$

$$True\ Negative\ Rate\ (TNR) = \frac{TN}{TN + FP} \tag{50}$$

$$Balanced\ Accuracy = {}^{(TPR + TNR)}\!/_{2} \tag{51}$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP + FP} \tag{52}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN + FN} \tag{53}$$

Therefore, the result of the TP values is given by the anomaly score obtained through OC-SVM, while the TN values are given by the combination of OC-SVM and the resulting samples above the reconstruction model threshold.

4.3 Experimental implementation and analysis of results

This section shows the implementation of the proposed methodology for the detection of anomalies, and discusses the results obtained using two experimental test benches described in Annex 1 and Annex 2. As previously mentioned, the objective is to develop an anomaly detection scheme based on deep-learning techniques applied to rotating industrial environments. Specifically, two schemes are combined to obtain a more robust anomaly membership. First, a clustering-based approach that improves the compactness of the clusters with the aim of generating a less dispersed mapping of the data and facilitating the characterization of the known states and later the detection of outliers. The second is about a scheme based on the reconstruction of the data, a membership is obtained based on the reconstructed output of the model.

4.3.1 Evaluation Scenarios

In order to evaluate the efficiency and adaptability of the anomaly detection scheme to different electromechanical systems, two different experimental test benches have been considered. First, an electromechanical system considering faults in different components with signals acquired in four operating conditions. Second, a test bench focused only on bearing faults, the available acquisitions consider various severities of bearing failures and different load conditions.

4.3.1.1 Multi-Fault Test Bench

The test bench used to evaluate the detection methodology is described in Annex 1. To carry out the assessment, four system fault conditions are considered and healthy condition. The faulty data considered are: a partially demagnetized motor (Df) was developed during the fabricate with a 50% of nominal flux reduction in one pair of poles. Fault in bearings (Bf); the inner and outer races of the non-terminal bearing have been scraped thoroughly in order to cause a generalized rough defect. A static eccentricity (Ef) was induced through a screw attachment in the gearbox output shaft. And slight degradation is generated on two gear teeth to impose a degree of smoothed on the reduction gearbox (Gf). The measurements were collected at different operating conditions corresponding to power frequency low, power frequency high (30 and 60 Hz), motor load low and motor load high (40 and 75% of the nominal load). Therefore, there are four resulting operating conditions: power frequency low—motor load low (C1), power frequency low—motor load high (C2), power frequency high—motor load low (C3) and power frequency high—motor load high (C4). For the healthy condition and each of the fault conditions, there are 200 segmented samples based on the acquisitions perform.

A total of fifteen test scenarios are considered for the evaluation of the detection scheme for this test bench. The distribution of the classes for each scenario is presented in **Table 4.3.1**. The analysis starts first considering only the healthy condition (S1), then fault conditions are added to the training set that correspond to the known conditions.

Table. 4.3.1 Experimental set for each training and testing scenario.

Label	Training Set	Testing Set	
		Known Set	Unknown Set
S1	He	He	Bf, Df, Ef, Gf
S2	He, Bf	He, Bf	Df, Ef, Gf
S3	He, Df	He, Df	Bf, Ef, Gf
S4	He, Ef	He, Ef	Bf, Df, Gf
S5	He, Gf	He, Gf	Bf, Df, Ef
S6	He, Bf, Df	He, Bf, Df	Ef, Gf
S7	He, Bf, Ef	He, Bf, Ef	Df, Gf
S8	He, Bf, Gf	He, Bf, Gf	Df, Ef
S9	He, Df , Ef	He, Df , Ef	Bf, Gf
S10	He, Df , Gf	He, Df , Gf	Bf, Ef
S11	He, Ef , Gf	He, Ef , Gf	Bf, Df
S12	He, Bf, Df, Ef	He, Bf, Df, Ef	Gf
S13	He, Bf, Df, Gf	He, Bf, Df, Gf	Ef
S14	He, Bf, Ef, Gf	He, Bf, Ef, Gf	Df
S15	He, Df, Ef, Gf	He, Df, Ef, Gf	Bf

4.3.1.2 Rolling Bearing Faults Test Bench

The test bench for bearing failure analysis is described in Annex 2. The acquired fault conditions are: single point fault in the ball (FB), the inner race (FI) and the outer race (FO), in addition to the healthy condition (HE). While for each of the faults, three severities corresponding to 0.007 inches, 0.014 inches and 0.021 inches are considered. Moreover, the signals were collected at different operating conditions corresponding to various motor loads *i.e*. 0, 1, 2, and 3 hp. All signal acquisitions were made with the sampling frequency of 12 kHz.

Under the operating conditions of the rolling bearing test bench, seven test scenarios are considered to evaluate the anomaly detection methodology. The distribution of conditions for each of the test scenarios is presented in Table 4.3.2.

Table. 4.3.2 Experimental set for each training and testing scenario for rolling bearing test bench.

Label	Training Set	Testing Set	
		Known Set	Unknown Set
SS1	HE	HE	FB, FI, FO
SS2	HE, FB	HE, FB	FI, FO
SS3	HE, FI	HE, FI	FB, FO
SS4	HE, FO	HE, FO	FB, FI
SS5	HE, FB, FI	HE, FB, FI	FO
SS6	HE, FB, FO	HE, FB, FO	FI
SS7	HE, FI, FO	HE, FI, FO	FB

The four classes are grouped into three sets: training set, known set, and unknown set. Each one of the scenarios corresponds to a stage of advancement of the detection methodology, from an initial knowledge of only the healthy condition (HE) with the four operating conditions, to a scenario where the information of three states, HE and two faults states. These scenarios simulate the application of the anomaly detection methodology in a real industrial setting, where initially only the healthy condition is available and progressively new fault states are detected and incorporated.

4.3.2 Application procedure

The four classes are grouped into three sets: training set, known set, and unknown set. Each one of the scenarios corresponds to a stage of advancement of the detection methodology, from an initial knowledge of only the healthy condition (HE) with the four operating conditions, to a scenario where the information of three states, HE and two faults states. These scenarios simulate the application of the anomaly detection methodology in a real industrial setting, where initially only the healthy condition is available and progressively new fault states are detected and incorporated.

Previous to the application of the anomaly detection methodology, a series of procedures must be carried out on the databases of electromechanical systems:

1) First, the data sets are divided into three subsets: data training, data validation and data test. Second, data is randomly selected for each of the subsets and a 5-fold cross-validation approach is applied to the data to generate statistical variation on the results.
2) The Fast Fourier Transform (FFT) is applied to each of the vibration signal samples to obtain the corresponding frequency amplitude. The frequency amplitude of the samples is scaled following the recommendations of [134].
3) For the multi-fault experimental test bench, the vibration data corresponding to the plane perpendicular to the rotation of the motor are used. Whereas for the bearing dataset, the frequency amplitude of only one side is taken. Therefore, the dimension of each of the input samples for the multi-fault test bench has an input length of 2048 and for the bearing dataset it is 1024.
4) Finally, three auto-encoders are used to build the deep-neural-network model. The hyperparameters of the DNN model, such as the coefficient for the L2 regularization term, the coefficient for the sparsity regularization term, the parameter for sparsity proportion and the unit numbers of the hidden layers are obtained through the GA optimization process, following the procedure described in Subsection 4.2.3. For the multi-fault test bench, the architecture of the deep-autoencoder is d-850-120-2, for the encoder part, where d is the dimension of input data. The decoder is a mirror of the encoder with dimensions 2-120-850-d. While the encoder architecture of DNN for the rolling bearing dataset is d-500-100-2, also d is the dimension of input data corresponding to this dataset.

4.3.3 Performance Results

In order to verify the effectiveness and performance of the proposed detection methodology, the obtained results are compared with three representative unsupervised anomaly detection methods. First, a reference method and also two simple variants of this anomaly detection proposal:

1) Reconstruction model based on deep-autoencoder.
2) Variant method 1: deep-autoencoder + oc-svm (without DCC).
3) Variant method 2: deep-compact clustering + oc-svm (without reconstruction method).

(1) is the reconstruction-based anomaly detection method as described in subsection 4.2.6. In this method, the threshold is used as a metric to identify anomalies. This method has been successfully applied in other research fields. (2) refers to the detection scheme that integrates the deep-autoencoder model without applying the feature space compaction process, then the anomaly detection method based on a support vector machine of a class is applied. (3) is a simplified version of the proposed anomaly method but without the integration of the reconstruction model (only one anomaly membership is computed).

4.3.3.1 Performance evaluation of multi-fault experimental test bench

After establishing the parameters and training each of the stages of the proposed methodology, as well as the reference methods for comparison, each one is evaluated on the validation data set that represents 20% of the database. The results of the performance indicators are shown in Table **4.3.3** and Table **4.3.4**. Each of the anomaly detection methods are presented, first the reconstruction model based solely on deep-autoencoder (DAE), then the mapping model of the DAE combined with oc-svm (DAE + oc-svm), then the model cluster compaction and oc-svm without including the reconstruction model (DCC + oc-svm) and finally the proposed method. Performances for each of the detection methods are shown in Table. 4.3.3 in terms of recall-based metrics such as TPR and TNR for each of the experimental test scenarios. In this sense, it can be observed that in the case of TPR, that is, the known samples, the best performance obtained is from the DAE method, since it is superior in thirteen of the fifteen scenarios (S1, S3–S14). However, regarding the detection capabilities, the DAE method is quite deficient, being the worst detection method, obtaining superior results only in one scenario (S15) and having a low performance in most of the scenarios. Similarly, the DAE + oc-svm detection method is superior in one case in TPR (S2), but has poor results in anomaly detection (TNR). Most TPR performance is acceptable, while TNR performance in most cases is less than 0.50, a low detection range.

In the case of the DCC + oc-svm detection method, it can be noted that with the compaction of the cluster space through DCC, anomaly detection improves considerably. As can be seen in Table 4.3.3, the TNR values increase relative to the uncompacted method. In fact, the results of the detection performance referring to the TNR values of the proposed methodology are superior in all test scenarios compared to the other methods. Therefore, by combining the DAE's reconstruction capabilities and compact representation capabilities, the ability to capture known information and detect anomalous conditions is greatly improved.

Table. 4.3.3 Performance of the methods for anomaly detection in terms of TPR and TNR on multi-faults test bench.

Label	DAE		DAE + OC-SVM		DCC + OC-SVM		Proposed method	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
S1	**0.982**	0.607	0.891	0.409	0.936	0.662	0.936	**0.894**
S2	0.908	0.003	**0.931**	0.573	0.862	0.797	0.862	**0.801**
S3	**0.988**	0.672	0.888	0.145	0.908	0.542	0.908	**0.877**
S4	**0.933**	0.770	0.911	0.138	0.873	0.677	0.873	**0.922**
S5	**0.983**	0.370	0.892	0.402	0.869	0.558	0.869	**0.874**
S6	**0.922**	0.020	0.917	0.459	0.852	**0.797**	0.852	**0.797**
S7	**0.920**	0.013	0.909	0.278	0.870	**0.811**	0.870	**0.811**
S8	**0.937**	0.000	0.910	0.496	0.916	**0.830**	0.916	**0.830**
S9	**0.967**	0.993	0.929	0.035	0.911	0.291	0.911	**0.997**
S10	**0.967**	0.500	0.930	0.17	0.934	0.435	0.934	**0.851**
S11	**0.949**	0.611	0.917	0.02	0.873	0.721	0.873	**0.887**
S12	**0.936**	0.030	0.924	0.322	0.879	0.800	0.879	**0.802**
S13	**0.938**	0.000	0.932	0.228	0.835	**0.822**	0.835	**0.822**
S14	**0.938**	0.000	0.923	0.262	0.860	**0.782**	0.860	**0.782**
S15	0.954	**1.000**	0.936	0.000	**0.973**	0.001	**0.973**	**1.000**

The balanced accuracy is presented below in **Table 4.3.4**. It can be noticed that the proposed methodology obtains the best performance than the other methods in each of the scenarios. It matches with the same precision in four scenarios (S6, S8, S13, S14) with the detection method of DCC + oc-svm. Only in the S9 test scenario, the reconstruction-based method DAE, is superior to the proposed methodology. In terms of percentage, the proposed methodology obtains an average of the fifteen scenarios of 87.70%, and is higher by 21.70%, 28.05% and 12.10%, compared to methods (1), (2) and (3), correspondingly. As can be observed, anomaly detection approaches that consider a single detection membership usually perform well on a single parameter, which can be capturing known information or anomaly detection. Barely a detection method with a single membership performs well on both metrics. Therefore, combining the membership obtained by the reconstruction error and the oc-svm membership obtained from the compact feature space significantly improves the overall metrics.

Table. 4.3.4 Balanced precision of anomaly detection methods for 15 test scenarios.

Label	Balanced Accuracy			
	DAE	DAE + OC-SVM	DCC + OC-SVM	DAECC-OC-SVM
S1	0.795	0.648	0.799	**0.915**
S2	0.456	0.752	0.830	**0.831**
S3	0.830	0.516	0.725	**0.892**
S4	0.851	0.524	0.775	**0.897**
S5	0.677	0.647	0.713	**0.871**
S6	0.476	0.688	**0.835**	**0.835**
S7	0.469	0.593	0.805	**0.840**
S8	0.461	0.703	**0.873**	**0.873**
S9	**0.979**	0.482	0.601	0.954
S10	0.736	0.550	0.653	**0.892**
S11	0.783	0.468	0.765	**0.880**
S12	0.483	0.623	0.839	**0.840**
S13	0.469	0.580	**0.828**	**0.828**
SS4	0.469	0.592	**0.821**	**0.821**
S15	0.977	0.468	0.487	**0.986**
Average	0.660	0.597	0.756	**0.877**

To confirm this assumption, **Table 4.3.5** shows the results for the detection of anomalies based on precision, that is, PPV and NPV. It can be noted that the proposed methodology is superior to other anomaly detection methods. In terms of PPV, the proposed methodology obtains the best score in the fifteen test scenarios. While in terms of VAN, it is superior in twelve of the fifteen test scenarios. In percentage terms, the proposed methodology obtains an average of the fifteen scenarios of 94.60% for PPV and 73.80% for NPV. The proposal is 12.30%, 16.70% and 7.70% higher for the VPP and is 30.80%, 24.67% and 14.32% higher for the NPV, compared to the method (1), (2) and (3), correspondingly.

Table. 4.3.5 Performance of the methods for anomaly detection in terms of PPV and NPV on multi-faults test bench.

Label	DAE		DAE + OC-SVM		DCC + OC-SVM		Proposed method	
	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
S1	0.777	**0.947**	0.595	0.760	0.740	0.909	**0.907**	0.932
S2	0.646	0.000	0.857	0.821	**0.908**	**0.720**	**0.908**	**0.720**
S3	0.888	**0.860**	0.677	0.415	0.763	0.704	**0.934**	0.821
S4	0.909	**0.820**	0.682	0.461	0.834	0.721	**0.957**	0.784
S5	0.783	0.578	0.789	0.709	0.870	0.609	**0.981**	**0.745**
S6	0.740	0.101	0.869	0.704	**0.992**	**0.644**	0.922	**0.644**
S7	0.739	0.085	0.800	0.563	0.920	0.670	**0.922**	**0.672**
S8	0.734	0.000	0.836	0.636	**0.938**	**0.777**	**0.938**	**0.777**
S9	0.997	**0.968**	0.746	0.195	0.791	0.426	**0.998**	0.796
S10	0.872	0.457	0.792	0.541	0.811	0.334	**0.942**	**0.745**
S11	0.894	0.755	0.740	0.129	0.827	0.605	**0.940**	**0.763**
S12	0.790	0.028	0.845	0.516	0.954	0.633	**0.955**	**0.634**
S13	0.789	0.000	0.828	0.460	**0.974**	**0.557**	**0.974**	**0.557**
S14	0.789	0.000	0.833	0.462	**0.916**	**0.598**	**0.916**	**0.598**
S15	**1.000**	0.840	0.789	0.000	0.795	0.018	**1.000**	**0.885**

4.3.3.2 Performance evaluation of bearing fault experimental test bench

The results of the detection methodology and the other detection methods on the bearing fault dataset are shown below. The performance results of each of the methods are detailed in **Table 4.3.6**. According to the results obtained in terms of TPR and TNR that are detailed in **Table 4.3.6**, it can be seen that the data corresponding to the detection of anomalies (TNR) of the methodological proposal, overcome the other methods listed. While the results obtained from TPR, the proposed method is superior in four of the seven analysis scenarios (SS1, SS5, SS6 and SS7). The reconstructed-based reference method (DAE + OC-SVM) without compression processing (DCC) achieves the best performance of the remaining three TPR scenarios. However, DAE + OC-SVM performs the worst for detecting anomalous samples.

Table. 4.3.6 Performance of the detection methods in terms of TPR and TNR on bearing-faults test bench.

Label	DAE		DAE + OC-SVM		DCC + OC-SVM		Proposed method	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
SS1	0.879	**1.000**	**0.986**	0.871	0.997	0.999	0.997	**1.000**
SS2	0.950	0.862	**0.984**	0.376	0.963	0.884	0.963	**0.980**
SS3	0.950	0.558	**0.998**	0.732	0.992	0.874	0.992	**0.937**
SS4	0.950	0.854	**0.985**	0.552	0.964	0.960	0.964	**0.983**
SS5	0.950	0.785	0.948	0.015	**0.987**	0.923	**0.987**	**0.929**
SS6	0.950	**1.000**	0.891	0.594	**0.982**	0.690	**0.982**	**1.000**
SS7	0.950	0.718	0.930	0.583	**0.960**	0.970	**0.960**	**0.995**

Based on the results obtained in **Table 4.3.7**, corresponding to the balanced accuracy, the proposed methodology reaches 97.6% (in percentage terms), overcome the other listed methods. Specifically, the balanced accuracy of the method based on reconstruction (DAE) is 88.2%, which is 9.4% lower than the proposal. The DAE + OC-SVM detection method obtains a balanced precision of 74.5%, while DCC + OC-SVM obtains 93.8% as a result, both lower than the proposed method in 23.1% and 3.8%, respectively. It should be noted that for each of the seven analysis scenarios, the proposed methodology overcome the anomaly detection performance, that is, TNR, of each of the methods considered. Only in the SS1 scenario, the DCC + oc-svm method obtains the same precision as the proposed method.

Table. 4.3.7 Balanced accuracy of the detection methods for bearing-faults test bench.

Label	Balanced Accuracy			
	DAE MSE	DAE + OC-SVM	DCC + OC-SVM	Proposed method
SS1	0.939	0.928	**0.998**	**0.998**
SS2	0.906	0.680	0.923	**0.972**
SS3	0.754	0.865	0.933	**0.965**
SS4	0.902	0.768	0.962	**0.973**
SS5	0.868	0.481	0.955	**0.958**
SS6	0.975	0.743	0.836	**0.991**
SS7	0.834	0.756	0.965	**0.978**
Average	0.882	0.745	0.938	**0.976**

On the other hand, Table XX details the results for the detection of anomalies based on the precision metrics, which correspond to positive predictive value (PPV) and negative predictive value (NPV). It can be noted that the methodological proposal is superior to the other anomaly detection methods. In terms of PPV, the proposal scores the best in all seven test scenarios. In the same way, in terms of PPV, it is superior in the seven test scenarios. This evaluation confirms that the best way to address the problem of anomaly detection is by considering several detection memberships. Furthermore, the proposed detection methodology was evaluated under two different test benches, confirming that its implementation can be generalized to different work environments.

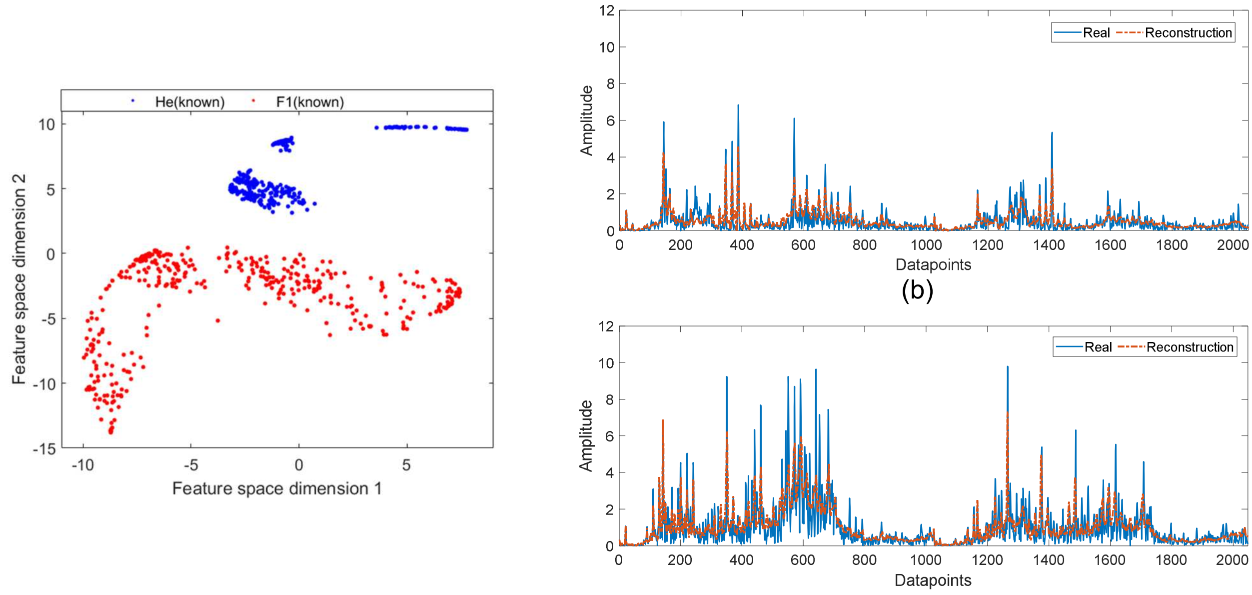
Table. 4.3.8 Performance of the detection methods in terms of TPR and TNR on bearing-faults test bench.

Label	DAE		DAE + OC-SVM		DCC + OC-SVM		Proposed method	
	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
SS1	1.000	0.892	0.912	0.963	0.999	**0.997**	**1.000**	**0.997**
SS2	0.936	0.893	0.782	0.885	0.944	0.922	**0.990**	**0.930**
SS3	0.834	0.767	0.896	0.991	0.940	0.984	**0.970**	**0.985**
SS4	0.933	0.892	0.860	0.932	0.980	0.931	**0.991**	**0.933**
SS5	0.930	0.839	0.742	0.088	0.974	**0.961**	**0.976**	**0.961**
SS6	1.000	0.869	0.868	0.647	0.905	0.930	**1.000**	**0.950**
SS7	0.910	0.827	0.870	0.735	0.989	0.892	**0.998**	**0.894**

4.3.4 Detection Model Performance Discussion

To give an idea of the effectiveness of the proposed anomaly detection method, an analysis of the interpretability of the main methods used during its execution is carried out. The analysis consists first of showing the characterization and feature mapping capabilities of the deep-autoencoder model. As previously mentioned, the deep-autoencoder model comply two functions in the presented detection scheme: first, to generate the mapping of features in a low-dimensional space through the bottleneck of the encoder layers; second, to generate the reconstruction of the signals learned during the training process. Through the mapping function, it is sought that the represented features are grouped in a more compact way among the samples of the same category. In contrast, samples belonging to a different category should ideally be clustered away from the other categories. Regarding the reconstruction of the signals, it is sought that the deep-autoencoder model obtains in the output the near reconstruction to the learned signals. Whereas for unseen signals, are expected to get a wrong reconstruction. In order to show the feature mapping capabilities, in **Fig 4.4.1** (a) you can see the feature mapping obtained from the deep-autoencoder model. The mapped clusters represent the healthy state (He) and a fault state (F1). It should be noted that each state of health presents different sets of clusters, this is because different operating conditions of the experimental test bench are considered. Regarding the reconstruction of the signals, **Fig 4.4.1** (b) and **Fig 4.4.1** (c) show a real and reconstructed signal of the healthy state and a fault signal, respectively. In a qualitative way, it can be observed that the signals reconstructed by means of the deep-autoencoder model have a high similarity with their corresponding real signal. Although it is not an exact reproduction, the reconstructed signal maintains the shape and trend of the most representative harmonics of the real signal for each of the states. The most notable differences are with respect to changes in amplitude in some harmonics, as well as the change in position of these, effects that are mainly due to the intrinsic properties of the system, such as existing non-linearities, as well as the presence of noise, oscillations or external interference. Quantitatively, the reconstruction error of the deep-autoencoder model obtained for the sample He has a Ω_{MSE} of 0.140, while for the fault sample F1, corresponding to the bearing fault, it has a Ω_{MSE} of 0.710. These Ω_{MSE} values correspond to the range of values obtained during training for each corresponding state of health.

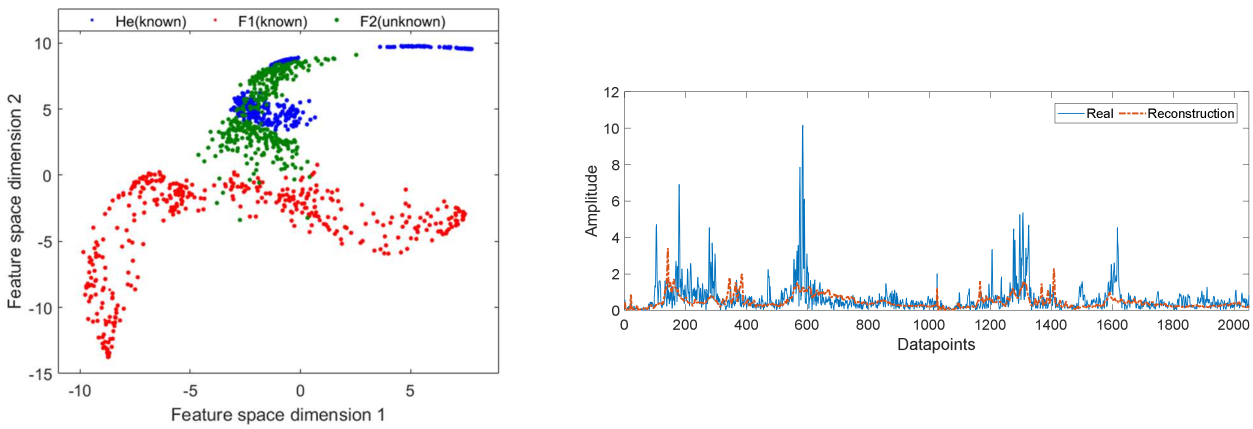
However, it must be taken into account that the mapping function learned by the deep-autoencoder model is specific to the distribution of training data, that is, the deep-autoencoder model will normally not succeed in data reconstruction that are significantly different from the data seen during training. Although there is a risk that the function learned by the deep-learning model can be adjusted to the data not seen in the training. In this case, the reconstruction error value produced by the deep-autoencoder model will be similar to the training data.



(a) (b) (c)

Fig. 4.4.1 Characterization of the healthy state and a fault state by means of the deep-autoencoder model: (a) mapping of the feature space of the encoder bottleneck (b) real and reconstructed signal of the healthy state; (c) real and reconstructed signal of a fault condition.

On one side, under this scenario, an anomaly detection scheme produces a misdiagnosis since the anomaly membership that is used is the reconstruction error metric and therefore will not be able to detect outlier samples. On the other hand, the feature space resulting from the deep-autoencoder model is also not optimized to produce compact clusters, generating sparse clusters instead. Therefore, when mapping unknown samples, these are hardly placed outside the sparse clusters, also making it difficult to detect outliers. To exemplify this issue, **Fig 4.4.2** (a) shows the feature space shown in the previous figure (**Fig 4.4.1**) but adding samples corresponding to an uncharacterized (unknown) condition. As can be seen, the unknown samples overlap with the clusters of the characterized samples.



(a) (b)

Fig. 4.4.2 Characterization of the healthy state and a fault state by means of the deep-autoencoder model: (a) mapping of the feature space of the encoder bottleneck (b) real and reconstructed signal of the healthy state; (c) real and reconstructed signal of a fault condition.

Furthermore, **Fig 4.4.2** (b) shows the characterization of a fault sample (F2), not seen in the training of the deep-autoencoder model. Similarly, qualitatively it can be noted that a sample not seen in the training of the model is not reconstructed effectively. However, although the reconstructed signal in **Fig 4.4.2** (b) does not fit the real signal, specifically in the range of 600 datapoints, the reconstruction error is slightly lower than that produced by the F1 fault. Therefore, although in some cases setting a novelty threshold would be enough to detect uncharacterized fault samples, in this case study, establishing a threshold would not be successful with some fault states, because the Ω_{MSE} of the anomalous samples is lower than that of some samples of training.

To address these shortcomings in anomaly detection schemes, deep-compact-clustering processing is introduced to improve the representation of functions resulting from the bottleneck of the deep-autoencoder model. The objective of DCC is to achieve a compact representation of the feature space mapping. As mentioned above, a traditional deep-autoencoder model is not optimized to generate a compact representation of the data, which affects the representation of unknown samples. Unknown samples produce an overlap with normal samples in feature space, this makes it difficult to detect these samples, which are considered abnormal.

To show the advantages of DCC processing over a traditional deep-autoencoder detection scheme, figures **Fig 4.4.3** and **Fig 4.4.4** compare the application of both schemes. In **Fig 4.4.3** (a) the boundary created through the OC-SVM method is presented over the known samples in the feature space of the deep-autoencoder model without compaction processing. **Fig 4.4.3** (b) shows the same feature space, also considering the placement of anomalous samples corresponding to an F2 fault. It can be noted that several anomalous samples overlap with the normal samples and are placed within the boundary determined to characterize the known states of the system. Furthermore, due to the clusters referring to the known samples are sparse, the oc-svms detection model also fails to capture the known samples, causing a high number of false negatives.

In contrast, when the deep-compact-clustering (DCC) introduces compactness in the data mapping, denser clusters are produced, and consequently the oc-svm-based scheme generates close-fitting boundaries, as can be illustrated in **Fig 4.4.4** (a). Consequently, when mapping samples corresponding to an unknown state, there is a reduction of data points overlapping with the normal samples, as illustrated in **Fig. 4.4.4** (b). It can be noticed that the feature space produced by DCC has grouped all the existing clusters in a compact way, thus, there is enough space for the anomalous samples to be positioned outside the space delimited by oc-svm. Consequently, while working with clustering-based schemes, generating a compact feature space considerably improves anomaly detection. It can also be noted that by having a space of features with compact clusters, the detection model more easily captures samples of known states, generating a low number of false negatives.

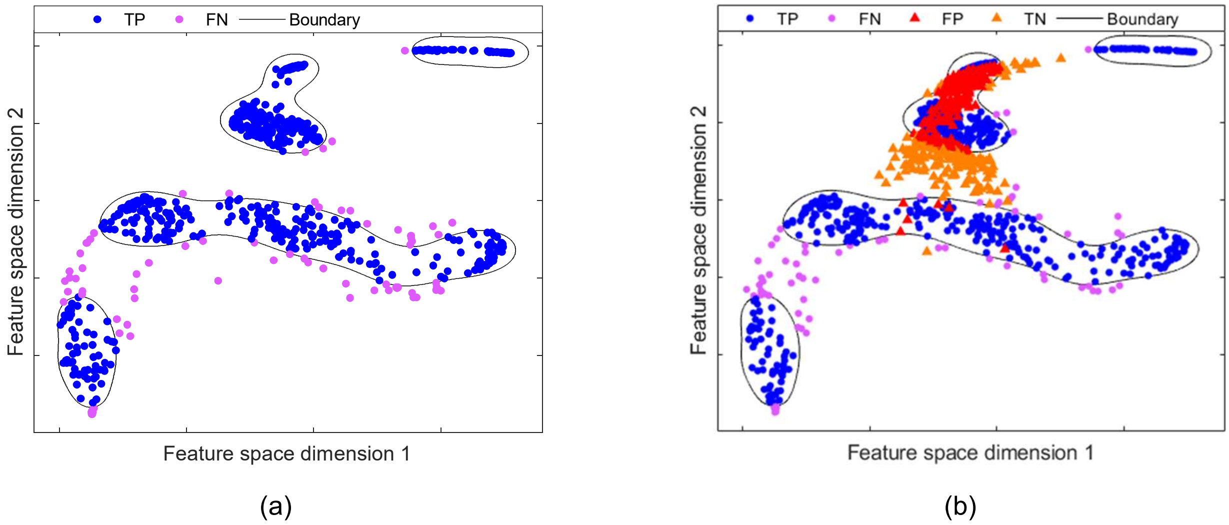


Fig. 4.4.3 Anomaly detection by oc-svm model on feature space mapping resulting from deep-autoencoder: (a) feature space mapping obtained in the training stage considering the He and F1 state.; (b) evaluation of the fault scenario F2 on feature space mapping of the deep-autoencoder model.

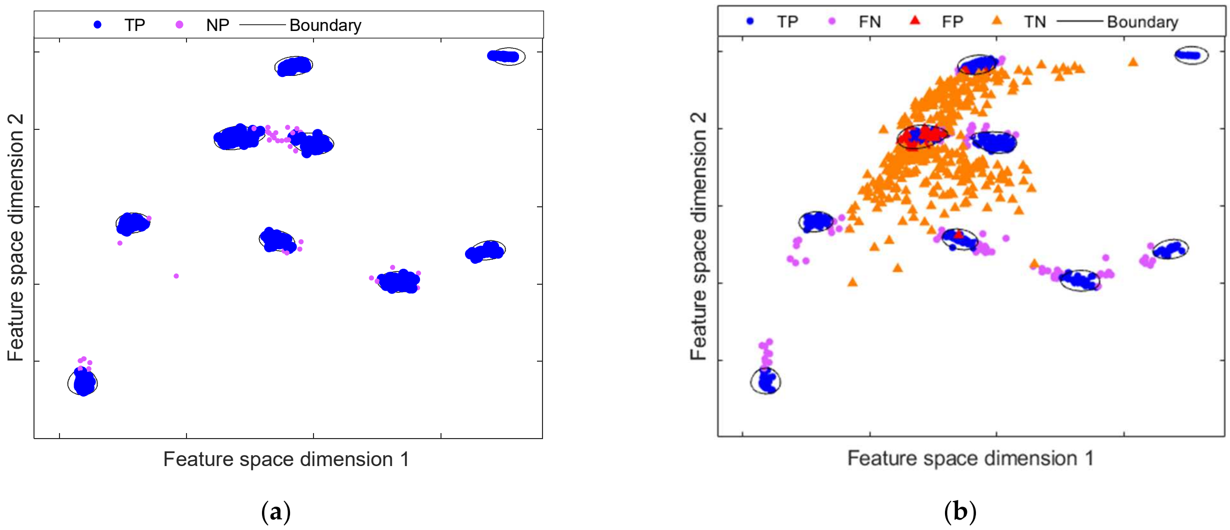


Fig. 4.4.4 Anomaly detection by oc-svm model on feature space mapping resulting from deep-compact-clustering: (a) feature space mapping obtained in the training stage considering the He and F1 state.; (b) evaluation of the fault scenario F2 on feature space mapping of the DCC model.

4.5 Discussion and conclusions

In this chapter a novel method for the detection of anomalies is presented, aligned with the state of the art about condition based monitoring of industrial rotary systems. The methodological proposal consists, first, on the extraction and learning of features through a deep-neural-networks method. Second, processing is applied to refine the compaction of the clusters in the feature space generated by the deep-neural-network method. Third, this compact representation is used as input for a classification scheme of a class, the compaction of clusters in the feature space mapping improves the implementation of detection schemes especially in clustering based schemes. Finally, the signal reconstruction capability of the deep-autoencoder model is combined to generate a more robust anomaly membership and consequently improve anomaly detection results.

It has been shown that the integration of different schemes is the most appropriate way to address anomaly detection, especially for monitoring environments in industrial systems. Due to the fact that multiple patterns have to be dealt with, both of the health condition, the presence of different faults and the different working conditions, it is necessary to apply methods with a great capacity for characterization. In this regard, deep-neural-network-based methods seem to be the natural answer to manage the complexities present in smart manufacturing environments, especially to deal with the non-linearities of rotating systems. Therefore, the main objective of this methodology is to propose a new anomaly detection scheme under an assembly environment. With the application of different methods, such as deep-neural-networks, one-classification schemes, cluster compaction embedded processing and reconstruction methods, detection performance is improved in practical application cases for monitoring schemes.

A comparison between the proposed detection methodology and other classical anomaly detection approaches highlights the superiority of the proposed scheme. Not only with respect to the detection task, but also with respect to the capture of known conditions. In contrast with classical approaches, which generally produce good performance in only one of the two aspects, detection or characterization of known scenarios, the hybrid approach proposed is capable of correctly and effectively characterizing the space of features referring to the different conditions of the electromechanical system, including the different states of faults and the different working conditions.

The detection of anomalies is a difficult task to face and is far from being solved. In the field of monitoring the condition of industrial systems, the detection of anomalies must face different problems, such as the physical configuration of the system, the different operating conditions and the presence of different faults, and therefore, must be a tool that allows solving inconveniences presented in the production processes. In this regard, this methodological process is presented for the detection of anomalies in electromechanical systems. The contributions presented in this methodology can be applied to re-al world settings and therefore generate additional research.

5.

Transfer Learning in Intelligent Fault Diagnosis

Data-driven fault diagnosis methods have presented significant advances, both for detection and classification functionalities. However, there are considerations that must be made so that these diagnostic schemes can be applied reliably in real world applications. In real industrial scenarios, variable work operations are presented, this difference in distribution has a significant impact on the performance of diagnostic schemes.

Transfer learning is a new paradigm of machine learning that has been proposed as the solution to address domain-shift issues. This chapter presents a novel strategy based on transfer learning to address this issue.

CONTENTS:

5. Transfer Learning in Intelligent Fault Diagnosis

5.1 Introduction

This chapter begins with a review of the state of the art regarding transfer learning and the functionalities it can provide within the field of condition based monitoring applied to industrial electromechanical systems. Specifically, the so-called “domain-shift” problems are addressed, derived from eventual changes of operating points related with industrial production systems and their patterns of operations (*e.g.* variation in speed or torque patterns). In this regard, different transfer learning strategies are proposed as solutions to different issues present in the data environments related to such electromechanical systems. Thus, the background and motivation are presented first, next, the state of the art about transfer learning is introduced, finally the proposed contribution of this work is described and discussed.

5.1.1 Background and motivation

In recent years, artificial intelligence-based techniques have been widely used to solve several problems for electromechanical industrial systems fault diagnosis [140]. However, in most real world fault diagnosis applications, variations in system working conditions have not been considered massively. The generalized assumption that a rotary system operates under only one speed regime and one workload represents a useful approach for data analysis, but an unrealistic scenario [141]. Industrial rotary systems often work under different operating conditions, which can be presented under two aspects. First, non-stationary operating patterns, that is, the consideration of speed and/or torque profiles that vary following a predetermined profile over time (*e.g.* ramp-up, ramp-down, starting curve). Second, stationary operating patterns which their characteristics vary (e.g. different values of stationary speed, different acceleration slopes on a ramp, different loads). Specifically, this work focuses on studying the variations in the patterns under stationary operating, that is, changes in the working conditions established. These changes in the working conditions (*e.g.* speed and torque), represent a challenge, and especially for the learning schemes implemented in fault diagnosis methodologies [142]. Most of the methods based on artificial intelligence work by learning a specific function for the data presented, hence a change in the working conditions represents changes in the behavior of the patterns of the learned data. This variance in pattern’s behavior is called “distribution-shift” or “domain-shift” [143]. A significant difference in the operating pattern (in terms of acquired physical magnitudes), will lead to reduce performance of the related fault diagnosis schemes, since the inference system will operate over not properly characterized data distributions. It should be noted that this change in distribution occurs essentially in complex industrial electromechanical environments, which are subjected to multiple operating conditions and are made up of various components that interact with each other.

Therefore, these complex environments provide greater uncertainty in the evolution of the patterns depending on the working point, consequently, it is difficult to infer the behavior of the patterns between one state of operation and another. Modern condition monitoring schemes, therefore, require generalize capabilities to maintain desired performances.

In this regard, deep-learning, as a branch of machine learning in the field of artificial intelligence, has achieved significant advances in different fields with the use of hierarchical architectures based on neural networks [144-146]. The ability to learn a higher-level representation —the process of making sense of complex data at an abstract, conceptual level— in an effective way has placed deep-learning at the center of multiple research lines [147-148]. However, deep learning technology presents some important limitations regarding the adaptation to complex real-world environments. First, deep-learning-based models requires, in general, abundant labeled data to perform their inferences. Especially in industrial environments, it is difficult to obtain massive labeled data representative of the different conditions (*i.e.* operating scenario but also fault and/or degradation conditions). Second, the training data and the test data must have the same distribution (*i.e.*, the same working conditions). A change in the distribution of the data will cause a significant reduction in performance or the prediction model will not work, especially if speed and/or load patterns change substantially.

In this regard, transfer learning represents a general learning method dealing with the need of extract common knowledge from multiple related (but different) application scenarios. Common knowledge must be understood, in the field of monitoring the condition of industrial systems, as the patterns that exist in common regardless of the variations in working conditions, always dealing with the same environment and the same health states. Therefore, a scenario refers to performing an analysis on the different operating states in the electromechanical system. Moreover, it has been shown as a promising tool to complement deep learning algorithms in their learning stage to overcome the limitations of some fields of applications (as described above). Hence, as defined by [149], transfer learning seeks to emulate the ability of humans to leverage the knowledge of approaching a problem to solve different but related problems. Transfer learning can materialize the logics related to monitoring schemes with better learning performance than classical approaches, mainly in front of limited training data.

5.1.2 State of art in transfer learning

A difference in the working conditions of an industrial electromechanical system in regard with the ones considered during the training stage, represents a risk of disrepair in the performance of the related fault diagnosis schemes. This is currently called the “shift-domain” learning problem [150]. Unfortunately, this problem is present in many real-world applications, mainly affecting the reliability of artificial intelligence-based methods used for fault diagnosis schemes. To address “domain-shift” issues and deal with performance degradation, transfer learning is being considered as potential solution by transferring the knowledge from the distribution of training samples used to build the model, defined as the source domain, to the new distribution of test samples, considered target domain [151].

The related literature in the field of electromechanical systems on deep-learning and transfer learning has been approached from different points of view. However, a new paradigm has emerged from the convergence between these two research branches, called "deep-transfer learning" (DTL) [152]. Being an emerging functionality, the categorization of deep-transfer learning is still not clear enough. Recent application approaches suggest three groups of DTL: (*i*) instance-based, (*ii*) model-based and (*iii*) feature-based DTL [153]. Instance-based DTL approaches are generally based on selecting instances or re-weighting instances. Model based DTL approaches have a common neural network structure and hyperparameters between the source domain and target domain. Feature-based DTL approaches are based on learning and sharing the feature space between source domain and target domain.

Instance-based DTL aims to train a deep transfer learning model generalizable from the transfer scenario where the source domain and the target domain have a different distribution caused by the limitation of labeled samples in the target domain, too limited to train a robust diagnostic model by its own. The main motivation for instance-based DTL approaches is that the direct combination of source data and target data could deteriorate resulting model performance for the target task due to the intrinsic data unbalance and, then, leading to a negative learning transfer.

Especially, the negative transfer occurs due to some labeled instances of the source domain have a significantly different probability distribution from instances of the destination domain. To deal with the negative transfer effect, the goal of the instance-based DTL approaches is to identify instances in the source domain that benefit target model generation through instance weighting strategies. Thus, the aim is to automatically learn the weights of the source domain instances and embed them in the objective function. In this regard, Pan *et al* [154] propose an approach based on deep belief network and a transfer learning strategy to deal with the lack of labeled training samples. The target sample data are obtained by collecting the coil current signal of the high-voltage circuit breaker (HVCB), and the auxiliary sample data are obtained through simulation based on an electromagnetic model of the HVCB. The fault diagnosis model uses the selective auxiliary data to augment the learning of target data by adjusting the weight of the training samples. Song *et al* [155] suggested a fault diagnosis method called domain adaptation network (DAN). Considering that machine damage always occurs under different circumstances, DAN retraining strategy seeks to minimize the difference between training and test data while maintaining maximum training accuracy. Another notable contribution of instance-based DTL is the one proposed by Zhang *et al* [156]. Proposed a domain adaptation method based on enhanced transfer joint matching for bearing fault diagnosis. This approach uses the maximum variance discrepancy (MVD) for combining with the maximum mean discrepancy (MMD) for the feature matching and therefore solve the problem under the condition of sample class imbalance.

In regard with the second DTL approach, that is, model-based DTL, this is focused on applications where the source domain and the target domain share some common knowledge at the model level. This implies that the knowledge transfer is integrated into the pretrained deep neural network model base on source data, whose structure and hyperparameters are useful to perform the target task. In the field of monitoring the condition of industrial electromechanical systems, it implies that the source domain data is substantial enough to generate a learning model due to the working conditions and failure categories used, therefore, to improve performance in target tasks (other work conditions), it is only necessary to explore which part of the already trained source model helps to improve the learning process of the model for the target domain.

The main objective of model-based DTL approaches is to take advantage of the knowledge learned in the source task, that is, to take advantage of the deep-neural model part to help improve the model learning process in the target domain. Model-based DTL algorithms work under the assumption that labeled instances of the target domain exist during the training stage. Based on this, there are two training subcategories for model-based DTL schemes: sequential training and joint training. On the one hand, in sequential training, there are a robust model trained under the instances of the source domain, which are numerous, then this source model is adjusted by retraining with the instances of the target domain, which usually has limited labeled instances. There are two stages in sequential training-based DTL approaches. The first is pretraining, and consists of initializing all the model parameters with the data from the source domain. The second stage consists of fine-tuning on the target domain, this can be done by freezing some layers of the correctly trained source model and adjusting only the rest of the layers with the instances of the target domain. Or reuse the entire model from source as a way to initialize the model from target by retraining under target instances. On the other hand, joint training consists of training the diagnostic model for the source and target tasks simultaneously. This type of training follows the objective of optimizing the performance of the target task by taking advantage of the common knowledge of the source task. There are two strategies for joining the source and target tasks. The first is based on solid sharing of hidden layer parameters, keeping task-specific layers separate. The second is the sharing of soft parameters, which change the coefficient of weights adapting to each specific task or adding regularization terms to the objective function.

In this sense, Shao *et al* [157], proposed a novel deep learning framework for machine fault diagnosis. The aim of the diagnostic scheme is to use transfer learning as a tool to accelerate the training of a deep neural network. The study framework uses labeled time-frequency (*i.e.* Wavelet transformation) images to pre-train the network and extract lower-level features. Then, time-frequency images of the target task are used to fine-tune the higher levels of the neural network architecture. In [158], a generic intelligent fault diagnosis system for bearings based on AlexNet deep network was proposed to address the problem of network overfitting.

A transfer learning strategy was used as a starting point to train a model and then transferring the knowledge to another task (target task), which represents a different working condition in the motor-driven system. Han *et al* [159], developed a transfer learning framework based on pre-trained convolutional neural network to tackle the problem of distribution imbalance between the instances of the source task (large datasets) and the target task (limited instances). The approach leverages the knowledge learned from the model trained with abundant data to facilitate diagnosing a new but similar task. Zhao *et al* [160], present a transfer learning framework based on deep multi-scale convolutional neural network to address the change in data distribution caused by changes in working conditions, as well as changes in internal and external environments. The approach is based on a multi-scale module that is used to learn the differential features, then a global pooling technology is adopted that replaces the traditional fully connected layer. Finally, the pretrained model weights are updated to other different but similar tasks following a fine-tuning instead of training a network from scratch. In [161], Haidong-Shao *et al*, propose a transfer learning scheme fed through infrared thermal images for the detection of faults in bearings under a variable work operation. The scheme is based on a modified convolutional neural network, which introduces a stochastic pooling and leaky rectified linear unit that improves the classical CNN. Then, parameter transfer is used to adapt the CNN model to the target domain.

Finally, in regard with the third DTL approach, that is, feature-based DTL, this aims to provide deep neural network models the ability to transfer knowledge at the feature space level, as opposed to instance-based or model-based approaches. The main idea of the feature-based DTL approaches is to learn a common mapping function between the source and target domains, in this way a common feature latent space can be obtained where the differences between domains are reduced, that is, domain-invariant and task-discriminative representations are obtained. The DTL feature-based methods applied to monitoring approaches of electromechanical systems have the objective of finding the characteristic patterns that are useful to recognize fault states and at the same time that are domain invariant, that is, that do not depend on only of working conditions. There are two variants within the feature-based DTL approaches: without or with adaptation to target domain. The approaches without adaptation are based on extracting low-level features using a pre-trained model from the source domain, then using the extracted features that are useful for the target domain because are closely related to the source domain. By contrast, the approaches with adaptation adjust the feature space of different domains using domain adaptation methods even for domains with significant difference. Since most application cases deal with cases with notable differences in domains, most feature-based DTL research focuses on adaptive approach scenarios.

The principal aim of feature-based DTL with domain adaptation is how to estimate and learn the invariance between the source domain and the target domain. That is, find the correlation between for each health condition of the rotating system operating in different operating conditions. To accomplish this aim, three strategies have recently been developed: discrepancy-based domain adaptation, adversarial-based domain adaptation, and reconstruction-based domain adaptation. First, the subcategory discrepancy-based domain adaptation is focused on aligning the change distribution in the features and improving the learning capacity by reducing the discrepancy based on distance metrics or criteria defined in the levels of representation between the source domain and the target domain. Metrics that have been applied for discrepancy-based domain adaptation include maximum mean discrepancy (MMD) [162], Kullback–Leibler divergence [163], correlation alignment (CORAL) [164] and Wasserstein Distances (WD) [165]. Second, the subcategory called adversarial-based domain adaptation that is inspired by Generative Adversarial Networks (GAN) [166]. GAN is a strategy based on two neural networks, a generator (G) and a discriminator (D). The process that G and D follow is a kind of competition with each other in a similar way to a zero-sum game, where the gain of one network implies the loss of another. For adversarial-based domain adaptation approaches, GAN is modified in order to endow the deep neural network with the capability of learning domain-invariant representations. That is, to guarantee that the features resulting from the difference of several domains cannot be distinguished. Although there are variants in adversarial-based domain adaptation approaches, called generative and non-generative, in most applications non-generative models have been opted for, due to focus mostly on learning the domain-invariant representations and not in generating new instances. Adversarial-based domain approaches modify GANs by introducing objective functions such as domain-confusion loss, starting from three main blocks: feature extractor (instead of the generator), the domain discriminator and the task-specific classifier. A promising strategy around non-generative adaptation schemes is the Domain Adversarial Neural Network (DANN) [167]. DANN introduces a new form of deep network training based on a Gradient Reversal Layer (GRL), which follows the objective of learning representations of different domains through maximizing the domain confusion loss. Third, the subcategory reconstruction-based domain adaptation, which uses the autoencoders functionalities together with a specific classification task focused on optimizing an encoder that captures domain-specific representations and another shared encoder that learns common representations between domains. Reconstruction-based domain adaptation schemes consist of a shared decoder between domains that learns to reconstruct instances with minimal reconstruction error. The traditional metrics of mean absolute error (MAE) and the root mean square error (RMSE) are used to accomplish this objective.

Table. 5.4.1 Categorization of Deep-Transfer Learning applied in industrial electromechanical systems.

Approach	Basic principles	Application references	Limitations
Instance-based DTL	Based on instance select or re-weight strategies. • Weight-estimation • Heuristic-reweighting	Estimation of network weights for fault diagnosis by estimate the adaptation matrix between the source and target instances, weighting the source instances that are negative for the target model.	Model performance depends on the number and quality of target instances.
Model-based DTL	Share the neural network structure and parameters. • Sequential training strategy • Joint training strategy	The problem of different working conditions is addressed by pre-training a deep model with source instances under different working conditions, and later tuning using the instances under the target working condition.	Fine-tune based models heavily relies on training data labeled in the target domain. Also, their performance would drop drastically if labeled instances are unavailable or insufficient.
Feature-based DTL	Share or learn the common feature representation. • Discrepancy-based domain adaptation • Adversarial-based domain adaptation • Reconstruction-based domain adaptation	It is sought that the schemes learn "universal" features under different working conditions from three aspects. From the aspect of domain adaptation based on discrepancy reduction. From the aspect of adversarial learning domain adaptation, GAN architectures have been implemented with the aim of learning features sensitive to tasks but insensitive to domain changes. From the aspect of domain adaptation through reconstruction, the application of encoder-decoder to generate models with the ability to generalize under different working conditions.	Requires building a complex model with different structures based on neural networks. It also depends heavily on the quality and quantity of data from the source domain and target domain to find the shared feature space.

5.1.3 Methodological proposal

It must be noted that techniques and strategies based on transfer learning have emerged recently, so their application, and especially around condition monitoring, represent a cutting-edge scenario. Considering the advances and the shortcomings in the state of the art about the application of transfer learning in the field of industrial electromechanical systems monitoring, in the development of this chapter an innovative fault diagnosis methodology is proposed to face the main problem present in the application of intelligent diagnostic schemes: the domain-shift. Although fault diagnosis under different working conditions represents a challenging problem, the combination of deep-learning with adversarial learning represents a powerful framework to explore. In this regard, the proposed contributions address the following points:

- The proposal of a framework applied to pattern recognition (*i.e.* fault diagnosis), combining both, deep-learning capabilities in feature learning, and transfer-learning strategy dealing with working points changes in the rotating systems under monitoring.
- The implementation of an innovative transfer learning scheme supported by domain adversarial learning to deal with both, changes between source domain and target domain, and the difficulty of obtaining labeled samples for different operating conditions.
- The consideration of a more realistic industrial scenario, that is, considering the imbalance in the categories of fault conditions, will allow generating a more robust diagnostic scheme.
- The proposal of a methodology that considers tasks under multiple domain combined with an adversarial learning strategy to provide diagnostic schemes with a capacity to adapt to changes in work conditions.

5.2 Theoretical approach and proposed contribution

5.2.1 Theoretical definition and preliminaries

In order to clearly state the issue around transfer learning applied to industrial systems monitoring to be solved, in this section, a basic notation, symbols and definitions are firstly introduced. According to the Transfer Learning, and taking the algorithms based on deep-learning as a starting point, it is explained how learning capabilities and feature extraction are used to transfer knowledge. In this sense, some concepts are defined to facilitate the understanding of related mechanisms and strategies [128, 153].

Domain, represented as $D = \{\mathcal{L}, P(X)\}$ incorporate two components: the feature space $\mathcal{L}$ and the probability distribution $P(X)$, where $X = \{x|x_i \in \mathcal{L}, i = 1, \dots, N\}$, is the dataset that contains N instances. An instance must be understood as the set of elements that defines x. For example, x is the set associated with the sample, each represented by the attributes label, the probability distribution, domain category, and others. Therefore, when referring to different domains, it implies that there is a difference in the feature space or a difference in the probability distribution between those two domains. Regarding machinery fault diagnosis, different working conditions, different sensor location, different fault categories or/and different machine can be considered as a different domain.

Task, represented as $T = \{\mathcal{Y}, f(\cdot)\}$ having a specific domain D, is composed of two components: a label space $\mathcal{Y}$ and a mapping function $f(\cdot)$, where $Y = \{y|y_i \in \mathcal{Y}, i = 1, \dots, N\}$, is a label set for each of the instances N in D. The mapping function $f(\cdot)$, also represent as $f(x) = P(y|x)$, corresponds to a nonlinear function which establishes the relationship between the input instance and the inferred prediction, since it establishes the learning of the patterns from the dataset. Similar to the concept of "domain", different tasks are established as there are different label spaces between each tasks. Different fault categories and severities (*i.e.* label space) can be defined as different tasks.

Transfer learning, given a specific source domain $D^S = \{\mathcal{L}^S, P^S(X^S)\}$ with the respective source task $T^S = \{\mathcal{Y}^S, f^S(\cdot)\}$ and a target domain $D^T = \{\mathcal{L}^T, P^T(X^T)\}$ with the respective target task $T^T = \{\mathcal{Y}^T, f^T(\cdot)\}$, which objective is to learn the best mapping function $f^T(\cdot)$ for the target task T^T through transfer learning from the source domain D^S. classical machine learning and deep learning problems assume that the domain and task between source and target are identical (*i.e.* $D^S = D^T$ and/or $T^S = T^T$), however, in transfer learning, this assumption is not considered, that is, it faces the problem where the domain and/or the task between source and target scenarios are different (*i.e.* $D^S \neq D^T$ and/or $T^S \neq T^T$).

Regarding the notation mentioned above, a definition of deep-transfer learning can be described as: given a transfer learning task $f^{S \to T}(\cdot)\colon X^T \to Y^T$ regarding $[D^S,\ D^T, T^S, T^T]$, the deep-transfer learning seeks to learn the mapping function $f^{S \to T}(\cdot)$ making use of the representation capabilities of a deep-learning model, that is, a model based on deep neural networks, trained following the strategies of the state of the art..

In **Fig 5.2.1**, the basic concept of the process that follows a strategy based on transfer learning is shown. It can be noted that it starts from a data source to train a model, commonly based on deep neural networks, focused on solving a specific task. Subsequently, this model is used through knowledge transfer, adapting it to solve a target task, which is different in domain.

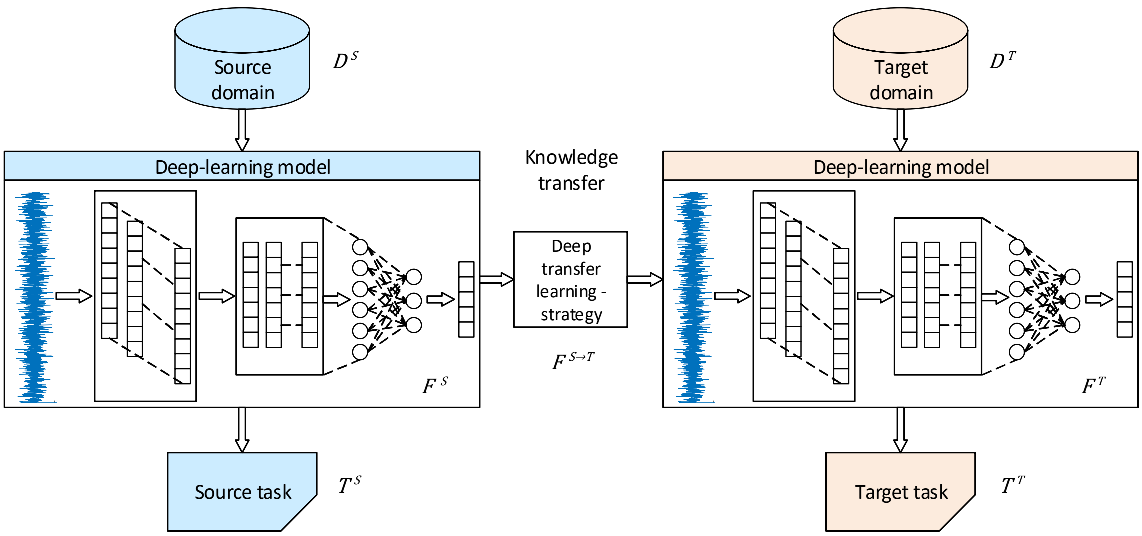


Fig. 5.2.1 Deep-transfer learning framework.

Taking into account the basic definitions around deep-transfer learning and the functionalities that they can offer within the field of condition based monitoring, the existing problems around industrial scenarios must be addressed. Specifically, in the industrial maintenance context, four main issues arise:

1) As previously mentioned, traditional deep-learning models obtain prominent performance against tasks in a single domain (one working condition). However, when faced with tasks that involve different domains, the performance of traditional deep-

learning models decreases as the patterns associated with fault categories diverge. This represents a particular problem in monitoring the current manufacturing industry since it is complex to deal with changes in working conditions, diversification of data and variable wear of components.

2) Industrial environments are subject to constant updates, as component replacement, and it is not always possible to collect data on each update or characterize faults of each of the components after some component incorporation. However, it can always be useful to reuse historical data and take advantage of existing diagnostic models, an alternative being reusing schemas and updating them through transfer learning strategies.
3) Diagnostic models are especially focused on classifying faults through of the labeled data presented during the training process, however, the detection of unknown/novel patterns, also called anomalies, remains a significant challenge to overcome. For practical application within industrial environments, it is essential that fault diagnosis schemes can automatically detect such novel patterns of operation to avoid false negatives/positives during the model execution.
4) Most of the research focus on the diagnosis of faults around the main components of rotating systems. Combined faults often occur in manufacturing environments, since different faults can evolve together, generating diversified behaviors. However, little research has been devoted to combined faults due to the complexity of their study.

Taking into account the aforementioned definitions around deep transfer learning and the associated issues within the field of condition-based monitoring, these problems should be addressed taking into account the most common industrial scenarios. Specifically, in the monitoring environment, four main issues arise that can be addressed through learning transfer strategies:

Improved generalization performance: In this case there is no difference between the label space of the target domain and the label space of the source domain, *i.e.*, $\mathcal{Y}^T \equiv \mathcal{Y}^S$, the types of faults considered are identical and the objective is focused on improving the generalization performance considering only a change in the operating conditions. Which implies the same electromechanical system, same database in terms of health categories considered, but different working conditions (*i.e.* torque and/or speed).

Partial fault diagnosis: In this case, the label space of target domain corresponds a un subset of the space of the source domain, *i.e.*, $\mathcal{Y}^T < \mathcal{Y}^S$, the challenge in this type of issue is to make use of the transfer to match label spaces which are identical. For example, in the task corresponding to the source domain, four categories of fault are present, while in the target domain, only three categories of fault are present, operating at a different working condition.

Emerging fault detection: In this case, the label space of the target domain is an open set of the label space of the source domain, *i.e.*, $\mathcal{Y}^T > \mathcal{Y}^S$ which implies that there are new categories of faults never exist in the source domain. For this issue, the priority is to detect new faults and correlate those that are identical. This case study refers to considering the presence of a new fault category that is present in the target domain, but is not present in the source domain, which implies implementing a detection process previously.

Compound fault analysis: In this case, the label space of the target domain has a singular variation with respect to the source domain, that is, in the target domain there are faults composed of single or multiple single faults of the source domain. The challenge in this issue involves diagnosing compound fault as a new category and not misclassifying it as a simple fault.

In this regard, the case study that is addressed in this thesis work corresponds to that of partial fault diagnosis, since it represents the scenario that is closest to a real industrial work environment. As depicted in **Fig. 5.2.2**, the main motivation behind the partial fault diagnosis scenario is that, in the field of monitoring, it is a promising functionality to use labeled historical data and the respective diagnostic models on new acquisitions of related scenarios. The model trained with the labeled data can be used to transfer knowledge from the large-scale source domain to a target domain that presents a smaller scale (a space of smaller fault categories). The main challenges to address partial domain fault diagnosis are due to two factors:

a) The lack of information regarding labels in the target domain due to it is expensive and unrealistic to acquire large amounts of data under fault conditions.
b) The presence of outlier faults in the source domain can lead to a negative transfer (decrease in the performance of a previously trained model), producing diagnostic errors with target domain acquisitions.

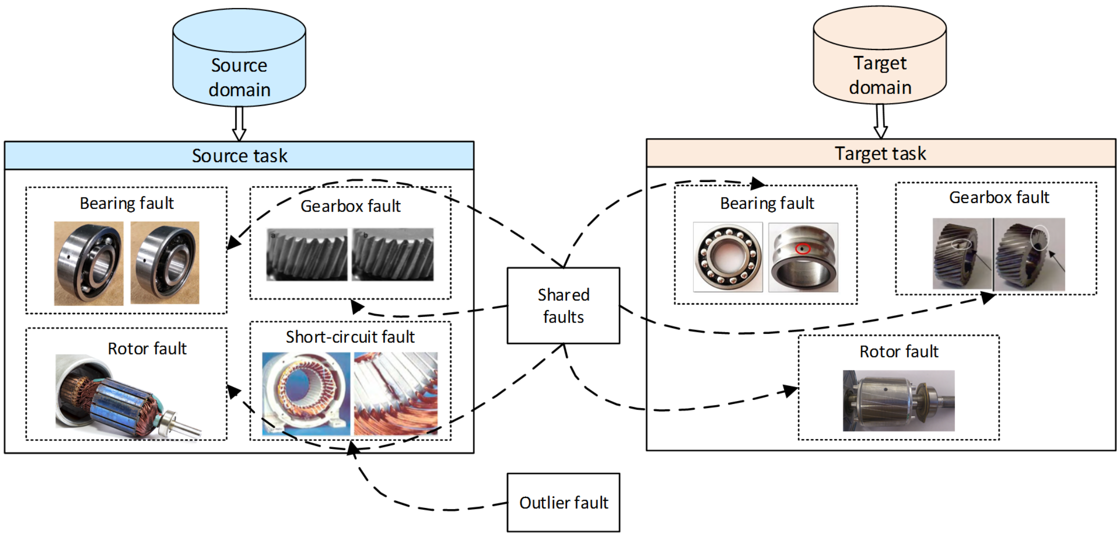


Fig. 5.2.2 Partial domain fault diagnosis scenario where the source task presents an outlier fault category that is not present in the target domain.

5.2.2 Contributions for transfer learning

The application of strategies based on transfer learning in the field of monitoring of electromechanical machines is a new topic and a rapidly growing area of research. The main functionality of transfer learning within the monitoring field is to provide the generalization capacity to fault diagnosis schemes. Domain adaptation is a transfer learning strategy focused on learn domain-invariant and discriminative features for an unlabeled target domain by leveraging the information of the labeled source domain, obtaining a well-performance model even if there is a shift between source and target domains. In addition, recent research shows that models based on deep networks can learn features with better invariance and discrimination that are valuable for being transferable and thus performing better domain adaptation. Therefore, deep domain adaptation (DDA) has been a topic of recent exploration with the aim of obtaining more satisfactory performance in transfer learning tasks. However, as it has been aforementioned, transfer learning has been applied to address various problems that still exist in the field of monitoring. In this regard, this chapter presents a methodology based on deep-domain adaptation to solve two of the main issues in industrial applications: (i) the improvement of generalization performance and, (ii) the label space imbalance problem, specifically, over partial fault diagnosis. In **Fig 5.2.3(a)**, the potential of techniques based on domain adaptation to solve the issue of data distribution difference is depicted. However, the assumption of the source domain and the target domain share the same label space, can lead to misclassifications. In **Fig 5.2.3(b)** a partial domain adaptation scenario with the different classes is shown, which means that the distribution of the two domains is different, and the target label space is a subset of the source label space.

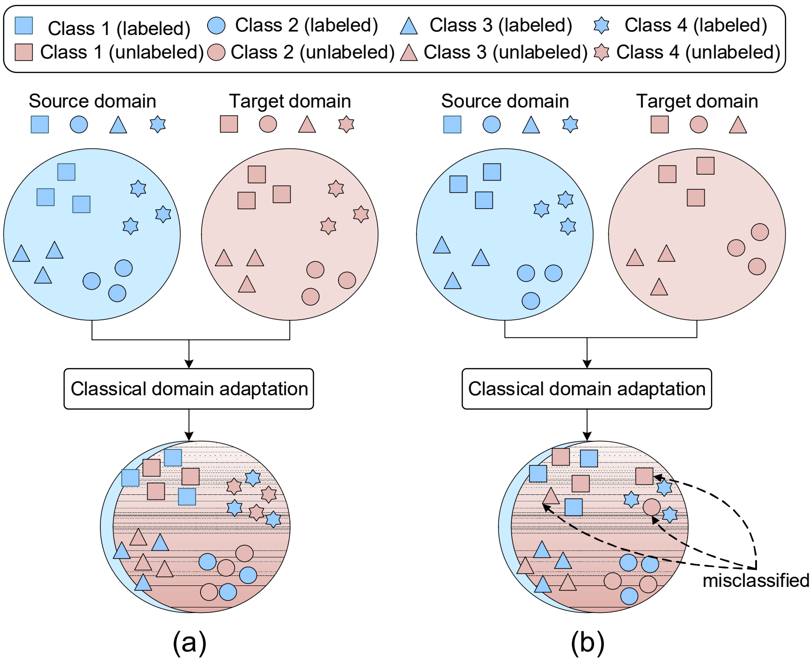


Fig. 5.2.3 Application of domain adaptation to transfer learning scenarios (a) domain adaptation with identical distribution of classes in source and target; (b) domain adaptation with different class distribution, specifically, Class 4, is not available in the target domain, which leads to misclassifications applying classical transfer learning procedures.

Nevertheless, the application of a classical domain adaptation model will lead to classification failures, since most of these schemes work under the assumption of the same label space, that is, the same categories of faults. This is an unrealistic industry scenario, since it is not feasible to wait till data for the same categories between the source domain and the target domain are collected. Considering these existing issues around the application of transfer learning strategies, a step-by-step workflow of the proposed methodology for domain adaptation is shown in **Fig 5.2.4**.

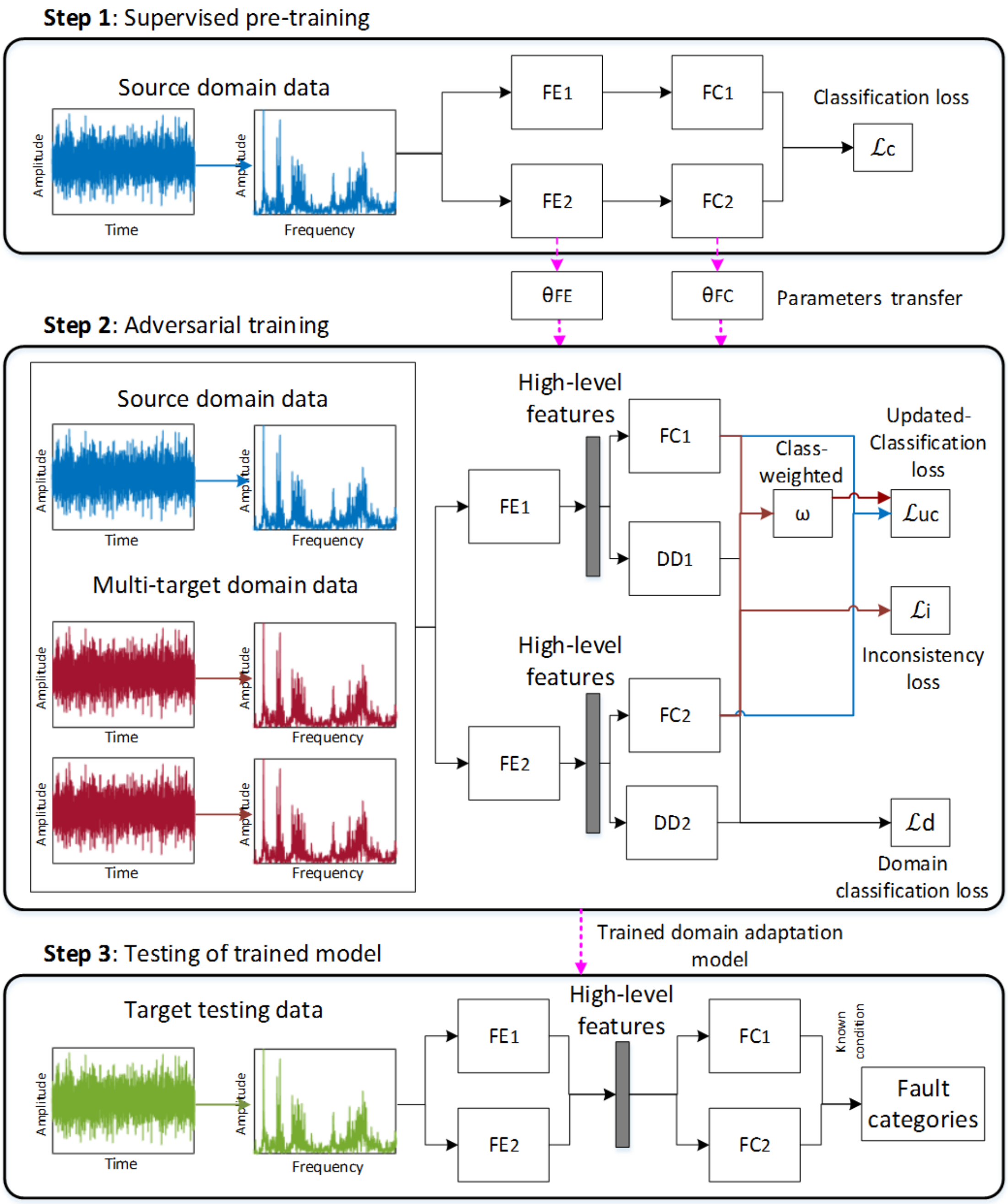


Fig. 5.2.4 Flowchart of the proposed methodology for fault diagnosis based on transfer learning. Supervised pre-training stage: Generation of the initial model from the source domain data, through feature extractor (FE) and fault classifier (FC) modules. Adversarial training stage: An adversarial-learning strategy is followed to perform domain adaptation between instances of different working conditions, which introduce a domain discriminator (DD) blocks. Testing of trained model stage: Validation of the model under new instances (different working conditions).

The proposed methodology is motivated by two key aspects:

- First, an accurate identification of the distribution of classes related to different target domains, and relate them to the categories of the source domain that are shared in the tag space.
- Second, the minimization of the distribution discrepancy between the categories shared between the target domains and the source domain, at the same time that the discriminatory features are learnt to carry out a proper classes' classification.

Therefore, the proposed domain adaptation-based model will be able to find a shared latent feature space between the source and different target domains in such a way that the discrepancy between domains is minimized. In addition, the proposed model contemplates solving the issue of the imbalance in the label space between the source and the different targets. The proposed transfer learning framework can be divided into the following three phases:

1) supervised pre-training, using source domain data;

2) adversarial adaptation training of N target domains and single source domain; and

3) test the domain adapted model.

Specifically, a dual model composed of two 1-dimensional convolutional neural network (1-D CNN) in order to adapt to vibration signals, is designed, which constitutes the general model. First, in the pre-training stage, the two networks are trained following the objective of correctly classifying the conditions of the source domain. Then, the weights of the two networks are transferred and the model is modified by adding a domain discriminator block that is used to calculate the target label distribution in conjunction with a loss function based on inconsistency. In addition, the loss function of the classifier is updated so that the network can focus on instances in the shared label space and downweigh instances that are not shared between the target domain and the source domain. Next, the two networks are retrained using three loss functions, which are described in the following subsections: (*i*) updated-classification loss, (*ii*) inconsistency loss and (*iii*) domain classification loss. In the following, all these steps are detailed.

5.2.3 Pre-processing

To carry out the implementation, analysis and validation of the proposed methodology related with the mentioned transfer learning scheme, an electromechanical systems driven by an electrical motor is considered, from which a set of vibration signals are collected for fault diagnosis. After this, a filtering and segmentation process follows to obtain each of the samples and thus make up the database. Immediately afterwards, fast Fourier transformation (FFT) is applied on each of the segmented samples to obtain the spectra of the signals. Due to the amplitude of the frequency components are too small to cause changes in the weights of the network, the samples are scaled before being used as input to the model following the recommendations of [134]. This process is followed for each of the considered domains: the source domain and the different target domains.

The source domain is considered as a wide database, with samples collected from different categories of faults with available labels. In contrast, the different target domains refer to databases which work at different operating conditions and samples of all faults categories have not been collected, causing an imbalance to carry out transfer learning. In **Fig. 5.2.5**, the scenario of transfer learning to handle multiple targets concurrently, starting from one source domain, is depicted.

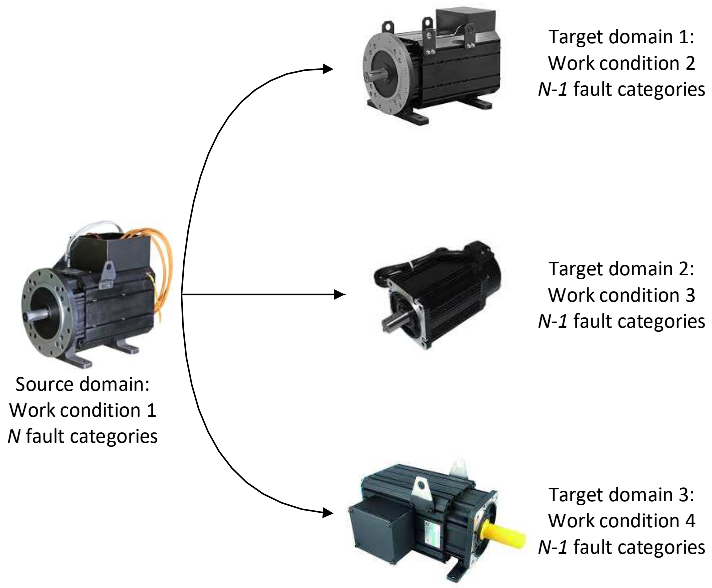
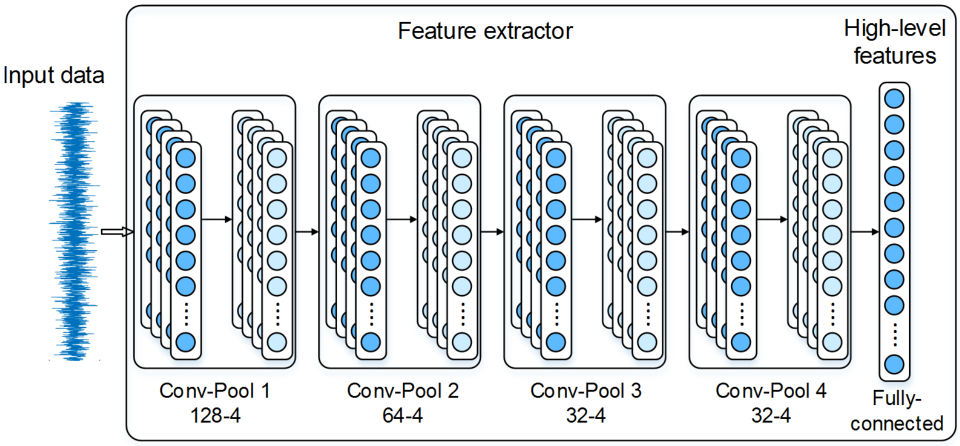


Fig. 5.2.5 Transfer learning scenario that considers multiple targets, which are given by different working conditions and imbalance in failure categories. Take into account this transfer learning scenario is more scalable and valuable for practical fault diagnosis problems.

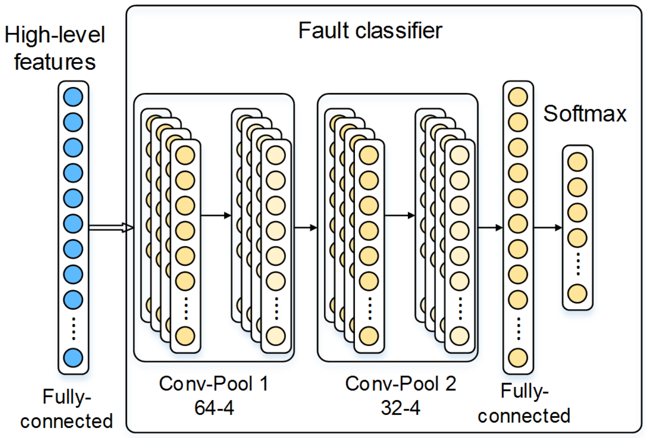
5.2.4 CNN architecture

The proposed network architecture is composed of different processing blocks based on convolutional neural networks (CNN). Specifically, three blocks are adopted: feature extractor, domain discriminator and fault classifier; which are part of the three steps considered during the proposed diagnostic procedure in **Fig. 5.2.6**.

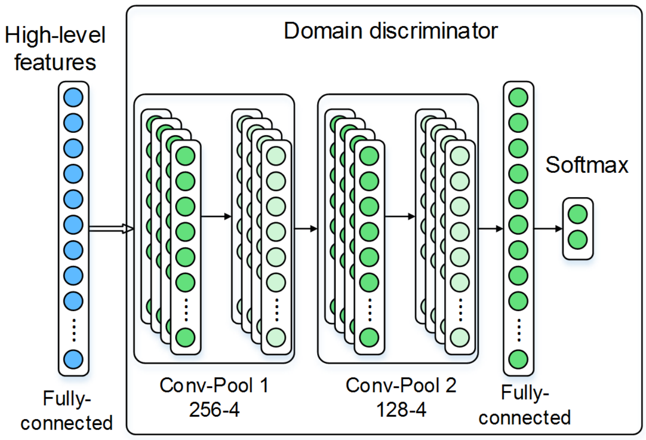
For step 1, corresponding to pre-training, two feature extraction modules (FE1 and FE2) are used with the same structure (same number of layers, number of neurons, batch size, kernel) and different parameters (due to each network is initialized randomly) which are denoted as θ_{FE1} and θ_{FE2}, respectively, each with its corresponding fault classification module (FC1 and FC2) whose parameters are denoted as θ_{FC1} and θ_{FC2}, respectively. For step 2, a transfer learning is made from the pre-trained network, it is also modified by including two domain discriminator blocks (DD1 and DD2) at the output of each feature extraction module, whose parameters are denoted as θ_{DD1} and θ_{DD2}, respectively. For step 3, only the feature extraction and fault classification modules are taken to make the prediction of each of the fault categories. The feature extractor blocks are composed of four convolutional layers, whose filter size is 4 and filter numbers are 128, 64, 32 and 32,



(a)



(b)



(c)

Fig. 5.2.6 Proposed network architectures CNN-based to perform the methodology for fault diagnosis based on transfer learning. (a) Feature extractor (FE): which complies with the functionality of learning features from the data. (b) Fault classifier (FC): it has the functionality of predicting the failure category. (c) Domain discriminator: aims to distinguish the source and target domains based on the extracted features.

Then, a fully-connected layer with 128 neurons is used. This feature mapping will be used later to feed the fault classification and domain discrimination blocks. The fault classifier structure consists of two convolutional layers with 64 and 32 filters of size 4. A fully-connected layer of 64 neurons and a softmax layer with a neurons output corresponding to the number of fault conditions complete the module for predicting fault categories. The structure of the domain discriminator consists of two convolutional layers with 256 and 128 filters of size 4, followed by one fully-connected layer with 512 neurons. At the output, two neurons are attached, representing the two different domain labels, and finally a softmax layer is used for domain classification. In addition, in order to prevent overfitting and create a more generalizable model, the techniques of dropout, batch normalization and the activation function of rectified linear units (ReLU), are adopted during training in most layers. The back-propagation (BP) algorithm by gradient descent is used for the updates of all the network parameters, and the Adam optimization method is applied. Each of the proposed architectures is shown in **Fig 5.2.4**.

5.2.5 Step 1: Supervised pre-training

The first stage of the proposed detection method is related to the fault identification model. A dual model based on two identical CNNs is built to classify the fault categories present in the source domain.

5.2.5.1 Pre-trained model

As shown in **Fig 5.2.4**, the fault identification model is trained from the source domain data by means of two feature extraction and two fault classifier blocks. For each fault classifier, F is obtained at the output of each classifier network, which for each sample, gives a class probability distribution that measures the label. Therefore, the optimization goal is defined by classification loss as follows:

$$\mathcal{L}_c = J_c(\text{F}_i(X_s), Y_s), \qquad i = 1,2 \tag{54}$$

where J_c represents the cross-entropy loss function, which is described as follows:

$$J_c = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{|y^S|} 1_{[k=y_i^s]} \log p(y = k|x_i^s) \tag{55}$$

where $1_{[k=y_i^s]}$ stands the value is 1 when $k = y_i^s$, in another way the value is 0. It can be noticed that although two networks are trained fed by the source domain data, the parameters of each one are different at the end of the training due to the initialization is random and individual. Therefore, the result is two networks that have identical architecture and different learning parameters.

5.2.6 Step 2: Adversarial training

The second stage consists of transferring the parameters of the networks from the previous stage, and modifying the model by adding a domain discriminator module, which is fed by the mapping obtained from the feature extractors. Once the two networks are initialized, the classification loss function is updated and both an inconsistency loss ($\mathcal{L}_i$) and a domain discrimination loss ($\mathcal{L}_d$) are adopted. The parameters of networks (θ_{FE}, θ_{FC}, θ_{DD}) are trained and updated iteratively using data from the source domain and from the different target domains.

5.2.6.1 Class weighted learning

In conventional domain adversarial learning, the aim is to reduce the domain-shift caused by a distribution change between the source domain and the target domain, considering that the label space is the same. In other words, it seeks to reduce the difference between the feature space caused by a work change in the machine, assuming that the health and fault states are identical. Besides, the assumption that the same label space is shared between the source domain and the target domain is an unrealistic scenario in the industrial environment. Since the partial transfer diagnosis scenario is more realistic in the field of monitoring, this methodology is based on resolving this practical issue. Directly applying a conventional domain adversarial learning scheme to the partial transfer learning problem results in performance degradation due to the source outlier classes, as shown in **Fig 5.2.3(b)**. The source outlier space refers to the label space that is not related to the label space of the target domain ($\mathcal{Y}^T \backslash \mathcal{Y}^S$), in contrast, the source shared space refers to the same label space of the target domain. Hence, the objective of the partial transfer diagnosis is to differentiate those samples from the source domain that belong to the outlier label space and which belong to the shared label space. From this, a strategy is generated that discriminates the samples belonging to the outlier label space and performs the adaptation of the domain between the shared label space of the source domain and target domain. First, the networks pre-trained with the data from the source domain are used, since the output of each network provides a class probability distribution $\digamma$ that computes label membership. In this way, the importance of each sample of the source domain can be quantified by means of the output $\tilde{y}$ of each of the target samples. Since there is no label relationship between the outlier space and the target domain label space, the chances of assigning the target samples to the outlier space are improbable.

Therefore, the label distribution weight is estimated using the two 1-D CNNs as follows:

$$\omega = \frac{1}{2n_t}\sum_{i=1}^{n_t}(\tilde{y}_{1i}^t + \tilde{y}_{2i}^t) = \frac{1}{2n_t}\sum_{i=1}^{n_t}\left(p_1(y|x_i^t) + p_2(y|x_i^t)\right) \quad (56)$$

where ω corresponds to a vector of $|\mathcal{Y}^S|$-dimensional and $\sum_{i=1}^{\mathcal{Y}^S}\omega_i = 1$ according to the definition; n_t is the number of target samples; $\tilde{y}_{1i}^t$ and $\tilde{y}_{2i}^t$ stands for the output of blocks from the classifier 1 and 2 for the *i*th target sample; $p_1(y|x_i^t)$ and $p_2(y|x_i^t)$ represent the specific output probabilities. Thus, the weights indicate the label distribution for the training target data and the classification loss function for the source domain is updated according to these weights as follows:

$$\begin{aligned}\mathcal{L}_{uc} &= \omega J_{c1}(\mathrm{F}_1(X_c), Y_c) + \omega J_{c2}(\mathrm{F}_2(X_c), Y_c) \\ &= -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{|\mathcal{Y}^S|}\omega_k 1_{[k=y_i^s]}\log p_1(y=k|x_i^s) \\ &\quad -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{|\mathcal{Y}^S|}\omega_k 1_{[k=y_i^s]}\log p_2(y=k|x_i^s)\end{aligned} \tag{57}$$

where ω_k is the *k*th element of ω. In (57) it can be seen that the purpose of the weight ω is to guide the model to focus on the samples in the shared label space and discriminate the samples from the outlier label space.

5.2.6.2 Domain adaptation by minimizing inconsistency

After the categorization of the shared classes, the learning of the domain-invariant and discriminative features is carried out. The domain shift is estimated through a strategy based on the measurement of inconsistency by means of two classifiers. This classification by inconsistency consists of guiding the model to learn discriminative and invariant representations between domains (work conditions) in order to correctly classify the fault categories of the unlabeled target data. If the function that measures the inconsistency is large, it means that the features of the target domain is far from the source. And therefore there are fault categories that coincide between the different domains. On the contrary, if the inconsistency measurement is small, it means that the features do not completely match and therefore there are outlier categories in the source domain. Therefore, a domain adaptation can be performed and at the same time the domain-invariant and discriminative features are learned by minimizing this inconsistency between two network-models outputs for target data. Specifically, the pairwise distance between two networks-models is adopted to calculate the inconsistency loss, mathematically, it is expressed as follows:

$$\mathcal{L}_i = \frac{1}{n_t}\sum_{i=1}^{n_t}\|p_1(y|x_i^t) - p_2(y|x_i^t)\|_1 \tag{58}$$

where p represents the output of the softmax layer of each of the blocks of the fault classifiers.

5.2.6.3 Domain discrimination weighted learning

Besides, to reduce the domain-shift, that is, the difference between the source domain and the different target domains, it is proposed to include a domain discrimination block, following the theory regarding adversarial domain adaptation. The adversarial training strategy follows two main objectives, on the one hand, the parameters of the domain discriminator (θ_{DD}) are optimized to minimize the loss of the domain, which consists of a binary classifier for two domains. On the other hand, the feature extractor is updated to maximize the loss of the domain discriminator.

Therefore, the final objective function of the proposed method based on adversarial training is defined as:

$$\mathcal{L} = \mathcal{L}_{uc} + \mathcal{L}_i + \mathcal{L}_d \tag{59}$$

where $\mathcal{L}_{uc}$ and $\mathcal{L}_i$ represents the fault classification loss and the inconsistency loss, respectively, previously estimated; and $\mathcal{L}_d$ denotes the domain classification loss. The loss function $\mathcal{L}_d$ represents the difference between the real domain label d and the predicted label.

On one hand, during the domain adaptation strategy, it is assumed that samples belonging to the same fault conditions (identical label) are clustered in the same regions. In contrast, the source outlier space classes are not clustered with the other domains, which implies high confidence in the prediction of the source domain outlier classes and therefore $\mathcal{L}_d$ produces a small value. On the other hand, for classes within the shared space, adversarial training is applied and the domain discriminator is less successful on domain label prediction. Consequently, the loss $\mathcal{L}_d$ is higher, and is defined as follows:

$$\mathcal{L}_d = \alpha_1 \omega_j J_d^s(DD_i(\mathbb{F}_i(X_s)), d_s) - \alpha_2 J_d^t(DD_i(\mathbb{F}_i(X_t)), d_t) \quad i = 1,2 \tag{60}$$

where J_d^s and J_d^t represents the cross-entropy loss function, for the source domain and the target domain, respectively. DD_i denote the output of each domain discriminator. α_1 and α_2 denote the penalty coefficients. d_s is the source-domain label and d_t is the target-domain label. ω_j denote the weights for the source classes to address partial transfer learning and is estimated as follows:

$$\omega_j = \frac{1}{n_{s,j}} \sum_{x_i \in D_j^s} J_d^s(DD_i(\mathbb{F}_k(X_s)), d_s) \quad j = 1,2, \dots, N_c \quad k = 1,2 \tag{61}$$

where ω_j represents the weight and D_j^s the samples for the jth source class, $n_{s,j}$ is the number of the samples in D_j^s, and N_c is the number of the source classes. Eq. (61) represents a way to find the classes of the source domain that are shared with the target domain. Therefore, ω_j is a way of guiding the network to focus on adapting shared classes, while atypical classes should be ignored.

The $\mathcal{L}_d$ optimization method is addressed in a different way than the other loss functions ($\mathcal{L}_{uc}$ and $\mathcal{L}_i$), since it cannot be addressed with the gradient descent optimization method. This is due to the adversarial learning process that the network follows, on the one hand it is expected to maximize the domain prediction loss by optimizing the feature extractor and at the same time minimizing the optimization of the domain discriminator.

To carry out this optimization problem, the Gradient Reversal Layer (GRL), is adopted to update the network parameters [156]. Specifically, GRL acts as identity mapping in the feed forward process, and in the back-propagation, it flipping the sign of the gradients received from the previous layer. Mathematically GLR can be represented by the function

R(*x*) for the two processes, in the feed forward and back-propagation.

$$R(x) = x,$$
$$\frac{dR(x)}{dx} = -\lambda\mathbf{I} \tag{62}$$

where $\mathbf{I}$ represents the identity matrix and λ is a penalty parameter commonly set as 1. With the implementation of the gradient reversal layer in conjunction with the stochastic gradient descent-based algorithms, the complete optimization of the network can be carried out. And therefore, all network parameters are updated during training at the same time.

In this regard, iteratively, the loss functions, (57), (58) and (60), are implemented to obtain the final networks for the prediction of fault categories for the target samples. During the training, it can be noticed that the joint minimization of (59) produces the alignment between the different domains and improves the diagnostic precision for the conditions of the different target tasks.

5.2.7 Step 3: Testing of trained model

Finally, in the test step, the evaluation of the model trained with the data collected from different target domains is carried out. Extensive experiments are carried out to validate the effectiveness of the proposed methodology based on adversarial domain adaptation to address domain-shift and partial fault-diagnosis problems. In this regard, different case studies are analyzed, establishing different working conditions (different domains), introducing imbalances in the fault categories (partial domain), performing cross validation (changing the outlier category in the target domain) and considering different objectives (multi-target). Thus, the evaluation consists of obtaining the probability of belonging to each fault category through the two fault classification blocks F_1 and F_2, respectively. Therefore, the prediction of the category is estimated as follows:

$$Fault\ category = J(\mathrm{F}_1, \mathrm{F}_2) \tag{63}$$

where J represents the cross-entropy function. It can be noticed, that the prediction is made by means of two classification networks, so the diagnosis has a high confidence regarding the class of the predicted fault. The results of the proposed methodology are shown in the following section.

5.3 Analysis and validation of the proposed methodology

This section shows the experimental implementation of the proposed methodology based on the transfer learning strategy called adversarial learning to address the domain change issue present in industrial environments over electromechanical systems. As previously mentioned, a change in the working conditions of the industrial system is considered a domain-shift, a practice that is common in real manufacturing environments. One way to address this problem is through strategies based on transfer learning, such as the proposed methodology.

Especially, an adversarial learning strategy is combined with an inconsistency adaptation strategy to carry out a more robust training of the diagnostic model. In this regard, a total of 36 experiments are performed, which include different partial diagnostic scenarios to evaluate the performance of the proposed methodology, which are detailed below.

5.3.1 Evaluation Scenarios

In order to evaluate the adaptation capacities and the performance of the proposed diagnostic scheme, different experiments are performed using the test bench described in Annex 1. A total of four operating conditions of the experimental test system were collected ($\mathrm{O_{PI}}$, $\mathrm{O_{PII}}$, $\mathrm{O_{PIII}}$, $\mathrm{O_{PIV}}$), which correspond to power frequency low (30 Hz) and power frequency high (60 Hz) in combination with motor load low and motor load high (40 and 75% of the nominal load). These operating conditions represent the different domains that are analyzed for the implementation of the methodology. The distribution of the four operating conditions is shown in **Table 5.4.1**. Regarding the state of health, for this study four states are included, which include the healthy state and four failure states: bearing fault, demagnetization fault and gear fault. The description of each of the states is included in **Table 5.4.2**. Since this methodology mainly concerns the partial transfer diagnosis issue, and also follows a training strategy based on single-source multiple target (1SmT), extensive experiments are carried out for analysis. A simple partial transfer diagnosis task is denoted as follows: $\mathrm{O_{PI-4}} \rightarrow \mathrm{O_{PII-3}}$; where $\mathrm{O_{PI-4}}$ represents the source domain from where the knowledge transfer is carried out, which corresponds to operation condition I considering 4 health states, while $\mathrm{O_{PII-3}}$ represents the target domain that corresponds to operation condition II and which only contains a subspace of 3 states of health.

In contrast, a multiple-target partial transfer diagnosis task is denoted as follows: $\mathrm{O_{PI-4}} \rightarrow \mathrm{O_{PII-3}}$ & $\mathrm{O_{PIV-3}}$; which involves a source domain and two target domains, the second target domain refers to the operation condition IV and similar to the first target domain contains a subspace of categories of 3 health states.

Table. 5.4.2 Operating conditions of the electromechanical system under study.

Index		Power supply	Load condition
I	$\mathrm{O_{PI}}$	30 Hz	40 %
II	$\mathrm{O_{PII}}$	30 Hz	70 %
III	$\mathrm{O_{PIII}}$	60 Hz	40 %
IV	$\mathrm{O_{PIV}}$	60 Hz	70 %

Table. 5.4.3 Fault categories considered for the study of the proposed methodology.

Index		State	Description	Operating condition
1	He	Healthy	Normal state without fault	
2	BF	Degraded bearings	The non-end bearing inner and outer races have been scraped thoroughly	Power supply frequencies: 30 Hz and 60 Hz
3	DF	Partially demagnetized motor	50% of nominal flux reduction in one pair of poles	Load conditions: 40% and 70% of the nominal load
4	GF	Degradation on the reduction gearbox	Two gear teeth were smoothed out to impose a degradation degree	

Therefore, in order to evaluate comprehensively the proposed methodology related to the partial transfer learning issue, different fault diagnosis tasks have been investigated. The detailed information of the concerned tasks is presented in **Table 5.4.3**. A total of 36 tasks are performed, of which the combinations between the different operating conditions and the different subspaces of the fault categories are considered.

Table. 5.4.4 Details of the fault diagnosis tasks for the partial transfer diagnosis.

Task	Source	Target 1	Target 2	Target test	Source class	Target 1 class	Target 2 class	Target test class
T1	O_{PI}	O_{PII}	O_{PIII}	O_{PIV}	All	1,3,4	1,2,4	1,2,3
T2	O_{PI}	O_{PII}	O_{PIV}	O_{PIII}	All	1,3,4	1,2,4	1,2,3
T3	O_{PI}	O_{PIII}	O_{PIV}	O_{PII}	All	1,3,4	1,2,4	1,2,3
T4	O_{PII}	O_{PI}	O_{PIII}	O_{PIV}	All	1,3,4	1,2,4	1,2,3
T5	O_{PII}	O_{PI}	O_{PIV}	O_{PIII}	All	1,3,4	1,2,4	1,2,3
T6	O_{PII}	O_{PIII}	O_{PIV}	O_{PI}	All	1,3,4	1,2,4	1,2,3
T7	O_{PIII}	O_{PI}	O_{PII}	O_{PIV}	All	1,3,4	1,2,4	1,2,3
T8	O_{PIII}	O_{PI}	O_{PIV}	O_{PII}	All	1,3,4	1,2,4	1,2,3
T9	O_{PIII}	O_{PII}	O_{PIV}	O_{PI}	All	1,3,4	1,2,4	1,2,3
T10	O_{PIV}	O_{PI}	O_{PII}	O_{PIII}	All	1,3,4	1,2,4	1,2,3
T11	O_{PIV}	O_{PI}	O_{PIII}	O_{PII}	All	1,3,4	1,2,4	1,2,3
T12	O_{PIV}	O_{PII}	O_{PIII}	O_{PI}	All	1,3,4	1,2,4	1,2,3
T13	O_{PI}	O_{PII}	O_{PIII}	O_{PIV}	All	1,3,4	1,2,3	1,2,4
T14	O_{PI}	O_{PII}	O_{PIV}	O_{PIII}	All	1,3,4	1,2,3	1,2,4
T15	O_{PI}	O_{PIII}	O_{PIV}	O_{PII}	All	1,3,4	1,2,3	1,2,4
T16	O_{PII}	O_{PI}	O_{PIII}	O_{PIV}	All	1,3,4	1,2,3	1,2,4
T17	O_{PII}	O_{PI}	O_{PIV}	O_{PIII}	All	1,3,4	1,2,3	1,2,4
T18	O_{PII}	O_{PIII}	O_{PIV}	O_{PI}	All	1,3,4	1,2,3	1,2,4
T19	O_{PIII}	O_{PI}	O_{PII}	O_{PIV}	All	1,3,4	1,2,3	1,2,4
T20	O_{PIII}	O_{PI}	O_{PIV}	O_{PII}	All	1,3,4	1,2,3	1,2,4
T21	O_{PIII}	O_{PII}	O_{PIV}	O_{PI}	All	1,3,4	1,2,3	1,2,4
T22	O_{PIV}	O_{PI}	O_{PII}	O_{PIII}	All	1,3,4	1,2,3	1,2,4
T23	O_{PIV}	O_{PI}	O_{PIII}	O_{PII}	All	1,3,4	1,2,3	1,2,4
T24	O_{PIV}	O_{PII}	O_{PIII}	O_{PI}	All	1,3,4	1,2,3	1,2,4
T25	O_{PI}	O_{PII}	O_{PIII}	O_{PIV}	All	1,2,4	1,2,3	1,3,4
T26	O_{PI}	O_{PII}	O_{PIV}	O_{PIII}	All	1,2,4	1,2,3	1,3,4
T27	O_{PI}	O_{PIII}	O_{PIV}	O_{PII}	All	1,2,4	1,2,3	1,3,4
T28	O_{PII}	O_{PI}	O_{PIII}	O_{PIV}	All	1,2,4	1,2,3	1,3,4
T29	O_{PII}	O_{PI}	O_{PIV}	O_{PIII}	All	1,2,4	1,2,3	1,3,4
T30	O_{PII}	O_{PIII}	O_{PIV}	O_{PI}	All	1,2,4	1,2,3	1,3,4
T31	O_{PIII}	O_{PI}	O_{PII}	O_{PIV}	All	1,2,4	1,2,3	1,3,4
T32	O_{PIII}	O_{PI}	O_{PIV}	O_{PII}	All	1,2,4	1,2,3	1,3,4
T33	O_{PIII}	O_{PII}	O_{PIV}	O_{PI}	All	1,2,4	1,2,3	1,3,4
T34	O_{PIV}	O_{PI}	O_{PII}	O_{PIII}	All	1,2,4	1,2,3	1,3,4
T35	O_{PIV}	O_{PI}	O_{PIII}	O_{PII}	All	1,2,4	1,2,3	1,3,4
T36	O_{PIV}	O_{PII}	O_{PIII}	O_{PI}	All	1,2,4	1,2,3	1,3,4

The experiments consist of transferring knowledge from the source domain and adapting to target domains 1 and 2. Different operating conditions are considered between the source domain and the two target domains, and the fault category subspace of each domain is also different. Additionally, the evaluation of a target test domain is performed, that corresponds to an operating condition and a subspace of unseen fault categories during the training. The objective of these experimental tests is to evaluate the generalization capabilities that is the main objective of diagnostic schemes based on domain adaptation. For this study, 100 labeled source domain samples under each machine condition are available for training in different scenario, and 80 unlabeled samples for each target domain under each considered condition are tested.

5.3.2 Cross-domain diagnostic results performance

For performed the evaluation of the performance of the methodology based on domain adaptation, firstly, the results obtained for the data corresponding to target domain 1 and target domain 2 are presented. Secondly, the results obtained using the source domain data are compared. And finally the diagnostic performance results for a target test not seen during training are presented.

5.3.2.1 Performance evaluation for target domains

The experimental results of the 36 partial transfer learning tasks for the diagnosis of faults between domains are presented in **Fig. 5.3.1**. For the first twelve experiments **Fig. 5.3.1**(a), the subspace of three categories is established for each of the two target domains (1,3,4 & 1,2,4) and a cross domain is performed. Similarly, in the following twelve experiments **Fig. 5.3.1** (b), the category subspaces are established for each of the target domains (1,3,4 & 1,2,3), while switching between each of the domains. And finally in the last twelve experiments **Fig. 5.3.1** (c), the subspaces of 1,2,4 & 1,2,3 are established for target domain 1 and target domain 2, respectively. For the 36 experiments conducted, the four fault categories available are used as source domain. The experiments consist of transferring knowledge from the source domain and adapting to target domains 1 and 2. Different working conditions are considered between the source domain and the two target domains, and the fault category subspace of each domain is also different. The objective of these experimental tests is to evaluate the generalization capabilities that is the main objective of diagnostic schemes based on domain adaptation. The reported experimental results are averaged by fifteen trials to reduce the effect of randomness, and the mean values and standard deviations are provided. It can be observed that under the proposed methodology fairly high testing accuracies are obtained in most of the transfer tasks, with the exception of some cases. The average value for the transfer tasks of target domain 1 is 88.67%, while the average value of accuracy for the transfer tasks of target domain 2 is 94.18%. Specifically, in transfer tasks 7, 10 and 11, **Fig. 5.3.1** (a), a low test performance (66%, 68.2% and 59.2%) of target domain 1 was obtained.

This is mainly due to the fact that the difference in power frequency between the domains causes that some fault categories cannot be distinguished correctly because their corresponding frequency spectra are quite similar under the working conditions that are studied.

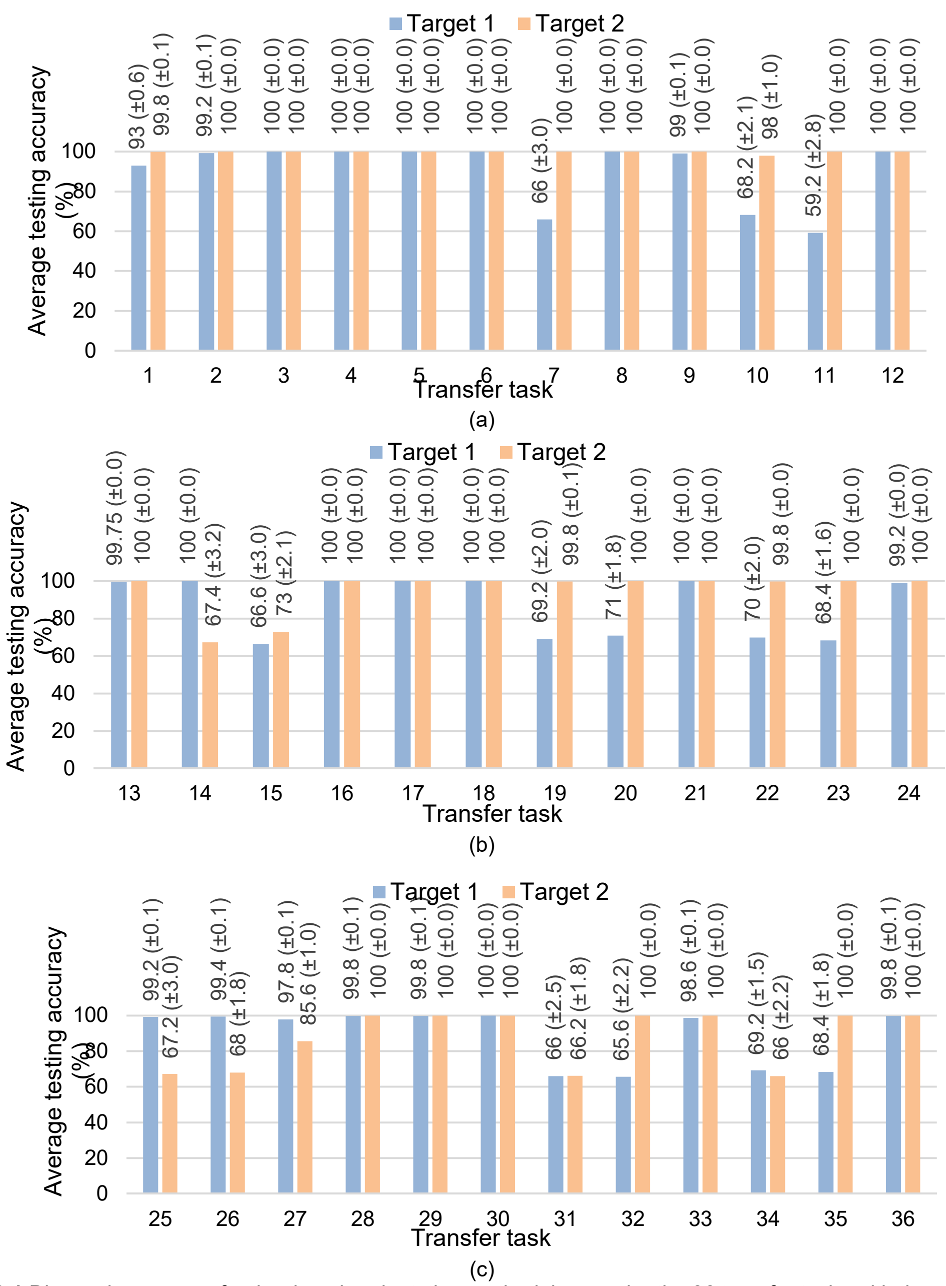


Fig. 5.3.1 Diagnosis accuracy for the domain adaptation methodology under the 36 transfer tasks with the respective deviation evaluating the corresponding working conditions of Target 1 and Target 2.

Similarly, the low performance in transfer tasks 15, 19, 22 and 23 of target domain 1 and transfer tasks 14 and 15 of target domain 2 is associated with the high difference in power frequency supply of the working conditions between the source domain and the respective target domain. It can be noticed, that although there is a low performance in the aforementioned transfer tasks compared to the rest of the tasks, this low performance is due to the fact that a fault category is being erroneously classified, on the contrary, all other categories are successfully classified despite the difference in power frequency.

Note that in most of the experiments a performance of 100% accuracy or close is obtained, which corresponds to 53 transfer tasks out of a total of 72, considering each of the target tasks. These results demonstrate the effectiveness of the proposal methodology to address the problem of transfer partial diagnosis and in general as a strategy to transfer knowledge from a diagnostic task to a different but similar task. Additionally, it can be noted that the standard deviations obtained present significantly stable values with small variations.

5.3.2.2 Performance evaluation for source domain

In addition, evaluating the data corresponding to the source domain, it can be shown that under the proposed approach there is no risk of negative transference, which is interpreted as the loss of performance of the source task by involving new target tasks. In this regard, **Fig. 5.3.2** shows the average accuracy corresponding to each of the transfer tasks considered. It can be observed that the performance of almost all the transfer tasks reaches 100% accuracy, which implies that while knowledge is incorporated into the diagnostic model, the performance for the source task continues to be preserved. Specifically, an average accuracy of 99.9% is obtained for the 36 transfer tasks. In a qualitative way, it can be compared the performance between the data from the source domain and the two target domains. While on average for the data of target 1 a performance of 88.67% is obtained and for the data of target 2 a performance of 94.18% is obtained. With the performances presented, it is demonstrated that the proposed methodology can correctly diagnose the fault categories present both in the source domain and in the different target domains even when there is a category mismatch (partial transfer) and addressing a domain change caused by the different operating conditions between domains.

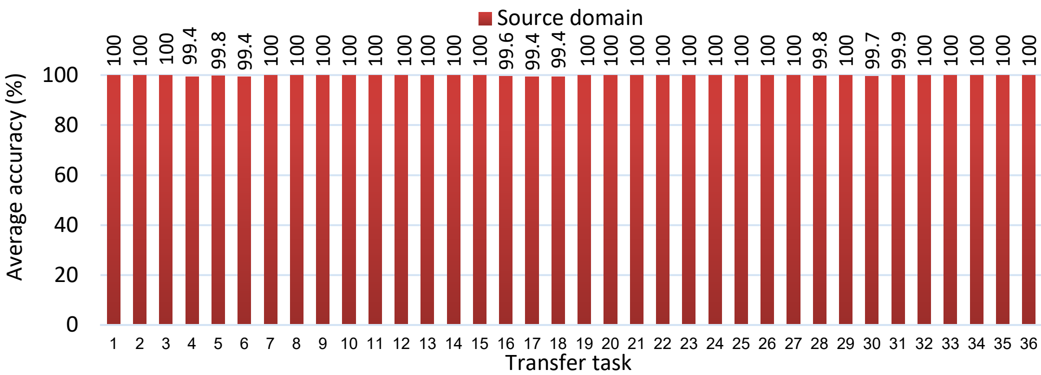


Fig. 5.3.2 Average accuracy corresponding to the labeled data of the source domain for the 36 proposed transfer tasks. In general, the fault diagnosis for the source domain data is successful, reaching an accuracy of around 100%. It implies that the diagnostic model does not "forget" the knowledge obtained to diagnose the main diagnostic task.

5.3.2.3 Performance evaluation for target test

Furthermore, test experiments are carried out with a target test, which corresponds to operating conditions and subspace of categories not seen during training. The objective of these tests is to analyze the generalization capabilities of the proposed methodology. In **Fig. 5.3.3** the diagnostic results for the experiments carried out for the target test are presented. In the same way as the experiments of the target domains 1 and 2, fifteen trials are carried out and averaged, in addition, standard deviations are also presented. The average accuracy obtained for the transfer tasks corresponding to the test experiments is 89.21%. It can be noted that in most experimental scenarios a good performance is obtained and accuracy close to 100% is achieved. These results show that the proposed knowledge transfer strategy is not only useful for the instances of the target domains used during training to perform the domain adaptation, but it is also useful for validating with instances under uncharacterized operating conditions. Thus, providing a generalization capacity to the monitoring schemes. Only in some diagnostic scenarios low performance are obtained. Low yields are only obtained in some diagnostic scenarios. For example, in target tasks 14, 25 and 26, the performance obtained is around 65%, however, when analyzing the operating conditions of the source domain and the target test domain, it can be seen that the difference in power frequency causes a fault category is misclassified. On the contrary, when the difference in the working condition presents a variation in the load, the diagnostic performance is more accurate.

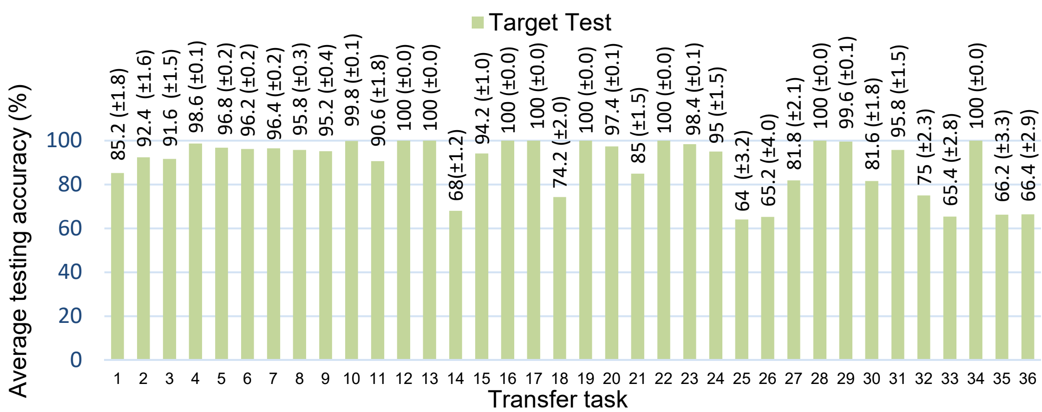
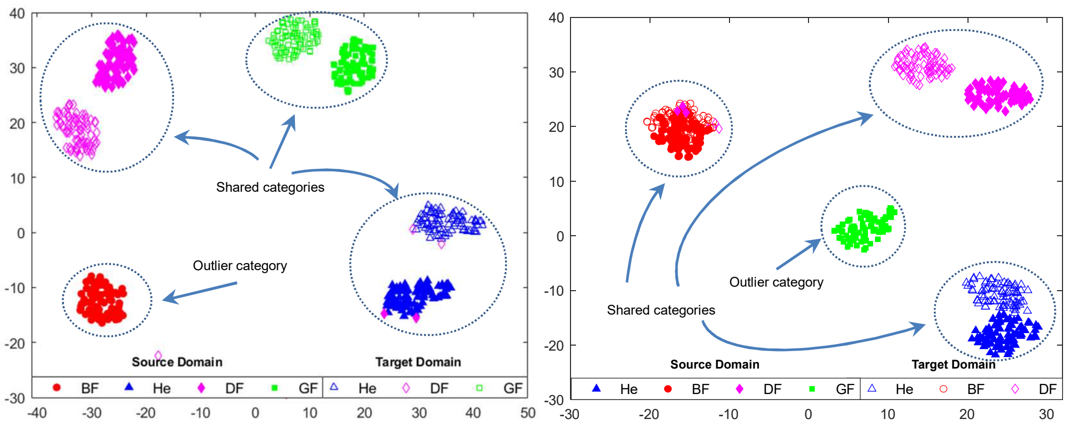


Fig. 5.3.3 Evaluation of the domain adaptation methodology under the 36 scenarios for the target test. The data from the target test correspond to a different working condition than those used in the training (source, target 1 and target 2). The performance obtained for most of the tasks reaches a good diagnostic accuracy.

5.3.3 Feature Visualization Analysis

In order to provide visual and interpretable insights into and perform a more intuitive analysis into the effects of transfer learning on the discrepancy or concurrency of the distribution of features from the source and target domains, the *t*-distributed stochastic neighbor embedding (*t*-SNE), technique is used to map the deep features into a 2-D space. Through this feature mapping, the domain adaptation methodology can be qualitatively assessed and thus a more instinctive understanding occurs. Under the 2-D space, the alignment of features between the source domain and the different target domains is analyzed. The first transfer task T9 $\mathrm{O_{PIII-4}} \rightarrow \mathrm{O_{PII-3}}$ is taken as an example to conduct a comprehensive analysis. The visualization results of the learned deep-features from the output of the fully connected layer of the extractor are displayed in **Fig. 5.3.4 (a)**. In the feature mapping it can be seen that the distributions of the source domain and the target domain that share the same fault categories are close to each other, while the outlier fault category is not aligned with any other. Even though the distributions are not completely aligned, they are drawn closer to each other and far from the other fault categories, which leads to a correct diagnosis. The reason why the instances are not completely aligned is mainly due to the difference in power frequency between the domains, while the source domain works under a power frequency of 60 Hz, the samples of the target domain 1 were acquired under a power frequency of 30Hz. In the same way, the visualization of the learned deep-features for the transfer task T20 $\mathrm{O_{PIII-4}} \rightarrow \mathrm{O_{PIV-3}}$ is presented in **Fig. 5.3.4 (b)**. It can be observed that for this test scenario, instances of the same fault categories are aligned and mostly overlap. The differences in the working conditions for this transfer task refer to the load, the source domain operates with a 40% load, while the target domain works under 70% of the nominal load, the power frequency conditions are the same. Therefore, it can be inferred that depending on the difference in operation conditions facilitate or hinder adaptation between domains. In any case, it is shown that through domain adaptation it is possible to reduce the discrepancy to carry out a correct diagnosis, providing generalization capacity to intelligent monitoring schemes.



(a) (b)

Fig. 5.3.4 Visualization results of the learned deep-features. (a) Transfer task T9. It can observe that the samples from each fault category are well separated. Additionally, the shared categories (▲,■,◆) are aligned or quite close to each other, while the samples from the outlier fault category (●) are separated, avoiding confusion with the other categories. (b) Transfer task T20. Similarly, for this transfer task. Shared categories (▲, ●, ◆). Outlier category (■).

Furthermore, an analysis including two target domains on the transfer task T28 $\mathrm{O_{PII-4}} \rightarrow \mathrm{O_{PI-3}}\ \&\ \mathrm{O_{PIII-3}}$ is performed. The corresponding learned deep-features from the fully connected layer output are displayed in **Fig. 5.3.5**. It can be noticed that the instances of the different domains are aligned or are drawn closer to each other which means that the same failure categories are grouped and classified well under all domains through the use of the proposed domain adaptation strategy. In contrast, the outlier fault category (which in this test scenario corresponds to the degaussing fault), does not align with any of the other clusters, which is desirable in partial transfer tasks. The results obtained indicate that the knowledge transfer methodology is an effective way to deal with the issues that have arisen around partial transfer learning problems.

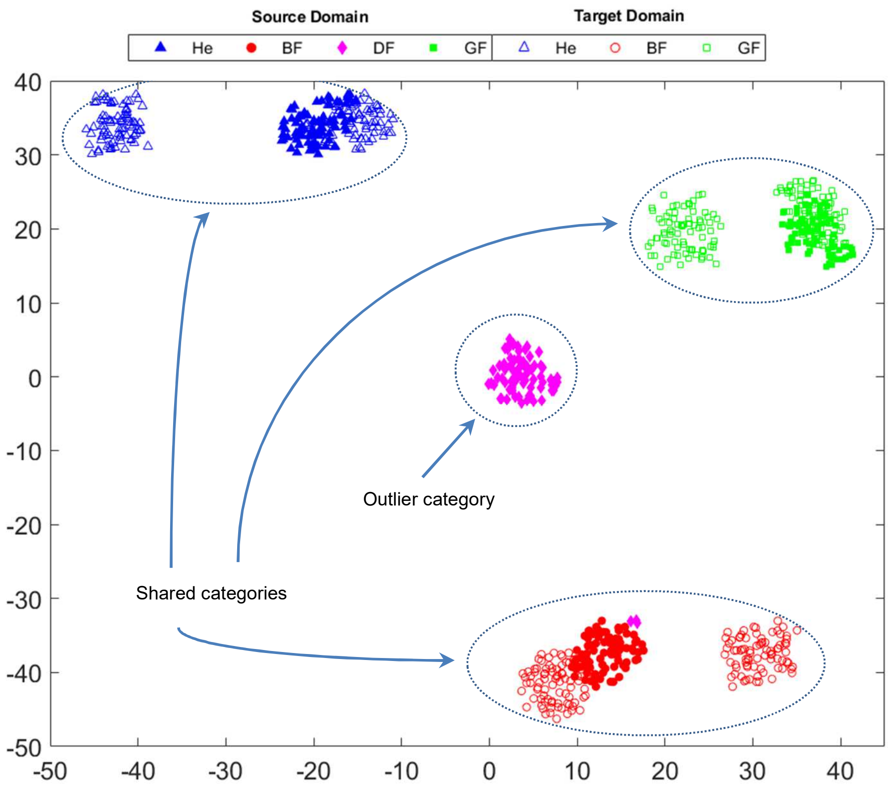


Fig. 5.3.5 Visualization results of the learned deep-features for multiple target instances. It can be seen that the shared categories are aligned with each other, while the instances of the outlier category are set apart without overlapping with the other categories.

5.4 Discussion and conclusions

In this chapter, a novel transfer learning scheme is introduced into the field of intelligent fault diagnosis of industrial rotatory machinery. The new intelligent method is developed based on a domain adaptation strategy and focuses on the partial transfer learning scenario, in which there is one source domain and multiple target domains containing data corresponding to a category subspace of the source domain. The proposed method consists, first, in a pre-training stage that includes two identical blocks of deep-convolutional networks that function as feature learning modules and two fault classifier modules, both trained with the source domain data. Second, the transfer of parameters from the previous modules is carried out and the model is modified by adding a domain discriminator block. Then, the model is trained following a strategy based on adversarial learning to perform domain adaptation between the shared fault categories of the source domain and multiple target domains. During this process, the aim is to reduce the loss due to classification and the loss due to inconsistency, and at the same time maximize the loss of domain classification, which guarantees that domain-invariant and discriminative features with respect to different health conditions are learned. Third, the diagnostic model is tested and its generalizability for different domains (operating conditions) is evaluated. Additionally, a visual and intuitive evaluation of the features learning process is carried out, through the generation of a low-dimensional mapping.

Extensive experiments have been conducted to evaluate the domain adaptation methodology proposed, validated through acquisitions of an electromechanical system working at different operating conditions and under different fault categories. The experimental results show that through a transfer learning scheme a correct diagnosis of faults between different domains can be performed, obtaining a resulting performance of 99.9% for the source domain instances, 94.18% accuracy for target 1 and 88.67% accuracy for target 2. In addition, following the strategy of adversarial learning and classification based on inconsistency, it is possible to align the category subspace of the target domains precisely and at the same time effectively separate the outlier source categories. Therefore, with the results presented, the effectiveness of the proposed methodology is demonstrated to address the problems surrounding the field of intelligent fault diagnosis. And especially, through transfer learning tools, monitoring schemes are endowed with greater robustness and greater generalization capacity.

As future work, the experimental tests could be extended to evaluate to what extent a generalizable fault diagnosis model can be developed and from what working conditions the scheme degrades its performance. In addition, carry out experimental tests to study the possibility of transferring knowledge from a machine environment to a different or similar one, following the state-of-the-art theory known as cross-machine.

6.

Conclusions and future work

The general conclusions on the main contributions of this thesis research are presented in this chapter, as well as the future lines of research and future work are introduced.

CONTENTS:

6. Conclusions and future work

6.1 Conclusions

This chapter presents the main conclusions in regard with the hypotheses and the proposed objectives of this Thesis, as well as the results obtained during the development of the research.

In this regard, this thesis project addresses condition monitoring in industrial electromechanical systems by proposing methodologies based on deep-learning strategies. Specifically, it contributes to three areas of monitoring: fault diagnosis, anomaly detection, and transfer learning. Specifically, a solution is given to the problem of loss of performance when dealing with industrial electromechanical systems, which involve complex systems made up of various components, operating with multiple working conditions and the presence of different faults.

In this regard, the main hypothesis of this thesis states that through the use of **advanced strategies based on deep-learning algorithms** the performance of detection, diagnosis and knowledge transfer schemes can be improved, in relation to existing traditional schemes. Additionally, the proposed methodologies have to be designed under generalizable principles, that is considering hyperparameters optimization and interpretable learning approaches. Thus, contributing to a massive deployment of improved monitoring schemes in manufacturing environments.

Initially, a review of the state of the art in the field of condition monitoring in industrial electromechanical systems was performed to identify existing problems and propose feasible avenues for improvement. This led to the identification and proposal of three main areas of contribution and the definition of strategies to investigate and develop solutions to current issues, *i.e.* fault diagnosis, anomaly detection, and transfer learning. Throughout the implementation of algorithms and their analysis and validation, several experimental test benches were considered. It is worth mentioning that the composition of the databases is independent of the proposed methodologies, that is, the types of motors (synchronous and non-synchronous), the types of faults and the working conditions. Therefore, no particular considerations are made, but it is intended to address industrial systems so that the proposed solutions are adapted to the system under monitoring, including the acquired signals. In this regard, the objective of the proposed methodologies is not to determine the best acquisition system, types of sensors or placement point, but from an existing monitoring system, to maximize the functionalities to carry out the diagnosis, the detection and learning. Therefore, the specific working conditions, the impact of the signals acquired, the type of components (asynchronous or synchronous motors), the power of the electromechanical chains or even the type of couplings are not part of the scope of this research. Consequently, the proposed methodologies are based on generic procedures and are adaptable to available systems and data.

Summarizing, the research focused on the contribution on these areas:

The first contribution refers to the improvement in fault diagnosis schemes for complex electromechanical systems: *"In order to improve the performance of condition monitoring schemes in electromechanical systems, advanced strategies based on deep learning algorithms can be implemented and adapted through methodological procedures that are generalizable and interpretable"*. During the review of the state of the art it was revealed that due to the complexity in manufacturing environments it is difficult to characterize the various states of condition in rotating systems. Furthermore, the interaction of multiple components, the consideration of different working conditions and the presence of multiple fault states cause a **highly clustered-space**, which results in non-Riemannian spaces, that is, characterizations that are not continuous over time and that each subset of data has a particular dispersion. In addition, it was found that there are no clear procedures for the application of deep-learning algorithms within the diagnostic schemes, and that they have been adapted according to the specific execution, which makes it difficult to apply these schemes in real industrial environments. In this regard, making use of the advantages of deep-learning algorithms, it was proposed to design a diagnostic methodology capable of automating the learning process and extracting characteristic faults information from the acquired signals. Through the proposed methodology, it was possible to achieve accuracy performance in the range of 85% and 100%, for the experimental cases presented, overcome classical methodologies and **improving the interpretability** of the application of schemes based on deep learning.

Furthermore, a deficiency was revealed in terms of the interpretability of the models based on deep-learning, which complicates understanding how the predictions are made and the relationship between the indicators with the fault states. An approach that integrates a validation module of the learning process of the fault diagnosis model was proposed. Therefore, as presented in the hypothesis, through the verification of both relevant and spurious indicators leads to the generation of mostly robust schemes.

The second contribution addresses the problem of dealing with highly clustered spaces to perform anomaly detection: "*The problem of detecting anomalies in highly clustered spaces can be addressed through the application of strategies that improve and compact feature spaces in combination with the advantages offered by deep-learning techniques to learn and reconstruct representations, which allows characterize known modes of operation and help detect outlier conditions"*. The state of the art revealed that there is a difficulty in trying to characterize new conditions of the system. The approaches that exist in this regard to address these problems are commonly called anomaly detection, novelty detection, or outlier detection. In electromechanical systems, anomaly detection is a challenging task, especially since machine faults are not so evident or may present only slight changes in the monitored variables. For this reason, there is increasing attention on deep-learning algorithms to take advantage of the representation functionality they have. Numerous methodologies have been developed using deep representations for clustering around anomaly detection. However, as the distribution of the data increases, the feature space becomes more congregate, making it difficult to detect new scenarios.

It has been shown that a not-compact feature space causes samples of the normal and abnormal class to overlap in the representation. Conversely, a dense and compact representation of features creates a wide margin of separation between normal and abnormal instances, such a representation can significantly improve final classification performance. For this reason, the adoption of strategies that improve and compact the clusters in the feature spaces will improve the application of tools that allow differences between normal conditions and abnormal conditions. Therefore, a **deep-learning-based anomaly detection methodology** was developed that uses automated feature learning capabilities in combination with strategies that improve cluster compactness. The use of deep-learning algorithms allows both the learning of features and achieve a compact representation in the feature space. In addition, it is also possible to use the functionalities of reconstruction to validate that proper characterization is being performed, which helps improve anomaly membership and understand the reasons for overlap between normal and abnormal instances that are more similar. Faced with the practical scenarios proposed (experimental case studies), a detection performance of around 82% and 95% was achieved, considering the known samples and the detection of anomalies (new samples).

The third contribution is regarding improving the robustness of monitoring schemes through incremental learning "*Improving the robustness and generalizability of condition monitoring approaches can be achieved through the adoption of incremental learning tools in combination with transfer learning strategies, in which the convergence of different operating conditions is considered*". Through the review of the state of the art, it was revealed that most of the monitoring schemes maintain high performance while operating under the same working conditions in which the models were trained. In contrast, when there are variations in the working conditions or the fault modes present significant changes such as greater wear or the presence of combined faults, the performance of the diagnostic schemes tends to decrease. Additionally, there is a challenge to deal with the imbalance between the training data set and the test data set, since in real industrial scenarios identical fault categories are not always available.

These types of challenges that involve a change in the operating conditions of the systems are commonly referred to as shift-domain problems. To deal with these drawbacks and provide diagnostic schemes with greater generalization capacity, strategies based on transfer learning have been adopted, which in other fields of application have given promising results. Through transfer learning, the aim is to take advantage of the knowledge gained to solve a task, and apply it to a different but similar task. Accordingly, a strategy based on transfer learning that leverage the advantages of deep-learning in feature representation was developed, with the aim of improving the schemes making them more reliable, robust and applicable. Especially, the problem of imbalance between failure categories and the change of domain caused by the different changes of conditions were addressed. A **transfer learning strategy** called adversarial learning was implemented, with which it is possible to extract domain invariant features while allowing perform an effective fault diagnosis. The implementation of a **methodology**

based on domain adaptation for the diagnosis of faults allows to significantly reduce the negative effects caused by shift-domain problems. Furthermore, strategies were implemented to deal with the imbalance of fault categories, referred to as partial fault diagnosis. Several experiments were performed to validate the proposed methodologies, obtaining satisfactory results that suggest that through domain adaptation it is possible to address current problems around health monitoring in rotating systems in industrial environments.

In conclusion, a complete framework of analysis and study for condition monitoring of industrial rotating systems was developed, applied and validated during the research of this Thesis. The results obtained show that, through the implementation of methodological strategies, some of the drawbacks still present in the state of the art related with industrial monitoring schemes can be solved, providing them with greater functionality and potential by effectively adapting and incorporating tools based on deep-learning. The methodology for the diagnosis of faults, specifically designed to apply a deep-learning procedure that enhances the interpretability of the model, together with the anomaly detection methodology and, finally, the development of strategies based on transfer learning to improve the generalization performance, represent a significant advance in the study of monitoring schemes and expand the state of the art, contributing to the inspiration of future research on this topic.

6.2 Future work

The results presented of this research thesis conducted include the proposal and development of a methodology for fault diagnosis, an anomaly detection scheme and the application of transfer learning strategies to performed a more effective monitoring of rotating systems that overcome the drawbacks presented in the state of the art. The development and adaptation of advanced artificial intelligence algorithms together with the application of guided procedures offer the generation of methods with greater precision, better generalization capabilities and greater scope for application in different industrial environments.

However, despite the successful advances of this and other research work in many applications, there is still a long way to go until monitoring schemes based on intelligent algorithms are widely adopted in practical industry systems. Therefore, certain improvements could be considered to further boost the methodologies resulting from the contributions of this thesis and other works in the state of the art:

- Reproducibility and reliability. Systematically, the generalization performance of fault diagnosis approaches has been improved by taking advantage of numerous algorithms based on artificial intelligence. However, this performance is conditioned under well-defined tasks and working conditions. For a monitoring model, an uncertain change in the input may represent a change that is reflected by a large change in the output. Furthermore, most of the reported works have not been shown to be reproducible due to the complexity of the training processes and the numerous hyperparameters that must be configured. Therefore, it continues to be a constant challenge to improve the stability of monitoring schemes so that they can be reproduced and also to increase the reliability in the application of intelligent algorithms, all of which requires advances not only in the sense of **increasing generalization performance**, but also in **creating application standards**.
- Interpretability and model configuration. One of the persistent flaws in monitoring schemes using algorithms based on deep-learning is that they are perceived as "black boxes". That is to say, it is not clear precisely how the models make the predictions, which leads to generating models that are not very interpretable. This problem hinders the applicability of this type of schemes in real industrial environments and is a low credibility point that must be addressed. Therefore, the design of these models is a crucial step, however there is still no clear procedure to establish the parameters, so the configuration of the models depends largely on the application. Accordingly, more effort should be put into deepening the **understanding of the models**, and thus, increasing the transparency on how the predictions are made. In addition, research should be focused on automating the configuration (**self-configuration**) of hyperparameters following more robust objectives, beyond obtaining high classification performance.

- Computational efficiency. It is well known that the use of algorithms based on deep-learning implies a high use of hardware and software tools due to the processing of large amounts of data that implies a large number of calculations, which is why the application of artificial intelligence models often it is expensive. Besides, computational speed represents a limitation when it comes to the application of real-time monitoring schemes, which is essential to guarantee the safety of manufacturing equipment, achieve early maintenance and avoid system failures. Therefore, work must be done to generate monitoring schemes that work optimally, avoiding **inefficient calculations**, including **self-configuration** tools and **reinforcement learning techniques** to enhance the application of intelligent monitoring schemes in real time.

- Cloud computing integration. Keeping the monitoring schemes on demand will allow to carry out in an optimal way the functionalities of fault diagnosis, detection of anomalies and contribute to incremental learning. This would be achieved through Cloud applications and through the development of cyber-physical architectures. In this way, it would allow the implementation of monitoring schemes in industrial environments with a broader scalability, improving the efficiency and reliability of production processes.

Future work will therefore include all these considerations mentioned above. In order to achieve schemes that are more applicable and more generalizable, that are reliable and safe, and that meet the requirements of the manufacturing industry, it is necessary to make significant efforts and overcome the challenges raised above.

7.

Thesis results dissemination

In this chapter, the direct publications resulting from this Thesis research are summarized, including publications in specialized journals and in international conferences. Additionally, the publications resulting from collaborations in related research projects are exposed.

CONTENTS:

7. Thesis results dissemination

7.1 Publications: Thesis contributions

Dissemination publications directly related to the contribution of this Thesis

Journal publications:

F. Arellano-Espitia, M. Delgado-Prieto, V. Martinez-Viol, J. J. Saucedo-Dorantes, and R. A. Osornio-Rios, "Deep-Learning-Based Methodology for Fault Diagnosis in Electromechanical Systems," *Sensors.*, vol. 20, 3949, 2020.

F. Arellano-Espitia, M. Delgado-Prieto, A.D. Gonzalez-Abreu, J. J. Saucedo-Dorantes, and R. A. Osornio-Rios, "Deep-Compact-Clustering Based Anomaly Detection Applied to Electromechanical Industrial Systems," *Sensors.*, vol. 21, 5830, 2021.

F. Arellano-Espitia, M. Delgado-Prieto, J. J. Saucedo-Dorantes, and R. A. Osornio-Rios, "Adversarial Learning-Based Multi-Target Domain for Partial Fault Diagnosis," *IEEE Transactions on Industrial Informatics*. **Under review.**

Conferences publications:

F. Arellano-Espitia, J. J. Saucedo-Dorantes, M. Delgado-Prieto and R. A. Osornio-Rios, "Autoencoder based feature reduction analysis applied to electromechanical systems condition monitoring," in *Proceedings of the 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Zaragoza, 2019.

F. Arellano-Espitia, A.D. Gonzalez-Abreu, M. Delgado-Prieto, J. J. Saucedo-Dorantes and R. A. Osornio-Rios, "Analysis of Machine Learning based Condition Monitoring Schemes Applied to Complex Electromechanical Systems," in *Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, 2020.

F. Arellano-Espitia, M. Delgado-Prieto, V. Martinez-Viol, A. Fernandez–Sobrino and R. A. Osornio-Rios, "Anomaly Detection in Electromechanical Systems by means of Deep-Autoencoder," in *Proceedings of the 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vasteras, 2021.

F. Arellano-Espitia, M. Delgado-Prieto, V. Martinez-Viol, J. J. Saucedo-Dorantes and R. A. Osornio-Rios, "Diagnosis Electromechanical System by Means CNN and SAE: An Interpretable-Learning Study," in *Proceedings of the 2022 5th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, Coventry, 2022.

F. Arellano-Espitia, M. Delgado-Prieto, Joan Valls Perez, J. J. Saucedo-Dorantes and R. A. Osornio-Rios, "Deep Learning-Based Partial Transfer Fault Diagnosis Methodology for Electromechanical Systems," in *Proceedings of the 2023 28th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sinaia, 2023. **Accepted**.

Book Chapters

F. Arellano-Espitia and L. Ruiz-Soto. "Novel methods based on deep learning applied to condition monitoring in smart manufacturing processes," in New Trends in the Use of Artificial Intelligence for the Industry 4.0. London, IntechOpen, 2020.

7.2 Publications: Collaborations and other works

Dissemination publications related to collaborations in the topic

Journals

J. J. Saucedo-Dorantes, **F. Arellano-Espitia**, M. Delgado-Prieto, and, R. A. Osornio-Rios, “Diagnosis Methodology Based on Deep Feature Learning for Fault Identification in Metallic, Hybrid and Ceramic Bearings,” *Sensors*., vol. 21, 5832, 2021.

Conferences

J. J. Saucedo-Dorantes, R. A. Osornio-Rios, R. J. Romero-Troncoso, M. Delgado-Prieto and **F. Arellano-Espitia**, “Novel condition monitoring approach based on hybrid feature extraction and neural network for assessing multiple faults in electromechanical systems,” in *Proceedings of the 2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED),* Toulouse, 2019.

A.D. Gonzalez-Abreu, J. J. Saucedo-Dorantes, R. A. Osornio-Rios, **F. Arellano-Espitia** and M. Delgado-Prieto, “Deep Learning based Condition Monitoring approach applied to Power Quality,” in *Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, 2020.

7.3 Collaboration in technologic transfer projects

Project Title:	Agrupació emergent Looming Factory
Funding entity:	Generalitat De Catalunya, cofinançat pel Fons Europeu de Desenvolupament Regional de la Unió Europea en el marc del Programa Operatiu FEDER de Catalunya 2014-2020.
Partners:	UPC, UOC UPF, UB, Leitat, Eurecat, CVC, Fundació i2Cat, Fundació CIM-UPC
Duration:	from 01/01/2019 to 31/12/2022
Task description:	Smart factory project. Development of a model based on predictive maintenance and for the prediction of energy consumption of industrial machinery.

References

References

[1] A. Choudhary, D. Goyal, S.L. Shimi, A. Akula, "Condition Monitoring and Fault Diagnosis of Induction Motors: A Review," Arch. Comput. Methods Eng. 26 (4) (2019) 1221–1238.

[2] L. Van De Ginste, E. Aghezzaf, J. Cottyn, "The role of equipment flexibility in Overall Equipment Effectiveness (OEE)-driven process improvement," *Procedia CIRP*, Vol. 107, pp 289-294, 2022.

[3] V. Goman, V. Prakht, V. Kazakbaev, and V. Dmitrievskii, ‘‘Comparative study of energy consumption and CO2 emissions of variable-speed electric drives with induction and synchronous reluctance motors in pump units,’’ *Mathematics*, vol. 9, no. 21, p. 2679, 2021.

[4] W. Vattanapuripakorn, S. Sonsupap, K. Khannam, N. Bamrungwong, P. Kaewkhiaw, J. Sarasamkan, B. Bubphachot, “Advanced Electric Battery Power Storage for Motors through the Use of Differential Gears and High Torque for Recirculating Power Generation,” *Clean Technol*, vol. 4, p. 987-1000, 2022.

[5] A. Richard, X. Roboam, F. Rougier, N. Roux, H. Piquet, "AC Electric Powertrain without Power Electronics for Future Hybrid Electric Aircrafts: Architecture, Design and Stability Analysis". *Appl. Sci.*, vol. 13, p. 672, 2023.

[6] G. J. Chen, J. Ni, T. T. Liu, M. Xu, and L. G. Lin, "Study on Structural Optimization and Servo Control for All-Electric CNC Broaching Machine," In *Applied Mechanics and Materials*, Trans Tech Publications, Ltd., vol. 380–384, pp. 562–566, 2013.

[7] Y. Tian, Chao C., K. Sagoe-Crentsil, J. Zhang, W. Duan, "Intelligent robotic systems for structural health monitoring: Applications and future trends," *Automation in Construction*, vol. 139, p. 104273, 2022.

[8] J. A. Antonino-Daviu, A. Quijano-López, M. Rubbiolo and V. Climente-Alarcon, "Advanced Analysis of Motor Currents for the Diagnosis of the Rotor Condition in Electric Motors Operating in Mining Facilities," in *IEEE Transactions on Industry Applications*, vol. 54, no. 4, pp. 3934-3942, 2018.

[9] L. C. Man Lau, E. C. Sing-Chui, G. C. Wai Man, Y. Xin, K. K. Wai-Ho, K. K. Kwan-Mak, M. T. Yun.Ong, S. W. Law, W. H. Cheung, P. S. Hang-Yung, "A novel image-based machine learning model with superior accuracy and predictability for knee arthroplasty loosening detection and clinical decision making," *Journal of Orthopaedic Translation*, vol. 36, p.p. 77-183, 2022.

[10] Y. Luo, N. Tang, G. Li, J. Tang, C. Chai and X. Qin, "Natural Language to Visualization by Neural Machine Translation," in IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 1, pp. 217-226, 2022.

[11] L. Servadei et al., "Accurate Cost Estimation of Memory Systems Utilizing Machine Learning and Solutions from Computer Vision for Design Automation," in *IEEE Transactions on Computers*, vol. 69, no. 6, pp. 856-867, 2020.

[12] H. Tang, H. Liu, W. Xiao and N. Sebe, "When Dictionary Learning Meets Deep Learning: Deep Dictionary Learning and Coding Network for Image Recognition With Limited Data," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2129-2141, 2021.

[13] D. W. Otter, J. R. Medina and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 604-624, 2021.

[14] S. Aradi, "Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 740-759, 2022.

[15] T. Kano, S. Sakti and S. Nakamura, "End-to-End Speech Translation with Transcoding by Multi-Task Learning for Distant Language Pairs," in *IEEE/ACM Transactions on Audio*, Speech, and Language Processing, vol. 28, pp. 1342-1355, 2020.

[16] V.G. Dhanya, A. Subeesh, N.L. Kushwaha, Dinesh Kumar Vishwakarma, T. Nagesh Kumar, G. Ritika, A.N. Singh, "Deep learning based computer vision approaches for smart agricultural applications," *Artificial Intelligence in Agriculture*, vol. 6, p.p. 211-229, 2022.

[17] A.A. Ansari, S.S. Rautaray, M. Pandey, "Deep Learning Trends and Inspired Systems in Natural Language Processing," *Data Science in Societal Applications*. Studies in Big Data, vol. 114. Springer, Singapore, 2022.

[18] Z. Song, "English speech recognition based on deep learning with multiple features," *Computing*, vol. 102, p.p. 663–682, 2020.

[19] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, Volume 89, p.p. 171-178, 2016.

[20] A. Krizhevsky, I. Sutskever, G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, p.p. 1097–1105, 2012.

[21] L. Liao, W. Jin and R. Pavel, "Enhanced Restricted Boltzmann Machine with Prognosability Regularization for Prognostics and Health Assessment," in *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7076-7083, 2016.

[22] T. Li, W. X. Zheng and C. Lin, "Delay-Slope-Dependent Stability Results of Recurrent Neural Networks," in *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2138-2143, 2011.

[23] G.E. Hinton, S. Osindero, Y.W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput*. vol. 18, p.p. 1527–1554. 2006.

[24] R. Zhang, H. Tao, L. Wu and Y. Guan, "Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions," in *IEEE Access*, vol. 5, pp. 14347-14357, 2017.

[25] A. D. Lewis, K. M. Groth, "Metrics for evaluating the performance of complex engineering system health monitoring models," *Reliability Engineering & System Safety*, vol. 223, p. 108473, 2022.

[26] M.A. Tnani, M. Feil, K. Diepold, "Smart Data Collection System for Brownfield CNC Milling Machines: A New Benchmark Dataset for Data-Driven Machine Monitoring," *Procedia CIRP*, vol. 107, p.p. 131-136, 2022.

[27] D. Zhang, J.J. Tsai, "Machine Learning and Software Engineering," *Software Quality Journal*, vol. 11, p.p. 87–119, 2003.

[28] D.T. Hoang, H.J. Kang, "A survey on Deep Learning based bearing fault diagnosis," *Neurocomputing*, vol. 335, p.p. 327-335, 2019.

[29] Zhao X, Jia M. "A novel unsupervised deep learning network for intelligent fault diagnosis of rotating machinery," *Structural Health Monitoring*. vol. 19, p.p. 1745-1763, 2020.

[30] Y. Jian, X. Qing, L. He, Y. Zhao, X. Qi, M. Du, "Fault diagnosis of motor bearing based on deep learning," *Advances in Mechanical Engineering*. vol. 11, 2019.

[31] Q. Wu, Y. Guo, H. Chen, "Establishment of a deep learning network based on feature extraction and its application in gearbox fault diagnosis," *Artif. Intell.*, vol 52, p.p. 125–149, 2019.

[32] Y. Zeng, Y. Guo and J. Li, "Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning," *Neural Comput. & Applic.*, vol, 34, p.p. 2691–2706, 2022.

[33] J. White, T. Kameneva and C. McCarthy, "Vision Processing for Assistive Vision: A Deep Reinforcement Learning Approach," in *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 1, pp. 123-133, 2022.

[34] D. W. Otter, J. R. Medina and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604-624, 2021.

[35] Y. Li, J. Wang, J. Huang, Y. Li, "Research on Deep Learning Automatic Vehicle Recognition Algorithm Based on RES-YOLO Model," *Sensors*, vol. 22, 3783, 2022.

[36] L. Duan, Y. Wang, J. Li, "Exploring the clinical diagnostic value of pelvic floor ultrasound images for pelvic organ prolapses through deep learning," *J. Supercomput*, vol., 77, p.p. 10699–10720, 2021.

[37] C. M. Judd, G. H. McClelland, C. S. Ryan, "Data Analysis: A Model Comparison Approach to Regression, ANOVA, and Beyond", 3er Edition, Harcourt Brace Jovanovich, New York, 1989.

[38] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[39] H. Kamyshanska and R. Memisevic, "The Potential Energy of an Autoencoder," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1261-1273, 1, 2015.

[40] P. Oza and V. M. Patel, "One-Class Convolutional Neural Network," in *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 277-281, 2019.

[41] F. Wu et al., "Regularized Deep Belief Network for Image Attribute Detection," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 7, pp. 1464-1477, 2017.

[42] K. Khalil, O. Eldash, A. Kumar and M. Bayoumi, "Economic LSTM Approach for Recurrent Neural Networks," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 11, pp. 1885-1889, 2019.

[43] R. Vargas, M. Amir, R. Ramon, "Deep learning: A review," In *Advances in Intelligent Systems and Computing*, Preprints: Basel, Switzerland, 2017.

[44] R. Ramachandran, D.C. Rajeev, S.G. Krishnan, P. Subathra, "Deep learning an overview," IJAER, vol. 10, 10, p.p. 25433-25448, 2015.

[45] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," *Nature*, vol. 521, p.p. 436-444, 2015.

[46] B. Bhattacharjee et al., "IBM Deep Learning Service," in *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 10:1-10:11, 1, 2017

[47] E. Haber and L. Ruthotto, "Stable architectures for deep neural networks," *Inverse Problems*, vol. 34, no. 1, pp. 1–23, Dec. 2017.

[48] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," in *IEEE Access*, vol. 6, pp. 14410-14430, 2018.

[49] D. Ryu et al., "DeepRegularizer: Rapid Resolution Enhancement of Tomographic Imaging Using Deep Learning," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1508-1518, 2021.

[50] Y. -H. Tu, J. Du and C. -H. Lee, "Speech Enhancement Based on Teacher–Student Deep Learning Using Improved Speech Presence Probability for Noise-Robust Speech Recognition," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2080-2091, 2019

[51] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang and X. Bai, "TextField: Learning a Deep Direction Field for Irregular Scene Text Detection," in IEEE Transactions on Image Processing, vol. 28, no. 11, pp. 5566-5579, 2019.

[52] I. Goodfellow, Y. Bengio and A. Courville, "Deep Learning", Hardcover, The MIT Press, 2016.

[53] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction", A Bradford Book, The MIT Press, Cambridge, MA, 1998.

[54] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning", *From Natural to Artificial Neural Computation*, pp. 195-201, 1995.

[55] G.E. Hinton, S. Osindero, Y.W. Teh, "A fast learning algorithm for deep belief nets," Neural Comput., vol. 18, p.p. 1527–1554, 2006.

[56] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[57] L. Jing, T. Wang, M. Zhao and P. Wang, "An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox", *Sensors*, vol. 17, no. 414, 2017.

[58] W. Li, Z. Chen and G. He, "A Novel Weighted Adversarial Transfer Network for Partial Domain Fault Diagnosis of Machinery", *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1753-1762, 2021.

[59] S. Sabour, N. Frosst and G. Hinton, "Dynamic routing between capsules", *Proc. Adv. Neural Inf. Process. Syst.*, pp. 3859-3869, 2017.

[60] G. Hinton, Deep Belief Nets. In Encyclopedia of Machine Learning, Sammut, C., Webb, G.I., Eds., Springer: Boston, MA., pp. 267–269, 2011.

[61] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, in *Artificial Intelligence and Statistics*, p.p. 448–455, 2009.

[62] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Computing and Informatics Journal*, vol. 3, 2, p.p. 334-340, 2018.

[63] S. Basodi, C. Ji, H. Zhang and Y. Pan, "Gradient amplification: An efficient way to train deep neural networks," in *Big Data Mining and Analytics*, vol. 3, no. 3, pp. 196-207, 2020.

[64] W.R. Howard, "Pattern Recognition and Machine Learning", *Kybernetes*, vol. 36 no. 2, pp. 275-275, 2007.

[65] M. Xia, T.Li, L. Liu, L. Xu, C. W. Silva, "Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder", *IET Science, Measurement and Technology*, vol. 11, 6, p.p. 687-695, 2017.

[66] H. Shao, H. Jiang, H. Zhao, F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis", *Mechanical Systems and Signal Processing*, vol. 95, p.p.187-204, 2017.

[67] Y. Qu, M. He, J. Deutsch, D. He, "Detection of Pitting in Gears Using a Deep Sparse Autoencoder", *Appl. Sci*. vol. 7, 515, 2017.

[68] E. Principi, D. Rossetti, S. Squartini and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," in *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, p.p. 441-451, 2019.

[69] M. Memarzadeh, B. Matthews, I. Avrekh, "Unsupervised Anomaly Detection in Flight Data Using Convolutional Variational Auto-Encoder", *Aerospace*, 7, 115, 2020.

[70] H. Shao, H. Jiang, H. Zhang and T. Liang, "Electric Locomotive Bearing Fault Diagnosis Using a Novel Convolutional Deep Belief Network," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727-2736, 2018.

[71] C. Lu, Z.Y. Wang, W.L. Qin, J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification", *Signal Processing*, vol. 130, p.p. 377-388, 2017.

[72] J. Sun, C. Yan and J. Wen, "Intelligent Bearing Fault Diagnosis Method Combining Compressed Data Acquisition and Deep Learning," in *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185-195, 2018.

[73] J. Dai, J. Tang, F. Shao, S. Huang, Y. Wang, "Fault Diagnosis of Rolling Bearing Based on Multiscale Intrinsic Mode Function Permutation Entropy and a Stacked Sparse Denoising Autoencoder", *Appl. Sci*. vol. 9, p. 2743, 2019.

[74] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification", *Measurement*, vol. 89, p.p. 171-178, 2016.

[75] C. Lu, Z. Wang, B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification", *Advanced Engineering Informatics*, vol. 32, p.p. 139-151, 2017.

[76] W. Sun, B. Yao, N. Zeng, B. Chen, Y. He, X. Cao, W. He, "An Intelligent Gear Fault Diagnosis Methodology Using a Complex Wavelet Enhanced Convolutional Neural Network," *Materials*, vol. 10, 790, 2017.

[77] L. Wen, X. Li, L. Gao and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990-5998, 2018.

[78] D. Verstraete, A. Ferrada, E. López-Droguett, V. Meruane, M. Modarres, "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings", *Shock and Vibration*, vol. 2017, 5067651, 17 pages, 2017.

[79] S. Li, G. Liu, X. Tang, J. Lu, J. Hu, "An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis," *Sensors*, vol. 17, 1729, 2017.

[80] L. Jing, T. Wang, M. Zhao, P. Wang, "An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox," *Sensors*, vol. 17, 414, 2017.

[81] T. Ince, S. Kiranyaz, L. Eren, M. Askar and M. Gabbouj, "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks," in *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067-7075, 2016.

[82] L. Jing, M. Zhao, P. Li, X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox", *Measurement*, vol. 111, p.p. 1-10, 2017.

[83] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, vol. 100, p.p. 439-453, 2018.

[84] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, "A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals," *Sensors*, vol. 17, 425, 2017.

[85] W. Sun, R. Zhao, R. Yan, S. Shao and X. Chen, "Convolutional Discriminative Feature Learning for Induction Motor Fault Diagnosis," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1350-1359, 2017.

[86] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14 (8), p.p. 1771–1800, 2002.

[87] T. Jie, L. Yilun, Y. Dalian, "Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion," *Shock and Vibration, Hindawi*, vol. 2016, 2016.

[88] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. E. Vasquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76–77, p.p. 283-293, 2016.

[89] H. Shao, H. Jiang, F. Wang, Y. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Transactions*, vol. 69, p.p. 187-201, 2017.

[90] SY. Shao, WJ. Sun, RQ. Yan, "A Deep Learning Approach for Fault Diagnosis of Induction Motors in Manufacturing," *Chin. J. Mech. Eng*. vol. 30, p.p.1347-1356, 2017.

[91] H. Oh, J. H. Jung, B. C. Jeon and B. D. Youn, "Scalable and Unsupervised Feature Engineering Using Vibration-Imaging and Deep Learning for Rotor System Diagnosis," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3539-3549, 2018.

[92] J. Deutsch and D. He, "Using Deep Learning-Based Approach to Predict Remaining Useful Life of Rotating Components," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 1, pp. 11-20, 2018.

[93] Z. Chen, S. Deng, X. Chen, C. Li, R. V. Sanchez, H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectronics Reliability*, vol. 75, p.p. 327-333, 2017.

[94] K. Khalil, O. Eldash, A. Kumar and M. Bayoumi, "Economic LSTM Approach for Recurrent Neural Networks," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 11, pp. 1885-1889, 2019.

[95] H. M. Lynn, S. B. Pan and P. Kim, "A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks," in *IEEE Access*, vol. 7, pp. 145395-145405, 2019.

[96] E. Messner, M. Zöhrer and F. Pernkopf, "Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964-1974, 2018.

[97] C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," in *IEEE Access*, vol. 5, pp. 21954-21961, 2017.

[98] M. Yuan, Y. Wu and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," 2016 *IEEE International Conference on Aircraft Utility Systems (AUS)*, Beijing, China, 2016, pp. 135-140.

[99] R. Zhao, R. Yan, J. Wang, K. Mao, "Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks", *Sensors*, vol. 17, 273, 2017.

[100] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen and J. Wang, "Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539-1548, 2018.

[101] H. Liu, J. Zhou, Y. Zheng, W. Jiang, Y. Zhang, “Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders”, *ISA Transactions*, vol. 77, p. p. 167-178, 2018.

[102] C. Chuang, L. Ningyun, J. Bin and X. Yin, "Condition-based maintenance optimization for continuously monitored degrading systems under imperfect maintenance actions," in *Journal of Systems Engineering and Electronics*, vol. 31, no. 4, pp. 841-851, 2020.

[103] A. H. Sodhro, S. Pirbhulal and V. H. C. de Albuquerque, "Artificial Intelligence-Driven Mechanism for Edge Computing-Based Industrial Applications," in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4235-4243, 2019.

[104] H. A. Hussein, M. E. Ammar and M. A. Hassan, "Three Phase Induction Motor's Stator Turns Fault Analysis Based on Artificial Intelligence", *International Journal of System Dynamics Applications* (IJSDA), vol. 6(3), p.p., 1-19, 2017.

[105] Y. Zhang, Z. Wang, H. Wang and F. Blaabjerg, "Artificial Intelligence-Aided Thermal Model Considering Cross-Coupling Effects," in *IEEE Transactions on Power Electronics*, vol. 35, no. 10, pp. 9998-10002, 2020.

[106] S. T. Spantideas, A. E. Giannopoulos, N. C. Kapsalis and C. N. Capsalis, "A Deep Learning Method for Modeling the Magnetic Signature of Spacecraft Equipment Using Multiple Magnetic Dipoles," in *IEEE Magnetics Letters*, vol. 12, pp. 1-5, 2021.

[107] S. Kumar, A. Pandey, K. S. Satwik, S. Kumar, S. K. Singh, A. K. Singh, A. Mohan, "Deep learning framework for recognition of cattle using muzzle point image pattern", *Measurement*, vol. 116, p.p. 1-17, 2018.

[108] W. X. Liao et al., "Automatic Identification of Breast Ultrasound Image Based on Supervised Block-Based Region Segmentation Algorithm and Features Combination Migration Deep Learning Model," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 984-993, 2020.

[109] D. W. Otter, J. R. Medina and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604-624, 2021.

[110] B. Zhang, D. Xiong and J. Su, "Neural Machine Translation with Deep Attention," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 154-163, 1, 2020.

[111] J. Liu et al., "Toward Robust Fault Identification of Complex Industrial Processes Using Stacked Sparse-Denoising Autoencoder with Softmax Classifier," in *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 428-442, 2023.

[112] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, "Deep learning and its applications to machine health monitoring", *Mechanical Systems and Signal Processing*, vol. 115, p.p., 213-237, 2019.

[113] J. Wang, P. Fu, L. Zhang, R. X. Gao and R. Zhao, "Multilevel Information Fusion for Induction Motor Fault Diagnosis," in *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 5, pp. 2139-2150, 2019.

[114] P. Gangsar, R. Tiwari, "Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms", *Mechanical Systems and Signal Processing*, vol. 94, pp., 464-481, 2017.

[115] X. Jiang, C. Xu, "Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data", *J. Clin. Med.*, vol. 22, 11, 5772, 2022.

[116] D. Zhang, Y. Chen, F. Guo, H. R. Karimi, H. Dong and Q. Xuan, "A New Interpretable Learning Method for Fault Diagnosis of Rolling Bearings," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021.

[117] J.S. Smith, "The local mean decomposition and its application to EED perception data," *J. R. Soc. Interface*, vol. 2, p.p., 443–454, 2005.

[118] R.O., Duda, P.E., Hart, D.G., Stork, "Pattern Classification", 2nd ed., Wiley: Hoboken, NJ, USA, 2000.

[119] C. H. Chang, "Deep and Shallow Architecture of Multilayer Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2477-2486, 2015.

[120] M.A. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, "A review of novelty detection", *Signal Process*, vol. 99, p.p., 215–249, 2014.

[121] F. Chan, Z.X. Wang, S. Patnaik, M.K. Tiwari, X.P. Wang, J.H. Ruan, "Ensemble-learning based neural networks for novelty detection in multi-class systems", *Applied Soft Computing*, vol. 93, p., 106396, 2020.

[122] C.C. Aggarwal, "Linear models for outlier detection," In *Outlier Analysis*, 2nd ed., *Springer Nature*: New York, NY, USA, pp. 65–111, 2017.

[123] V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey", ACM Comput. Surv., vol. 41, p., 1–58, 2009.

[124] 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 International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives* (SDEMPED), Toulouse, France, pp. 330-337, 2019.

[125] J. Guo, G. Liu, Y. Zuo and J. Wu, "An Anomaly Detection Framework Based on Autoencoder and Nearest Neighbor," 2018 15th *International Conference on Service Systems and Service Management* (ICSSSM), Hangzhou, China, pp. 1-6, 2018.

[126] Z. Yang, D. Gjorgjevikj, J. Long, "Sparse Autoencoder-based Multi-head Deep Neural Networks for Machinery Fault Diagnostics with Detection of Novelties," *Chin. J. Mech. Eng*. vol. 34, 54, 2021.

[127] T. Ergen and S. S. Kozat, "Unsupervised Anomaly Detection with LSTM Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 3127-3141, 2020.

[128] Y. Bengio, A. Courville, P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach*. Intell., vol. 35, p.p., 1798–1828, 2013.

[129] M. Ribeiro, A. E.Lazzaretti, H. S. Lopes, "A study of deep convolutional auto-encoders for anomaly detection in videos", *Pattern Recognition Letters*, vol. 105, p.p., 13-22, 2018.

[130] X. Peng, S. Xiao, J. Feng, W.Y. Yau, Z. Yi, "Deep subspace clustering with sparsity prior," In *Proceedings of the International Joint Conference on Artificial Intelligence*, New York, NY, USA, 9–16 July 2016, pp. 1925–1931.

[131] F. Tian, B. Gao, Q. Cui, E. Chen, T.Y. Liu, "Learning deep representations for graph clustering," In *Proceedings of the AAAI Conference on Artificial Intelligence*, Québec City, QC, Canada, 27-31 July 2014, pp. 1293–1299.

[132] J. Xie, R.B. Girshick, A. Farhadi, "Unsupervised deep embedding for clustering analysis", In *Proceedings of the International Conference on Machine Learning*, New York, NY, USA, 19–24 June 2016; pp. 478–487.

[133] M. Sakow, K. Marchelek, "Design and optimisation of regression-type small phase shift FIR filters and FIR-based differentiators with optimal local response in LS-sense", *Mechanical Systems and Signal Processing*, vol. 152, 107408, 2021.

[134] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang and T. Zhang, "Deep Model Based Domain Adaptation for Fault Diagnosis," in *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296-2305, 2017.

[135] D.P. Kingma, J. Ba, "ADAM: A method for stochastic optimization", In Proceedings of the 3rd International Conference for Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.

[136] X. Glorot, Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," In Proceedings of the 13th *International Conference on Artificial Intelligence and Statistics, Sardinia*, Italy, 13–15 May 2010; pp. 249–256.

[137] M. Ribeiro, M. Gutoski, A. E. Lazzaretti and H. S. Lopes, "One-Class Classification in Images and Videos Using a Convolutional Autoencoder with Compact Embedding," in *IEEE Access*, vol. 8, pp. 86520-86535, 2020.

[138] L. Van der Maaten, G.E. Hinton, *Visualizing High-Dimensional Data Using t-SNE*, *J. Mach. Learn. Res*., vol. 9, p.p. 2579–2605, 2008.

[139] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt, "Support vector method for novelty detection". In: *Proceedings of the 12th International Conference on Neural Information Processing Systems*. 1999, 582–588.

[140] X. Zhang, Y. Hu, J. Deng, H. Xu and H. Wen, "Feature Engineering and Artificial Intelligence-Supported Approaches Used for Electric Powertrain Fault Diagnosis: A Review," in *IEEE Access*, vol. 10, pp. 29069-29088, 2022.

[141] Z. Zhao et al., "Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-28, 2021.

[142] Z. He, H. Shao, X. Zhang, J. Cheng and Y. Yang, "Improved Deep Transfer Auto-Encoder for Fault Diagnosis of Gearbox Under Variable Working Conditions With Small Training Samples," in *IEEE Access*, vol. 7, pp. 115368-115377, 2019.

[143] X. Li, W. Zhang, N. X. Xu and Q. Ding, "Deep Learning-Based Machinery Fault Diagnostics With Domain Adaptation Across Sensors at Different Places," in *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6785-6794, 2020.

[144] L. Li, Y. Huang, Y. Han and J. Jiang, "Use of deep learning genomics to discriminate Alzheimer's disease and healthy controls," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico, 2021, pp. 5788-5791.

[145] B. Z. Hussain, I. Andleeb, M. S. Ansari, A. M. Joshi and N. Kanwal, "Wasserstein GAN based Chest X-Ray Dataset Augmentation for Deep Learning Models: COVID-19 Detection Use-Case," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 2058-2061.

[146] Y. Panagakis et al., "Tensor Methods in Computer Vision and Deep Learning," in Proceedings of the IEEE, vol. 109, no. 5, pp. 863-890, May 2021.

[147] D. Tran, H. Nguyen, B. Tran, Fast and precise single-cell data analysis using a hierarchical autoencoder. Nat Commun 12, 1029, 2021.

[148] D. J. Chalmers, R. M. French, D. R. Hofstadter, "High-level perception, representation, and analogy: A critique of artificial intelligence methodology", in Journal of Experimental & Theoretical Artificial Intelligence, vol. 4, pp.185-211, 2007.

[149] L. Shao, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 5, pp. 1019–1034, 2014.

[150] J. Zhang, W. Li and P. Ogunbona, "Joint Geometrical and Statistical Alignment for Visual Domain Adaptation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5150-5158.

[151] S. J. Pan, I. W. Tsang, J. T. Kwok and Q. Yang, "Domain Adaptation via Transfer Component Analysis," in *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199-210, 2011.

[152] C. Tan, F. Sun, T. Kong, "A survey on deep transfer learning", Proc. Int. Conf. Artif. Neural Netw. pp. 270–279, 2018.

[153] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, K. Gryllias, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges", *Mechanical Systems and Signal Processing*, vol. 167, 2022.

[154] Y. Pan, F. Mei, H. Miao, "An Approach for HVCB Mechanical Fault Diagnosis Based on a Deep Belief Network and a Transfer Learning Strategy", *J. Electr. Eng. Technol*. vol. 14, pp., 407–419, 2019.

[155] Y. Song, Y. Li, L. Jia and M. Qiu, "Retraining Strategy-Based Domain Adaption Network for Intelligent Fault Diagnosis," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6163-6171, 2020.

[156] Z. Zhang, H. Chen, S. Li, Z. An, "Unsupervised domain adaptation via enhanced transfer joint matching for bearing fault diagnosis", *Measurement*, vol. 165, 108071, 2020.

[157] S. Shao, S. McAleer, R. Yan and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446-2455, 2019.

[158] T. Lu, F. Yu, B. Han and J. Wang, "A Generic Intelligent Bearing Fault Diagnosis System Using Convolutional Neural Networks with Transfer Learning," in *IEEE Access*, vol. 8, pp. 164807-164814, 2020.

[159] T. Han, C. Liu, W. Yang, D. Jiang, "Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions", *ISA Transactions*, vol. 93, p.p. 341-353, 2019.

[160] B. Zhao, X. Zhang, Z. Zhan, S. Pang, "Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains", *Neurocomputing*, vol., 407, p.p. 24-38, 2020.

[161] H. Shao, M. Xia, G. Han, Y. Zhang and J. Wan, "Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3488-3496, 2021.

[162] B.K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, G.R. Lanckriet, "Hilbert space Embeddings and metrics on probability measures", *J. Mach. Learning*, vol. 99, p.p., 1517–1561, 2010.

[163] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, M. "Kawanabe, Direct importance estimation for covariate shift adaptation", *Ann. Inst. Stat. Math.*, vol. 60, 4, pp. 699–746, 2008.

[164] B. Sun, K. Saenko, "Deep Coral: Correlation alignment for deep domain adaptation", in *Proc. European Conf. Computer Vision*, pp. 443–450, 2016.

[165] M. Arjovsky, S. Chintala, L. Bottou, "Wasserstein generative adversarial networks", in *Proceedings of the 34th International Conference on Machine Learning (PMLR)*, pp. 214–223, 2017.

[166] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville0, Y. Bengio, "Generative adversarial networks", *Commun ACM*, vol. 63(11), pp. 139–144, 2014.

[167] Y. Ganin, V. Lempitsky, "Unsupervised domain adaptation by backpropagation", in *Proceedings of the 32nd International Conference on International Conference on Machine Learning (ICML)*, vol. 37, pp. 1180–1189, 2015.

[168] 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.

Annexes

The annexes include the descriptions of the experimental test benches used for the development, application and validation of the thesis proposal.

CONTENTS:

A.I Multi-Fault Experimental Test Bench

In order to validate the proposed methodologies of this thesis, an experimental electromechanical system is monitored and data are acquired from this. The experimental test bench diagram is shown in **Fig. AI. 1.** This test bench consists on a kinematic chain composed by two identical facing motors, *i.e.*, one motor under monitoring and other motor that works as a load. Both motors are connected via a screw linked to a gearbox. The screw has a moving part that is driven by the output shaft of the gearbox. The motor that provides the rotation is driven by ABB power converters, model ACSM1. The two motors are of the surface permanent magnet synchronous motor (SPMSM) type, which have three pairs of pole with a rated torque of 3.6 Nm, 230 Vac, and rated speed of 6000 rpm provided by ABB Group.

The measurement equipment is focused on the acquisition of stator currents and vibrations. The three stator currents were acquired using Tektronix current probes placed at the output of the stator phases of the power converter. The vibration signal is acquired from the perpendicular plane of the motor axis, using a tri-axial accelerometer ENDEVCO Isotron KS943B.10. The measurements were done to a PXIe 1062 acquisition system provided by national instruments (NI). Sampling frequency is set to 20 kS/s during 1 s for each experiment.

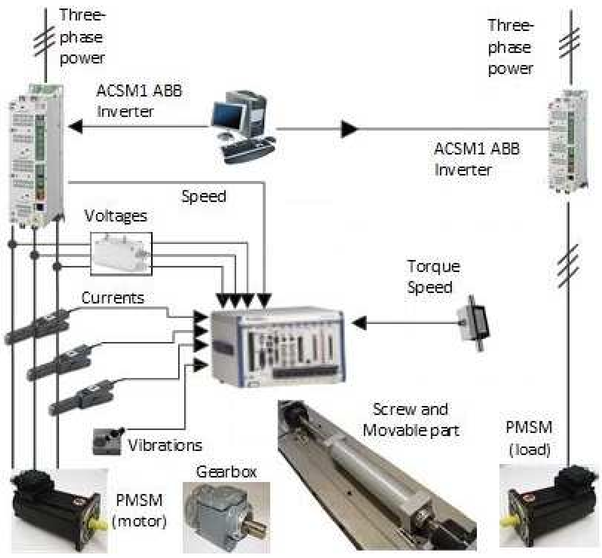
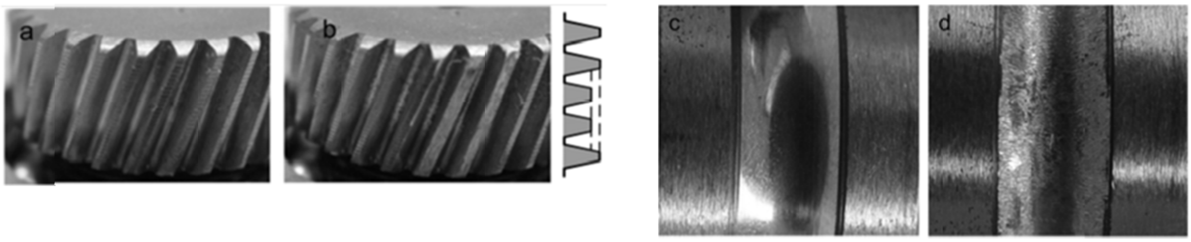


Fig. AI.1. Experimental electromechanical test bench used for experimental validation of the methodologies.

A set of five health categories have been considered, including the normal state (no fault) and four types of fault states. First of all, data was acquired from complete healthy electromechanical actuator. Second, a motor with an induced demagnetization corresponding to 20% nominal flux reduction in a pair of poles was manufactured. Third, widespread degradation is generated in the inner and outer raceways of a ball bearing in order to cause a generalized rough defect. Fourth, a degradation was made to two teeth of the gearbox in order to produce a smoothing.

The wear produced to smooth two teeth of the gearbox is shown in Fig. AI.2 (a). In addition, the degradation made to the metallic bearing is shown in Fig. AI.2 (b).



(a) **(b)**

Fig. AI.2. (a.b**)** Appearance of healthy gear teeth, (a.b) Appearance of damaged gear teeth. (b.c) Appearance of healthy bearing inner-race, (b.d) Appearance of damaged bearing inner-race.

The acquisitions were collected at different operating conditions corresponding to power frequency low, power frequency high (30 and 60 Hz), motor load low and motor load high (40 and 75% of the nominal load). Thus, for each of the five health states there are acquisitions of four operating conditions: low power frequency—low motor load (C1), low power frequency—high motor load (C2), low power frequency— high—motor load low (C3) and power frequency high—motor load high (C4).

A.II Rolling Bearing Faults Experimental Test Bench

The bearing fault test bench experimental database is provided by the center of the Case Western Reserve University (CWRU) [168]. The experimental test bench, is shown in **Fig. AII.1**. It is composed by a 2-hp motor, a torque encoder, a dynamometer, and control electronic. The data sets are composed of different vibration measurements generated from the experimental bench. A total of five health states are presented, which include normal state and four faults induced separately at the inner raceway, rolling element, and outer raceway. For each of the faults, three types of severity were induced with a single point with diameters of 0.007, 0.014, and 0.021 inches.

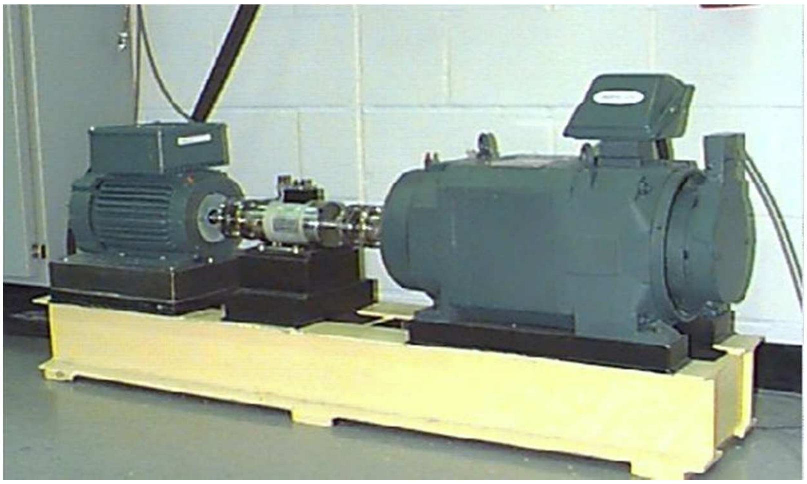


Fig. AII.1. Overview of experimental bearing fault test bench provided by CWRU.

The vibration data are collected by some accelerometers attached to the housing with magnetic bases. Each vibration data of the healthy and fault categories was acquired under different operating conditions corresponding to various motor loads of 0, 1, 2, and 3 hp (engine speeds from 1720 to 1797 RPM). In this experiment, the vibration signals are from the end of the drive, and all data was acquired at a 12 kHz sampling rate. It took 10 seconds of each condition to obtain 40 volumes of 250ms with 3000 samples which were used to validate part of the proposed methodologies.

A.III Pulley-Belt Electromechanical Test Bench

The experimental electromechanical system based on a pulley-belt system is shown in **Fig. AIII.1**. It is composed by a 971-W three-phase induction motor model WEG00136APE48T, the motor has one pair of poles and supports 220 VAC as a power supply. The control of the motor speed is carried out by means of a variable frequency driver model WEGCFW08. Furthermore, the motor is coupled by means of a pulley-belt system to an ordinary alternator. The automotive alternator works as a mechanical load and generates a nominal load with variations between the range of 25% to 35% on the induction motor.

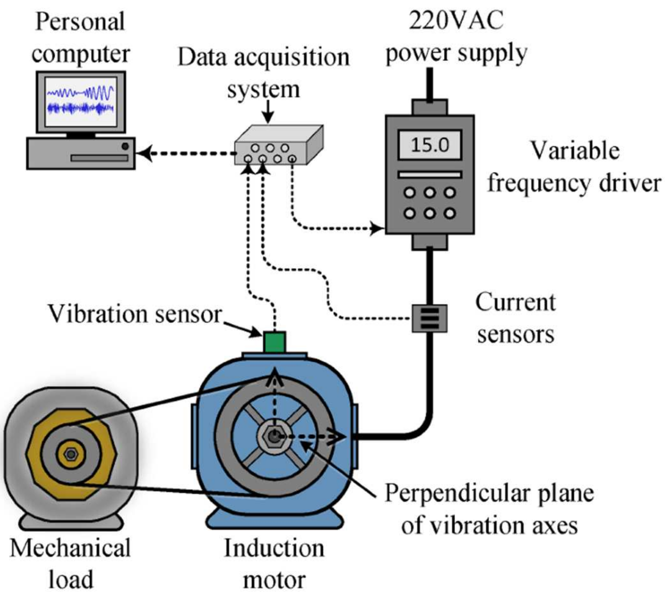
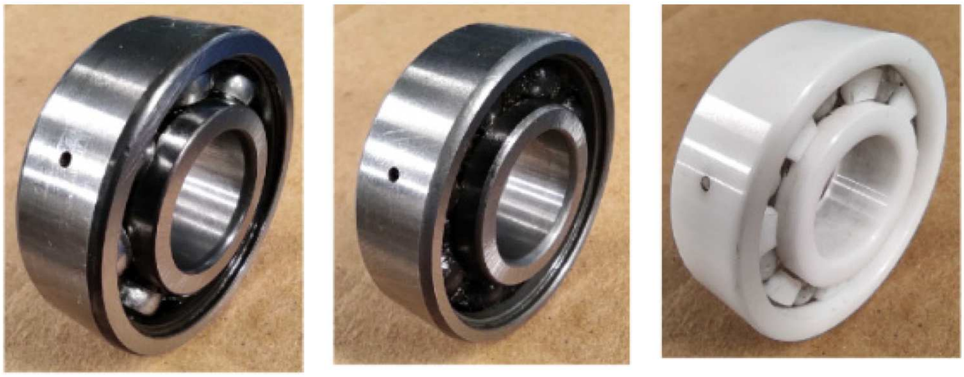


Fig. AIII.1. Scheme of the experimental setup based on a pulley-belt system.

The data acquisition is carried out by means of a data acquisition system based on FPGA (field-programmable gate array), collecting data from two vibration axes and a stator current. The vibration signals are acquired by means of an accelerometer model LIS3L02AS4, while the stator current is measured by means of a hall effect sensor of Tamura Corporation model L08P050D15. The sampling frequency to collect the vibration signals and the stator current was set at 3000 Hz. For each experiment carried out, 300 s of continuous operation are acquired and stored in a database for further processing.

The experimental objective of this test bench is to evaluate the different bearing technologies. In this regard, three types of bearings are evaluated including: metal bearings, ceramic bearings, and hybrid bearings (metal races and ceramic balls). Two health states are evaluated: the normal state and fault in the outer race. The defects in the three bearings are fabricated generating artificial damage by drilling a hole with a tungsten drill bit of 1/16 of diameter.

The damaged bearings can be seen in **Fig. AIII.2**, corresponding to the three types of materials. Data acquisitions are made for each of the bearings under different supply frequencies, that is, 5 Hz, 15 Hz, 50 Hz and 60 Hz.



(a) (b) (c)

Fig. AIII.2. Appearance of bearings damaged in the outer race: (a) metallic bearing, (b) hybrid bearing, (c) full ceramic bearing.