Chapter

Trends of Digital Transformation in the Shipbuilding Sector

Alejandro Sánchez-Sotano, Alberto Cerezo-Narváez, Francisco Abad-Fraga, Andrés Pastor-Fernández and Jorge Salguero-Gómez

Abstract

The new paradigms of Industry 4.0 force all the industrial sectors to face a deep digital transformation in order to be on the edge in a competitive and globalized scenario. Following this trend, the shipbuilding industry has to establish its own path to adapt itself to the digital era. This chapter aims to explore this challenge and give an outlook on the multiple transformative technologies that are involved. For that reason, a case of study is presented as a starting point, in which the digital technologies that can be applied are easily recognized. A social network analysis (SNA) is developed among these key enabling technologies (KETs), in order to stress their correlations and links. As a result, artificial intelligence (AI) can be highlighted as a support to the other technologies, such as vertical integration of naval production systems (e.g., connectivity, Internet of things, collaborative robotics, etc.), horizontal integration of value networks (e.g., cybersecurity, diversification, etc.), and life cycle reengineering (e.g., drones, 3D printing (3DP), virtual and augmented reality, remote sensing networks, robotics, etc.).

Keywords: digital transformation, key enabling technologies, shipbuilding 4.0, Industry 4.0, artificial intelligence, complex projects

1. Introduction

In the twenty-first century, industrial organizations are expanding their business lines to offer maintenance, repair, and checkup services related to their products, as well as technical support, and are paying more and more attention to these services [1]. In this environment, shipyards nowadays comprise of designing, engineering and building, procurement and logistics, assembling and commissioning, as well as maintaining and repairing and transforming and advancement of vessels and marine equipment, among many others.

Ships, ferries, and offshore platforms are complex products with long service lives and high costs of construction, manning, operating, maintaining, and repairing [2]. In addition, these are usually built to order and involve complex production processes, with large-scale but short series production, high degree of customization, and intensive labor. In return, they provide high value-added but requiring large and fixed capital investments although they have long life cycles [3]. However, most of them do not always evolve in line with the development of the latest technology [4].

Due to the aforementioned context, productivity in shipbuilding sector is developing slower than other manufacturing industries. Many factors may be identified as the root causes for this lack of timing, as companies are focusing on their short-term profits, usually ignoring outside benchmarks. This creates a barrier to change, in addition of conservative regulations, that makes difficult the entry of disruptive innovations, causing a lack in terms of competitiveness [5].

This lack of productivity, which affects project-based industries (as shipbuilding), has been steadily discussed by both academics and practitioners [6–8], which have been suggesting and proposing measures to increase their performance. At an early stage, innovative working methods from better organization of the processes are involved [9], such as the promotion of a more efficient split of work in order to improve the coordination within and across companies involved through the supply chain [10]. Then, due to the introduction of the Industry 4.0 paradigm, emerging technological capacities, to design better products, improve the efficiency of their services, and offer new value-added processes, were applied. As a consequence, self-managed processes, people, machines, and systems are communicating and cooperating [11].

To achieve the Industry 4.0 paradigm, a number of key enabling technologies (KETs) are used. These technologies, both from real and virtual world, were first described by the Boston Consulting Group [12]. With the aim of transforming the current production system, technologies like autonomous robots, additive manufacturing, horizontal and vertical integration, Big Data, Internet of things, cybersecurity, cloud, augmented reality, and simulation were included.

In addition to the initial set of KETs, other technologies, such as autonomous guided vehicles [13], blockchain (BCH) [14], or artificial intelligence (AI) [15], own a great potential to be crucial in the digital transformation of industries. Particularly, a European Commission report [16] arises the AI as a transverse technology both to be applied in software-based systems (virtual world) and be embedded in hardware devices (real world). Using data gathered from the available sources, the integration of the AI with the other KETs will improve overall performance through better automatic decision-making based on analyzed data.

This chapter is structured as follows: Section 2 presents the objectives of the research. Section 3 develops the literature review. Section 4 relates the research method. Section 5 describes its implementation in a case study. Section 6 shows its findings, discussing the results obtained. Section 7 concludes the chapter, summarizing the contributions and proposing further research.

2. Objectives

The main purpose of this research is to explore the challenge of facing a deep digital transformation by the shipbuilding industry, in order to be on the edge in a competitive and globalized scenario. This chapter also aims to give an outlook on the multiple and transformative technologies that are involved, analyzing the importance of the digital transformation (digitalization, automation, exploitation, and integration) in complex projects and its application in the context of Industry 4.0, discussing the results of its potential implantation.

For that reason, a case of study is presented as a starting point, in which digital technologies applied are recognized. Afterwards, a social network analysis (SNA) is developed, in order to highlight the correlations and links between KETs, aiming to confirm the AI as a support to the others. Among those, vertical integration of production systems, horizontal integration of value networks, and life cycle reengineering are stressed. The research framework is summarized in **Figure 1**.

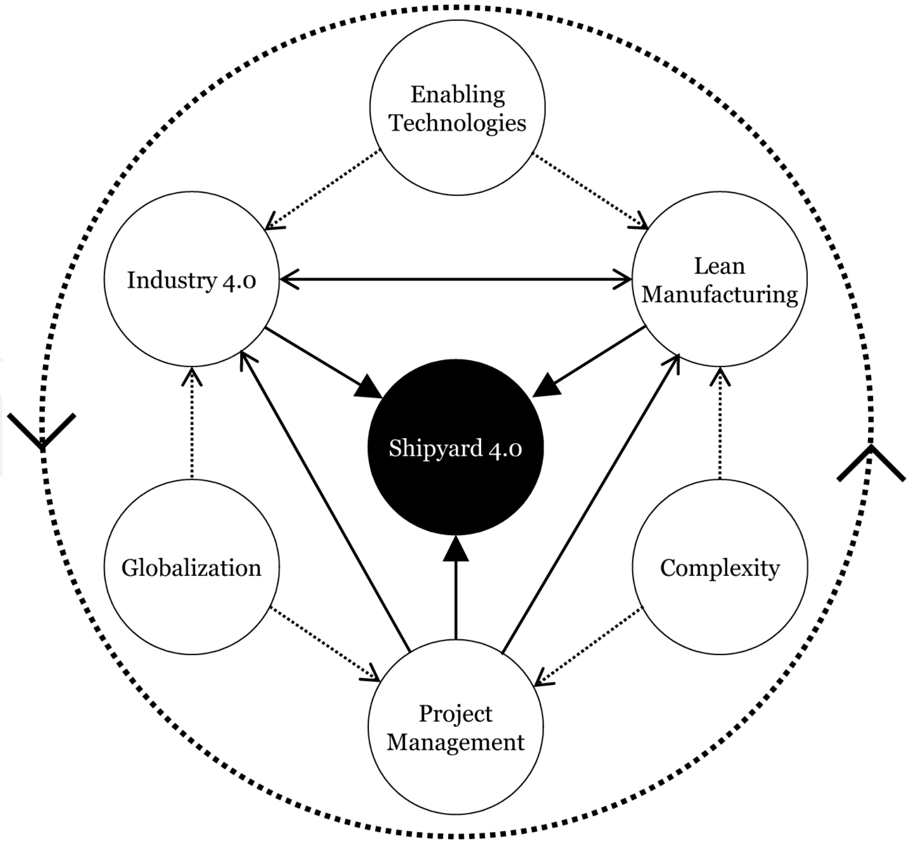

Figure 1.
Research framework.

3. Literature review

The shipbuilding sector is characterized by complex manufacturing processes, with a wide range of involved elements, low-volume serial production, and results of a high added value [17]. Faced with unpredictable conditions and intense competitors, the sector is forced to restructure its long-term objectives [18], as the most dynamic shipyards, which show a greater adaptation to the global market, get better results. In order to achieve this, they adopt research, development, and innovation (RDI) philosophies, launching bold business initiatives to counter these uncertainties using technology-driven practices that create infrastructure and empowerment, preparing them for the upcoming challenges [19].

3.1 Complexity in shipbuilding projects

Complexity is the property of projects that make them difficult to understand, foresee, and keep under control their overall behavior, even when given reasonably complete information about the system [20]. Every project has a degree of complexity, becoming one of the most important factors of their failure. Furthermore, project complexity presents additional challenges to achieve objectives, although some significant indicators can be chosen to measure and asses it [21], such as compliance and authorization, project organization, targets, resources, change orders, technology familiarity, and location, among others.

The two most common types of complexity within projects concern the organization and the technology [22]. Organizational complexity is caused by the engagement of several diverse and separate organizations for a limited period of time (both suppliers and consultants as well as temporary structures to manage the projects), depending on the hierarchical structures and organizational units [23].

In contrast, technological projects depend on the result produced, mainly due to the diversity of tasks [24]. Furthermore, although complexity is usually expressed by the means of cost, duration, or people involved, these criteria do not correlate well with how they are managed [25].

In summary, complex projects consist of ambiguity and uncertainty, interdependency, nonlinearity, unique local conditions, autonomy, emergent behaviors, and unfixed boundaries. According to these properties, projects can be classified as simple, complicated, chaotic, and complex [26]. On the other hand, complex projects are also influenced by significant external changes [27], from misaligned stakeholders' view of success, in which current tools and decision processes are unsuitable for analyze it. To respond positively to this complexity, it is necessary to imply both organizations and practitioners [28].

It can be noticed that complex projects undertaken by traditional methods, practices, and frameworks usually result inadequate in terms of scale, rate of change, heterogeneity, multiple pathways, and ambiguous objectives [29]. In this context, project management decouples and modularizes the complexity, freezing its components and controlling the variability associated [30]. In addition, the understanding of project complexity helps to identify problems, develop the business case and choice processes, and improve managerial capacities [31].

Increasing competitiveness on product quality, cost, and delivery while maintaining flexibility during the whole project (including design, engineering, and production) are a few of the challenges that many organizations currently encounter in the shipbuilding industry [32]. In settings of complex projects (as those from shipbuilding sector), the ability to make proper decisions when solving problems is essential in the production efficiency of the derived operations. In this context, shipyards must face these challenges from a combination of constraints, among which the technical level of their production facilities and the practices, techniques, and tools at the disposal of their staff stand out [33].

3.2 Lean manufacturing in the shipbuilding sector

Lean manufacturing has been the most remarkable methodology for improving the operational performance in manufacturing organizations in the last two decades [34], increasing their productivity and decreasing their costs [35]. Lean manufacturing helps industrial companies to transform themselves in order to add higher value, due to the use of a considerable set of tools, methodologies, and procedures focused on boost their performance [36], waste reduction, and better communication. This combination of information acquisition and management with new design and manufacturing techniques allows companies to redirect towards new trends that respond quickly to market changes [37]. If new features must be introduced to meet these demands, companies cannot compromise their efficiency. In fact, they will try to improve it despite these challenges [38].

There are different points of view in the literature related to how lean manufacturing and Industry 4.0 interact together to influence the performance of processes involved. Some studies suggest that lean manufacturing is a mediator of their relationship [39, 40], while other suggests that Industry 4.0 is a moderator [41]. Others investigate their supportive effects without hypothesizing which of the two is the moderator [42, 43], and even other studies emphasized the interaction between them in many contexts, depending on industry and company size [44].

If shipbuilding manufacturers want to operate with lean production principles, they must establish the shipbuilding project management plan based on optimized production and overall resource balance, decomposing product tasks according to zone, stage, and type and clarifying the relationship between tasks and resources [45]. In this context, Industry 4.0 opportunities are used as a methodological and strategic tool to accelerate the engagement of shipbuilding suppliers. In these cases, lean tools mostly aim to introduce and motivate the implementation of these concepts into practice through the entire supply chain, whereby the objectives are needed to be fully understood and cross-functional teams are expected to be active in the value stream creation [46]. However, other requirements are needed, as design and assembly building methods [47].

If arbitrariness and uncertainty (affecting quality, production, operation, and logistics) are not faced, low productivity and management efficiency are the most probable result. To successfully address these challenges, shipbuilding companies must enhance their technology and management innovation, as well as actively adopt advanced production systems, for improving their efficiency [48].

3.3 Industry 4.0 in the shipbuilding sector

Industry 4.0 is a vital evolution for the survival of any industrial organization. Particularly those which target global markets, pursue a strategic distinction that supports the necessary excellence in their deliverables [49]. This implies a top-down transformation that applies to a wide range of methods, tools, and techniques involved in production management, improved processes and workplaces, and developing staff's skills [50]. Industry 4.0 modernizes the organizational processes and makes them more efficient. This involves the entire company, from operational to strategic management. In this competitive context, industrial companies need to redesign their strategies, enabling not only better resource allocation but also infrastructure investment and quality systems [51].

Industrial companies aiming to reach flexible manufacturing, with very low waste and high quality in their deliverables, are constantly evolving, in order to set them apart from their competitors. In that sense, they try to get higher levels of efficiency and productivity, associating new technologies within their processes. This use of disruptive methodologies helps them to create value, connecting and sharing information between companies and customers [52] and increasing also their applied innovation to offer complete solutions [53].

Among the Industry 4.0's main points of interest for the shipbuilding industry are artificial intelligence (pattern recognition, process automation, simulation, etc.), compatibility systems and task reassignment (occupational health and safety, decision-making, etc.), virtual and augmented reality, additive manufacturing and Internet of things, and more, specifically, the automatic generation of timelines, the creation of mathematical analysis models and evaluation of production processes, the integration of high-quality algorithms with computer-aided design (CAD) and with product life cycle management systems (PLM). In this context, the digital transformation of the shipbuilding industry optimizes the production and the operational efficiency, through the analysis and integration of storing, connecting, and organizing the information generated by different sources [17, 54].

This necessary transformation has led the shipbuilding sector to adopt the concept of Industry 4.0. The concept of "Shipyard 4.0" [55] is described as the result of the application of the Industry 4.0 to this sector. The Shipyard 4.0 involves deep changes in the shipyard production system including facilities, advanced product design, management changes, and the implementation of the digital technologies. Therefore, the Shipyard 4.0 initiative has to be the response of the shipbuilding sector to the digital transformation.

4. Case study

This research has opted for a case study since there is almost no previous research on the topic and the empirical observations are insufficient to turn it into a quantitative study. Probably, this is expected mainly due to confidentiality and competitive reasons. Companies do not tend to share the information that would be required for a more extensive analysis. In fact, when there is only limited theoretical knowledge, an inductive strategy leads to an emerging theory from a case study which can be a good starting point [56].

Building a theory from a case study is a research strategy that involves using the case to create theoretical constructions, propositions, and/or empirical evidence of midrange theory [57]. If a theoretical sampling of a single case is chosen, they must be unusually revelatory and extremely exemplar or represent unique opportunities to acquire research insights [58].

The case company is Navantia, a Spanish state-owned (and worldwide as well) reference in the design, construction, and integration of high technology military and civilian naval platforms [19]. Navantia is an ETO manufacturer that offers design, engineering, manufacturing, and project management of products (e.g., frigates, aircraft carriers, submarines, patrol vessels, logistic ships, defense systems, and wind power) and services (e.g. life support, repairs, maintenance, modernization, training, and simulation) [59]. Navantia has facilities in Spain and Australia. It also has offices in Brazil, India, Norway, Saudi Arabia, Turkey, and the USA.

The organization model applied by the company is mostly a line organization, in which department leaders are part of the project team and allocate tasks to their own staff on a periodic basis type with only a few people allocated specifically per each project. However, this type of organization is not usually associated with engineering to order contexts, where large and complex project environments have already been usually adopted [60].

Navantia is immersed in a major transformation process directed towards to increase the company sustainability in the twenty-first century market, in which technological innovation and digitalization are essential to change, encompassing all areas of the organization. The key to transformation lies not only in the implementation of innovative solutions but also in the transformation of processes and people themselves: a more agile organization, an interactive management culture, and a renewed talent management, both internally and externally, are fundamental to success [61]. Since 2015, Navantia has been striving to shape digitalization in the shipbuilding sector. This new concept of the connected Industry 4.0 emphasizes the exploitation of the potential of new technologies based on product and service innovation, client-centric approach, data value, and operational excellence.

Navantia's Shipyard 4.0 concept includes processes and products, which are integrated to operate ecologically, efficiently, and flexibly, and has an advantage over traditional systems, which are based on [62]:

- Vertical integration of the shipbuilding production processes (connectivity, additive manufacturing, Internet of things (IoT), radiofrequency, collaborative robotics, etc.), to guarantee production that is safe, fast, and adapted to the context, with a better price-performance ratio, operates online, consumes less energy, and better protects the environment

- Horizontal integration of value creation networks (cybersecurity, innovation, diversification, etc.), to attend to the needs of the interested parties in an integrated way, responding individually to them

- Reengineering of the value chain (drones, 3D/4D printing, artificial intelligence (AI), virtual and augmented reality (VAR), remote sensing networks, robotics, etc.), introducing changes that affect the lifecycle

Navantia's transformation involves an improvement in its tools and processes throughout the value chain and renewing its production centers, fully integrating them in a new digital ecosystem: the Shipyard 4.0. This change to smart factories is carried out focusing on equipment and products, applications, the company itself, and people as the main field of action [61]. At the moment, Navantia consider 13 KETs, as shown in **Figure 2**. Through these technologies, which are described below, the company is facing the digital transformation in different areas of the system and manufacturing process, regardless of whether new emerging technologies can also be introduced in the future.

4.1 Navantia key enabling technologies

4.1.1 3D printing (3DP)

3D printing is a new manufacturing process which is also known as additive manufacturing. It consists on the manufacturing of a part adding material layer by layer. This technology is getting a lot of attention nowadays, and it is expected

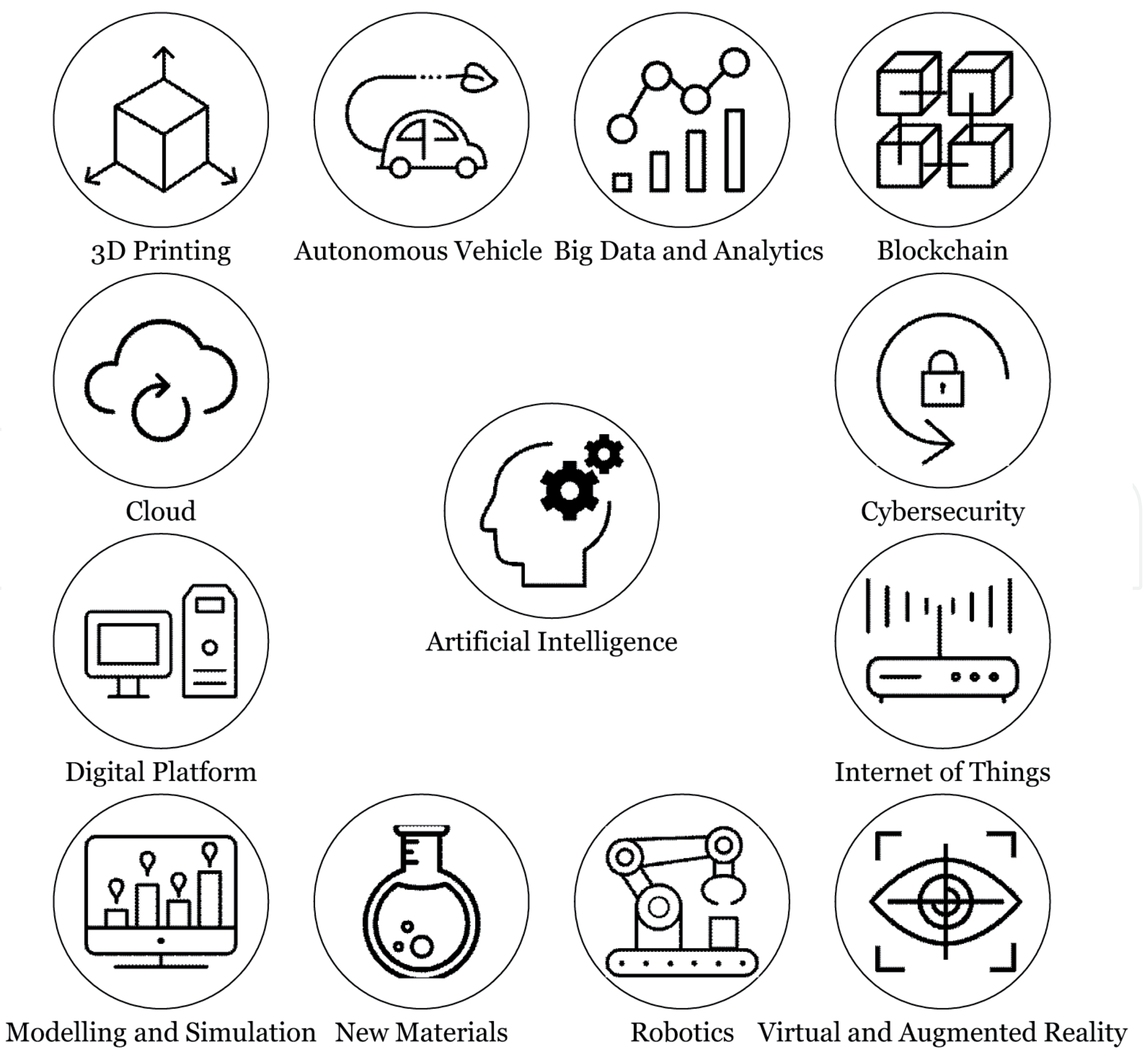

Figure 2.
Key enabling technologies by Navantia. Based on [61].

to become a major revolution in different industrial sectors. Particularly, in the shipbuilding industry, there are recent studies in which they use a polymeric-based additive manufacturing technology [63]. This technology is being used to make large, nonstructural parts, reducing the overall manufacturing costs, which also reduces the manufacturing time.

On the other hand, wire arc additive manufacturing (WAAM) technology is also under research. In this case, the polymer is replaced by a metal wire melt due to the heat produced by an electric arc [64]. This technology has the potential to replace components of the vessels which still needs to be made of metal, reducing the manufacturing costs. This assumption leads to the inevitable redesign of the ship to evaluate which parts are able to change its manufacturing technology. Therefore, it is clear that this technology still needs other changes to have the impact it is supposed to have.

4.1.2 Autonomous guided vehicles (AGV)

Autonomous guided vehicles are used for processing and transporting goods inside a factory environment [65]. They are considered smart due to their capability onboard of making decentralized decision to avoid collisions and stablish the best path planning possible to reach its destination.

Therefore, the technology of the autonomous guided vehicles makes the smart factory possible due to the flexible logistic and transport of materials within the workshop. Its application mainly affects the internal supply chain, with the aim of delivering components just in time, which has implications with the lean manufacturing system and a direct impact on the overall performance. The use of these technologies, along with simulation and artificial intelligence [13], makes the decision-making more reliable.

4.1.3 Big Data analytics (BDA)

The growing expansion of the information available due to the evolution of systems, digital products, and the development of the IoT has introduced the concept of Big Data. These are technologies which allow the capture, aggregation, and processing of the amount of the ever-growing data received by the different systems [66]. This volume of data is increasing at higher speed than the previous technologies which were capable of processing and getting valuable information of it. For that reason, the Big Data analytics is needed.

Big Data analytics is the set of techniques that make the vast amount of information generated by the other KETs manageable. At the same time, it models the data in order to get knowledge, supporting the decision-making process and even generating new solutions [66].

This amount of data, mainly gathered by sensors and the IoT, is usually storage and can be analyzed in the cloud (in real time or later) [67], which makes a very close relationship between these three technologies. Moreover, Big Data analytics has implications with other KETs too, such as additive manufacturing [68], AI [69], or simulation [70], which make the Big Data analytics one of the core driver technologies of the Industry 4.0, having also connections in the shipbuilding industry [71].

4.1.4 Blockchain (BCH)

Blockchain is a technology that can be used in any digital transaction that ends up taking place in the future Shipyard 4.0. As it is a decentralized data base in which all data are checked and confirmed by different actors before adding new information ("blocks") to the data chain, this technology improves tracking and reliability of the information due to the impossibility to change isolated information [72].

Blockchain has capabilities to promote resilience, scalability, security, autonomy, and trustworthiness [14] to every information exchange. Therefore, applications in the supply chain operation can take advantage of this technology through the smart contracts, increasing the automation and avoiding the use of intermediaries [73].

4.1.5 Cloud

The cloud is essentially a network infrastructure that supports the interconnection of Industry 4.0 through servers and cloud computing technologies [74]. It allows large data applications such as storage space, computing capacity, and resource sharing, among others. It also provides worldwide access to the information accordingly with specific access type and service provided, which can be split in different layers, named as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS), granting different kinds of access to the cloud [75].

As the industry becomes increasingly digital in manufacturing environments, the cloud computing concept has evolved into cloud manufacturing, in which users can request services during the whole lifecycle of the product. This is a change of mind-set for industrial companies as the approach differs from the previous production-oriented to the newly service-oriented concept of manufacturing, increasing flexibility during the design process [75].

Due to the remote access to the information and the application of cyber-physical systems in distributed manufacturing systems, the concept of collaborative cloud manufacturing is possible. This means that organizations with different production units connected through a collaborative network are able to synchronize themselves, multiplying the overall capacities without further investment [76]. According to this, the cloud has a fundamental role in the smart factory concept for the shipbuilding industry, in which complex projects that are undertaken in long-term can reduce the overall production time to meet on-demand expectations.

4.1.6 Cybersecurity (CS)

The huge amount of information that is sent from different devices to the cloud and backwards creates new opportunities and vulnerabilities to the industrial companies. This scenario compromises confidential information due to the banishment of the physical boundaries [77]. For this reason, the evolution of security towards the virtual world is needed, giving birth to the cybersecurity, which aims to increase the security levels in IoT environments.

The cybersecurity, by definition, is a process consisting three objectives: to protect, detect, and respond to cyber-attacks [78]. Particularly, the two main objectives are the ones that rely on data protection and are given more attention since Internet of things networks have to be built in a safe environment that allows a safe interoperability between the facilities. But not only the information is at risk. As the manufacturing units are connected to the network, they can also be shut down, change its normal behavior, or even modify the product design. All of these factors lead to an enormous economic loss and should therefore be avoided [79].

In summary, this technology needs to be addressed and takes an important role in the context of any enterprise, which aims to carry a deep digital transformation out. To achieve a successful smart factory, the concept of "security by design" is mandatory in which both, data and cyber-physical systems, are adequately protected [77].

4.1.7 Digital platform (DP)

The digital platform is the answer that Navantia has given to the horizontal and vertical integration. Horizontal and vertical integrations involve every stakeholder that takes part in the production process, including marketing, supply chain operations, or engineering, among others. Referring to each integration to the intercompany or intracompany, respectively, the global output is a real-time data sharing among every part implied in the process [75].

To make this integration possible, a digital platform, aided by cloud computing, is the perfect answer to gather all the agents, both from the supplier or the client, as it can be accessed remotely from different geographical areas to collaborate in the process, updating the information needed in real time and resulting in a fully updated system, which can give further information according to all the data received. Therefore, Big Data analytics and cybersecurity are also connected with the digital platform.

4.1.8 Internet of things (IoT)

The IoT refers to the connectivity of every device within a network that is able to generate data from sensors or embedded electronic devices, which are sent afterwards to the cloud through a transmission system [80]. As every "thing" is generating data, the connection between IoT and Big Data analytics is clear. This technology also includes the concept of cyber-physical systems, being the gateway to fuse the real world with the virtual world, bringing physical objects into the network.

In the industrial sector, the application of the IoT is known as the industrial Internet of things, having particular implications and principles that must be fulfilled [81]. These principles include, among others, interoperability, wireless communication, decentralization, real-time feedback, or system-wide security to avoid outsider's intromission, which can put all the data on risk. In this way, cybersecurity technology acquires an important role in protecting the industrial environment.

A study on the implications this technology can handle in complex engineering projects, namely, the ones carried out in the shipbuilding industry, is also under investigation [82]. This research concludes that it is possible to create a "digital construction site" for shipbuilding, in which the IoT plays a strategic role as it is being used in specific applications.

4.1.9 Modeling and simulation (M&S)

By definition, a simulation is the imitation of the operation of a real-world process or system over time. Simulation involves the generation of an artificial history of the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented [83]. Therefore, almost any real world can be modeled into the virtual one, in order to study and predict its behavior after developing and applying specific events. In this way, many kinds of simulations appear, each regarding one different area of study [84].

Although Industry 4.0 represents a new paradigm, this can be also accomplished by simulation. Due to high levels of digitization and the increased integration of the whole product lifecycle, the traditional stand-alone simulations are not able to fulfill those new requirements.

In this challenging scenario, the concept of digital twin appears, which consists of the digital representation of an asset that can alter its properties and behavior by information and data [85]. This is the result of adopting a system design approach, which allows to train on a virtual machine and to identify potential issues with the real machine if it is combined with a model predictive analysis, deep learning, and AI. Besides, this enables to optimize its own performance, because it will be able to predict faults and coordinate with other machines, thanks to machine-to-machine interaction.

These technologies are being applied also in the shipbuilding industry, in which CAD/CAM/CAE solutions are already in use; meanwhile, discrete event simulation as the previous step of the digital twin is under development. Moreover, the application of finite element methods for new materials is also a technology with a huge potential to advance them.

4.1.10 New materials (NM)

The development of new materials, such as those based on composite carbon fiber- and fiber-reinforced plastic, polymers, or new metal alloys [86], facilitates to redesign the shipbuilding sector's product to add or replace several components. The use of these materials can offer a weight reduction, leading to a decrease in fuel consumption, which would end up making the vessels eco-friendlier.

The advantages of introducing these materials can also strengthen the corrosion resistance [87]. This can be achieved thanks to the use of new materials which are resistant to the corrosive action of salt water, leading to an increase in the added value of the ships provided.

4.1.11 Robotics (robot)

The robotics is one of those technologies from the third industrial revolution that holds a paradigm change with this new industrial revolution. In that sense, the manufacturing paradigm, from mass production towards customized production, makes the robots need to be more flexible and autonomous [75]. On the other hand, the use of advanced sensors makes the integration between robot and operator possible, resulting in collaborative robots or cobots [88].

Despite this technology is mostly used to undertake very easy repetitive actions, like in a production line, advanced shipyards have achieved to introduce this technology within its manufacturing system, increasing drastically its performance [89]. Furthermore, new advances have been managed to develop robots for specific shipbuilding tasks, such as pipe inspection or hull cleaning. In the case study, Navantia has also researched regarding robotic welding [90].

4.1.12 Virtual and augmented reality (VAR)

The VAR could be englobed within modeling and simulation technologies [84]. However, as this technology involves partial or complete human immersion, as well as pursues a different aim, the VAR has been treated separately.

On the one hand, the virtual reality implies a full immersion of the human being within a virtual world using a special device connected with a simulation. In this virtual world, the user can interact with virtual elements in order to train and improve the operator knowledge significantly. It has also applications in product testing and validation of complex products [91].

On the other hand, the augmented reality converges the real world with the virtual one through a device, adding data from the virtual system (or digital twin), exactly where needed. This technology is useful not only in the manufacturing processes but also in maintenance tasks. Using augmented reality also offers applications in assuring quality control, location of products and tools, warehouse management, and support for the visualization of hidden areas [92], among others.

In the shipbuilding industry, both technologies are already being used in small applications for training and part positioning.

4.1.13 Artificial intelligence (AI)

The AI is one of the Industry 4.0 driver technologies. According to the European Commission, AI refers to "systems that display intelligent behavior by analyzing their environment and taking actions (with some degree of autonomy) to achieve specific goals" [16]. Its application in the industrial sector has resulted in the "intelligent manufacturing" concept [93], which, along with the other recent emerging Industry 4.0 KETs, will allow more flexible and efficient operations in the smart factory [15]. In order to achieve a good implantation of this technology, the industrial AI framework is also proposed with a clear structure, methodology, and ecosystem [15].

In the shipbuilding industry, there are already some applications in terms of design vessels for optimizing the overall performance [94]. The applications of AI are mainly related with the development of other technologies, acting as an enabler to impulse the potential of each one of the other KETs [95]. This is shown in the interaction between AI and the particular effect it deploys.

4.2 Social network analysis (SNA)

Due to the existing correlation between the KETs selected in this case study, it is possible to develop a social network in order to confirm the links among them. A social network is defined as a finite set of actors (such as people, organizations, or technologies) and the relationship among them [96]. The social network perspective focuses on these relations as an important addition to the standard social research, which is mainly concerned in the attributes of the social units. The social network analysis (SNA) is similar to the mind map technique, which allows to represent the ideas and their relationships. This method has already been used to the study the Industry 4.0 enablers [97]. The SNA is an innovative technique and research tool that has already been successfully used to find the relationship between different technologies and resources relative to the Industry 4.0 [80].

The MoSCoW method is used [98] to establish the network of Navantia's KETs. This method stands for "Must, Should, Could and Won't Have" criteria, and it is mainly used to stablish a priority list. In this case, a variant of the method is considered to weight the different interaction possibilities:

- Must have. Numeric value 3: The technology studied needs the crossed technology one mandatory.

- Should have. Numeric value 2: The technology studied can have major connection with the crossed technology.

	3DP	AuV	BD	BCH	Cloud	CybS	AI	DP	IoT	M&S	NM	Robot	VAR
3DP	0	0	1	1	1	1	2	1	1	3	2	1	0
AuV	1	0	1	1	2	2	2	0	2	1	0	0	1
BD	0	0	0	2	3	2	2	1	3	0	0	0	0
BCH	0	0	0	0	1	2	2	0	0	0	0	0	0
Cloud	0	0	1	1	0	2	2	1	1	0	0	0	0
CybS	0	0	1	2	1	0	2	1	0	0	0	0	0
AI	0	0	2	1	2	2	0	1	2	1	0	0	0
DP	0	0	1	1	3	1	1	0	1	0	0	0	0
IoT	0	0	2	1	3	1	2	1	0	1	0	0	0
M&S	0	0	1	0	1	0	2	1	1	0	0	0	1
NM	2	0	2	0	0	0	2	0	0	2	0	0	0
Robot	1	1	2	1	1	1	2	0	2	2	1	0	1
VAR	0	0	2	1	2	1	2	1	1	3	0	0	0

Table 1.
Relationship among Navantia's key enabling technologies.

- Could have. Numeric value 1: The technology studied can have minor connection with the crossed technology.

- Won't have. Numeric value 0: The technology studied does not need the crossed technology.

In the first place, a nonsymmetric matrix is created, in which the nonlinear dependencies between the KETs are shown. Each row shows the dependency of a KET with the others. For example, VAR is dependent of M&S, but it is not the same in the other way around. These binary and paired comparison assessments were completed by the expert committee of Navantia, as summarized in **Table 1**. Once the data is ready, it is introduced in the software UNICET (version 6.675) [99], which will return the analyzed data (from the social network) and graphic representation.

5. Results

Once the data is analyzed, the results give information regarding the relationship between KETs, as betweenness, centrality, closeness, or density. However, the measures of centrality and betweenness are the ones to be taken into account. Centrality is the grade of each actor which is linked with the others. In a nonsymmetric matrix, the difference between ins and outs means the necessity of other technologies have of the chosen one (ins) and the necessity of the chosen technologies have of the others (outs). In addition, betweenness is the possibility that an actor has to intermediate the communications between pairs of actors. These are also known as bridge actors. The grade of centrality and betweenness is summarized in **Table 2**, where the main results are shown in bold.

These results show that both the AI and the cloud are the most demanded technologies among the other KETs (more than 0.55 relatively), while the individual

	Outdeg	Indeg	nOutdeg	nIndeg	Betweenness	nBetweenness
3DP	14.000	4.000	0.389	0.111	5.667	4.293
AuV	13.000	1.000	0.361	0.028	0.000	0.000
BDA	13.000	16.000	0.361	0.444	2.250	1.705
BCH	5.000	12.000	0.139	0.333	0.000	0.000
Cloud	8.000	**20.000**	0.222	**0.556**	2.000	1.515
CybS	7.000	15.000	0.194	0.417	0.917	0.694
AI	11.000	**23.000**	0.306	**0.639**	**10.417**	**7.891**
DP	8.000	8.000	0.222	0.222	0.583	0.442
IoT	11.000	14.000	0.306	0.389	3.583	2.715
M&S	7.000	13.000	0.194	0.361	**9.500**	**7.197**
NM	8.000	3.000	0.222	0.083	0.000	0.000
Robot	15.000	1.000	0.417	0.028	2.500	1.894
VAR	13.000	3.000	0.361	0.083	0.583	0.442

Table 2.
Centrality and betweenness grade of Navantia's KETs.

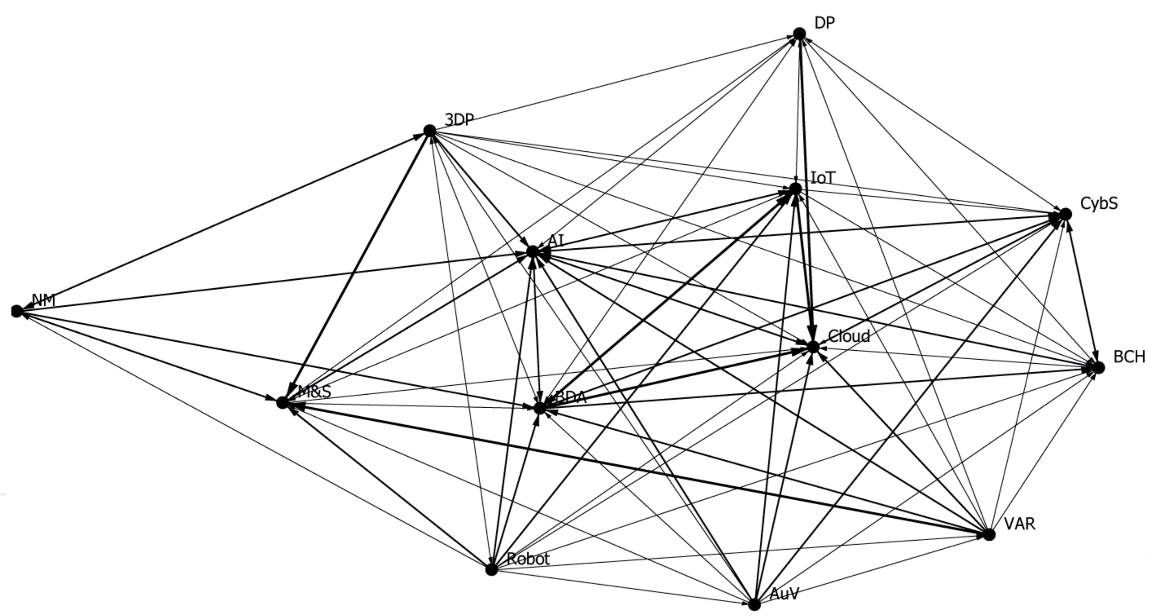

Figure 3.
Social network among Navantia's key enabling technologies.

dependency of each KET on the others is not too high, being robotics the most dependent. In terms of betweenness, the AI stands out again, followed closely by modeling and simulation. These are the two technologies with more capacity to stablish interactions between other technologies, which is an important added value to consider. The social network result is drawn in **Figure 3**, in which the connections between the technologies are represented.

The network shows the four main technologies on which the rest of the technologies revolve: artificial intelligence (AI), cloud, Big Data analytics (BDA), and Internet of things (IoT). This is consistent with the principles of the digital transformation and with the implications that the use of AI has to achieve a further development of the other KETs either due to direct integration or as an enabler linker for other technologies. These results also show the importance each of the technology has in the shipbuilding industry. This could be used to stablish a criterion, in order to support one technology over another.

6. Conclusions

In this book chapter, a state of the art of the shipbuilding industry is carried out. This includes a literature review in shipbuilding complex projects, lean manufacturing implications in shipbuilding, and the introduction of the fourth industrial revolution into this industrial sector, and there is a need to overcome the difficult situation that it is currently facing. To go further in this research, a study case is presented. The Spanish state-owned shipyard Navantia is chosen to study how a shipyard is challenging the digital transformation and introducing KETs in its production system. Afterwards, a revision of the advances and the integration of these technologies in the shipbuilding industry are presented.

Moreover, due to the relationship that exists among the KETs, a SNA is performed. This analysis has confirmed the main technologies that the Industry 4.0 has to prioritize during its implementation. From the nine original technologies, Big Data analytics, Internet of things, and cloud are highlighted. On top of those, artificial intelligence appears to join the cloud as the technology that will have the biggest impact in the Industry 4.0, due to its potential to increase the benefits of the other key enabling technologies.

Acknowledgements

This work is part of the advances of an industrial doctoral thesis within the framework of the “Agreement between the University of Cadiz and Navantia S.A., S.M.E. for collaboration in the promotion of the training of research staff for the completion of doctoral theses in companies,” acknowledging both entities to enable, finance, and facilitate its development.

Conflict of interest

The authors declare no conflict of interest.

Author details

Alejandro Sánchez-Sotano[1]*, Alberto Cerezo-Narváez[1], Francisco Abad-Fraga[2], Andrés Pastor-Fernández[1] and Jorge Salguero-Gómez[1]

1 Department of Mechanical Engineering and Industrial Design, University of Cadiz, Spain

2 Navantia S.A. S.M.E. Bahía de Cádiz Shipyard, Spain

*Address all correspondence to: alejandrojavier.sanchez@uca.es

References

[1] Herterich MM, Uebernickel F, Brenner W. The impact of cyber-physical systems on industrial services in manufacturing. Procedia CIRP. 2015;**30**:323-328

[2] Favi C, Campi F, Germani M, Manieri S. Using design information to create a data framework and tool for life cycle analysis of complex maritime vessels. Journal of Cleaner Production. 2018;**192**:887-905

[3] Baginski L, Pitassi C, Barbosa JGP. Technological capability in the Brazilian naval industry: A metric for offshore support vessels. RAI—Revista de Administração e Inovação. 2017;**14**(2):109-118

[4] Kolić D, Fafandjel N, Zamarin A. Lean manufacturing methodology for shipyards. Shipbuilding. 2012;**63**(1):18-29

[5] Eriksson T, Vahlo J, Rissanen T, Koponen A. Management myopia and project management. Insiders vs outsiders. In: ISPIM Innovation Symposium. Fukuoka, Japan: The International Society for Professional Innovation Management (ISPIM); 2018. pp. 1-9

[6] Vrijhoef R, Koskela L. The four roles of supply chain management in construction. European Journal of Purchasing & Supply Management. 2000;**6**(3-4):169-178

[7] Ng ST, Rose TM, Mak M, Chen SE. Problematic issues associated with project partnering—The contractor perspective. International Journal of Project Management. 2002;**20**(6):437-449

[8] Xue X, Wang Y, Shen Q, Yu X. Coordination mechanisms for construction supply chain management in the Internet environment. International Journal of Project Management. 2007;**25**(2):150-157

[9] Hansen MJ, Vaagen H. Collective intelligence in project groups: Reflections from the field. Procedia Computer Science. 2016;**100**(1877):840-847

[10] Aloini D, Dulmin R, Mininno V, Ponticelli S. Key antecedents and practices for Supply Chain Management adoption in project contexts. International Journal of Project Management. 2015;**33**(6):1301-1316

[11] Iordache O. Implementing Polytope Projects for Smart Systems, Studies in Systems, Decision and Control. 1st ed. Vol. 92. Cham: Springer; 2017. p. 197

[12] Lorenz M, Rüßmann M, Strack R, Lueth K, Bolle M. Industry 4.0. The Future of Productivity and Growth in Manufacturing Industries. Boston, USA; 2015

[13] Yao F, Keller A, Ahmad M, Ahmad B, Harrison R, Colombo AW. Optimizing the scheduling of autonomous guided vehicle in a manufacturing process. In: 16th International Conference on Industrial Informatics (INDIN). Porto, Portugal: IEEE; 2018. pp. 264-269

[14] Viriyasitavat W, Da Xu L, Bi Z, Sapsomboon A. Blockchain-based business process management (BPM) framework for service composition in industry 4.0. Journal of Intelligent Manufacturing. 2018:1-12. (In press)

[15] Lee J, Davari H, Singh J, Pandhare V. Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. Manufacturing Letters. 2018;**18**:20-23

[16] European Commission. Artificial Intelligence for Europe. Brussels, Belgium: Digital Single Market; 2018

[17] Rouco Couzo M, Quiroga Pazos M, Cebral Fernández M, Morgade Abeal R, García del Valle A, Morgade Abeal D. El Astillero 4.0: Modelado y Simulación del Astillero de Navantia-Ferrol. In: 56º Congreso de Ingeniería Naval e Industria Marítima. Madrid: Asociación de Ingenieros Navales y Oceánicos de España (AINE); 2017. pp. 1-15

[18] Cagalj A, Veza I, Markovina R. Interactive networked company in shipbuilding industry. Strojarstvo. 2009;**51**(1):15-26

[19] Ramsay S. A case for a new warship-building strategy. Maritime Affairs: Journal of the National Maritime Foundation of India. 2012;**8**(1):150-156

[20] Vidal LA, Marle F, Bocquet JC. Measuring project complexity using the Analytic Hierarchy Process. International Journal of Project Management. 2011;**29**(6):718-727

[21] Dao BP. Exploring and Measuring Project Complexity. College Station, TX, US: Texas A&M University; 2016

[22] Wiskerke J, Veeke H, Pruijn J, Groen C, Hopman H. Knowledge framework on the execution of complex projects. The development of a functional framework using a systems approach. In: International Conference on Information and Knowledge Engineering (IKE). Athens, Greece: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp); 2016. pp. 73-78

[23] Hall DJ, Saias MA. Strategy follows structure! Strategic Management Journal. 1980;**1**(2):149-163

[24] Baccarini D. The concept of project complexity: A review. International Journal of Project Management. 1996;**14**(4):201-204

[25] Duncan J. The structure of cognition: Attentional episodes in mind and brain. Neuron. 2013;**80**(1):35-50

[26] Bakhshi J, Ireland V, Gorod A. Clarifying the project complexity construct: Past, present and future. International Journal of Project Management. 2016;**34**(7):1199-1213

[27] Hayes S, Bennett D. Managing projects with high complexity. In: Cooke-Davies T, Crawford L, Patton JR, Stevens C, Williams TM, editors. Aspects of Complexity: Managing Projects in a Complex World. Newtown Square: Project Management Institute; 2011. pp. 17-27

[28] Cooke-Davies T. Complexity in project management and the management of complex crojects. In: Cooke-Davies T, Crawford L, Patton JR, Stevens C, Williams TM, editors. Aspects of Complexity: Managing Projects in a Complex World. Newtown Square: Project Management Institute; 2011. pp. 1-13

[29] Sistemas y Tecnología de IBM. Watson-Un sistema Diseñado para Respuestas. El futuro del diseño de sistemas optimizados de cargas de trabajo. Madrid; 2011

[30] Loch C, Payne FC. Strategic management: Developing policies and strategies. In: Cooke-Davies T, Crawford L, Patton JR, Stevens C, Williams TM, editors. Aspects of Complexity: Managing Projects in a Complex World. Newtown Square: Project Management Institute; 2011. pp. 41-55

[31] Geraldi J, Maylor H, Williams T. Now, let's make it really complex (complicated). International Journal of Operations & Production Management. 2011;**31**(9):966-990

[32] Kjersem K, Jünge GH. Improving planning process for

ETO-projects: A case study. In: Nääs I, Vendrametto O, Mendes Reis J, Gonçalves RF, Silva MT, von Cieminski G, et al., editors. Advances in Production Management Systems, APMS 2016, IFIP Advances in Information and Communication Technology. Vol. 488. Cham: Springer International Publishing; 2016. pp. 927-934

[33] Hellgren S. The Bayesian Model for Cruise Shipbuilding: A Process for Production efficiency and Organization. Vol. 240. Helsinki, Finland: Aalto University; 2016

[34] Found P, Bicheno J. The Routledge companion to lean management. In: Netland TH, Powell DJ, editors. The Routledge Companion to Lean Management. New York, NY: Routledge; 2016. pp. 23-33

[35] Sanders A, Elangeswaran C, Wulfsberg J. Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing. Journal of Industrial Engineering and Management. 2016;**9**(3):811

[36] NG T, Ghobakhloo M. What derives lean manufacturing effectiveness: An interpretive structural mode. International Journal of Advances in Applied Sciences. 2017;**4**(8):104-111

[37] Ghobakhloo M, Azar A. Business excellence via advanced manufacturing technology and lean-agile manufacturing. Journal of Manufacturing Technology Management. 2018;**29**(1):2-24

[38] Sartal A, Vázquez XH. Implementing information technologies and operational excellence: Planning, emergence and randomness in the survival of adaptive manufacturing systems. Journal of Manufacturing Systems. 2017;**45**:1-16

[39] Tortorella G, Miorando R, Caiado R, Nascimento D, Portioli Staudacher A. The mediating effect of employees' involvement on the relationship between Industry 4.0 and operational performance improvement. Total Quality Management and Business Excellence. 2018:1-15. (In press)

[40] Kamble S, Gunasekaran A, Dhone NC. Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies. International Journal of Production Research. 2019:1-19. (In press)

[41] Tortorella GL, Giglio R, van Dun DH. Industry 4.0 as a moderator on the relationship between lean and operational performance. In: 25th International Annual EurOMA Conference: To Serve, to Produce and to Servitize in the Era of Networks, Big Data and Analytics. Budapest, Hungary: University of Twente; 2018. pp. 1-10

[42] Tortorella GL, Fettermann D. Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies. International Journal of Production Research. 2018;**56**(8):2975-2987

[43] Rossini M, Costa F, Tortorella GL, Portioli-Staudacher A. The interrelation between Industry 4.0 and lean production: An empirical study on European manufacturers. International Journal of Advanced Manufacturing Technology. 2019;**102**(9-12):3963-3976

[44] Pinho C, Mendes L. IT in lean-based manufacturing industries: Systematic literature review and research issues. International Journal of Production Research. 2017;**55**(24):7524-7540

[45] Jiang G, Feng D, Zhu W. Toward efficient merchant shipbuilding based on the lean production methodology. Journal of Ship Production and Design. 2016;**32**(4):245-257

[46] Beifert A, Gerlitz L, Prause G. Industry 4.0—For sustainable development of lean manufacturing companies in the shipbuilding sector. In: Kabashkin I, Yatskiv I, Prentkovskis O, editors, Lecture Notes in Networks and Systems. Vol. 36. Cham: Springer International Publishing; 2018. pp. 563-573

[47] Jiang G, Feng D, Zhu W. Lean shipbuilding for project-based manufacturing. In: Industrial and Systems Engineering Research Conference. Nashville, USA: Institute of Industrial and Systems Engineers (IISE); 2015. pp. 1730-1739

[48] Buer S-V, Strandhagen JO, Chan FTS. The link between Industry 4.0 and lean manufacturing: Mapping current research and establishing a research agenda. International Journal of Production Research. 2018;**56**(8):2924-2940

[49] Agrawal A, Schaefer S, Funke T. Incorporating Industry 4.0 in corporate strategy. In: Analyzing the Impacts of Industry 40 in Modern Business Environments. Hershey, USA: IGI Global; 2018. pp. 161-176

[50] Ragulina YV, Shkodinsky SV, Mishchenko VV, Romanova YA. Scenarios of development of Industry 4.0 in the conditions of knowledge economy's formation and their consequences for modern economic systems. In: Studies in Systems, Decision and Control. Cham, Switzerland: Springer; 2019. pp. 227-234

[51] Paranitharan KP, Babu R, Pandi PA, Jeyathilagar D. An empirical validation of integrated manufacturing business excellence model. International Journal of Advanced Manufacturing Technology. 2017;**92**(5-8):2569-2591

[52] Gunasekaran A, Papadopoulos T, Dubey R, Wamba SF, Childe SJ, Hazen B, et al. Big data and predictive analytics for supply chain and organizational performance. Journal of Business Research. 2017;**70**:308-317

[53] Costa N, Patrício L, Morelli N, Magee CL. Bringing service design to manufacturing companies: Integrating PSS and service design approaches. Design Studies. 2018;**55**:112-145

[54] Hmeshah K, Erbach N, Bronsart R. Impact of production requirements on high-quality ship product data models. Journal of Engineering for the Maritime Environment. 2016;**230**(3):508-522

[55] Recamán RÁ. Navantia's Shipyard 4.0 model overview. Cienc y Tecnol buques. 2018;**11**(22):77

[56] Siggelkow N. Persuasion with case studies. The Academy of Management Journal. 2007;**50**(1):20-24

[57] Eisenhardt KM. Building theories from case study research. The Academy of Management Review. 1989;**14**(4):532

[58] Yin RK. Case Study Research: Design and Methods. 3rd ed. SAGE: Thousand Oaks; 2003. p. 183

[59] Navantia. Products and Services [Internet]. 2018. Available from: https://www.navantia.es/en/products-and-services/

[60] Gosling J, Naim MM. Engineer-to-order supply chain management: A literature review and research agenda. International Journal of Production Economics. 2009;**122**(2):741-754

[61] Navantia. Shipyard 4.0. 2018 [Internet]. 2018. Available from: https://www.navantia.es/en/navantia-4-0/shipyard-4-0/

[62] Fraga LP. Enabling Technologies and Cyber-Physical Systems for Mission-Critical Scenarios. Universidade da Coruña; 2017

[63] Moreno Nieto D, Casal López V, Molina SI. Large-format polymeric pellet-based additive manufacturing for the naval industry. Additive Manufacturing. 2018;**23**:79-85

[64] Knezović N, Topić A. Wire and arc additive manufacturing (WAAM)—A new advance in manufacturing. In: Karabegović I, editor. New Technologies, Development and Application, Lecture Notes in Networks and Systems. Cham, Switzerland: Springer International Publishing; 2019. pp. 65-71

[65] Mehami J, Nawi M, Zhong RY. Smart automated guided vehicles for manufacturing in the context of Industry 4.0. Procedia Manufacturing. 2018;**26**:1077-1086

[66] Bi Z, Cochran D. Big data analytics with applications. Journal of Management Analytics. 2014;**1**(4):249-265

[67] Jun C, Lee JY, Kim BH. Cloud-based big data analytics platform using algorithm templates for the manufacturing industry. International Journal of Computer Integrated Manufacturing. 2019;**32**(8):723-738

[68] Majeed A, Lv J, Peng T. A framework for big data driven process analysis and optimization for additive manufacturing. Rapid Prototyping Journal. 2019;**25**(2):308-321

[69] Tao F, Qi Q, Wang L, Nee AYC. Digital twins and cyber–physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison. Engineering. 2019;**5**(4):653-661

[70] Taylor SJE. Distributed simulation: State-of-the-art and potential for operational research. European Journal of Operational Research. 2019;**273**(1):1-19

[71] Bao J, Zheng X, Zhang J, Ji X, Zhang J. Data-driven process planning for shipbuilding. Artificial Intelligence for Engineering Design Analysis and Manufacturing AIEDAM. 2018;**32**(1):122-130

[72] Lin C, He D, Huang X, Choo K-KR, Vasilakos AV. BSeIn: A blockchain-based secure mutual authentication with fine-grained access control system for industry 4.0. Journal of Network and Computer Applications. 2018;**116**(March):42-52

[73] Chang SE, Chen Y-C, Lu M-F. Supply chain re-engineering using blockchain technology: A case of smart contract based tracking process. Technological Forecasting and Social Change. 2019;**144**:1-11

[74] Wang S, Wan J, Li D, Zhang C. Implementing smart factory of Industrie 4.0: An outlook. International Journal of Sensor Networks. 2016;**12**(1):3159805

[75] Alcácer V, Cruz-Machado V. Scanning the Industry 4.0: A literature review on technologies for manufacturing systems. Engineering Science and Technology, an International Journal. 2019;**22**(3):899-919

[76] Rauch E, Seidenstricker S, Dallasega P, Hämmerl R. Collaborative cloud manufacturing: Design of business model innovations enabled by cyberphysical systems in distributed manufacturing systems. Journal of Engineering. 2016;**2016**:1-12

[77] He H, Maple C, Watson T, Tiwari A, Mehnen J, Jin Y, et al. The security challenges in the IoT enabled cyber-physical systems and opportunities for evolutionary computing & other computational intelligence. In: 2016 IEEE Congress on Evolutionary Computation (CEC). Vancouver, BC, Canada: IEEE; 2016. pp. 1015-1021

[78] Piedrahita AFM, Gaur V, Giraldo J, Cardenas AA, Rueda SJ. Virtual incident response functions in control systems. Computer Networks. 2018;**135**:147-159

[79] Tsuchiya A, Fraile F, Koshijima I, Ortiz A, Poler R. Software defined networking firewall for industry 4.0 manufacturing systems. Journal of Industrial Engineering and Management. 2018;**11**(2):318

[80] Da Costa MB, Dos Santos LMAL, Schaefer JL, Baierle IC, Nara EOB. Industry 4.0 technologies basic network identification. Scientometrics. 2019;**121**(2):977-994

[81] ur Rehman MH, Yaqoob I, Salah K, Imran M, Jayaraman PP, Perera C. The role of big data analytics in industrial Internet of Things. Future Generation Computer Systems. 2019;**99**:247-259

[82] Lopes Miranda Junior H, Albuquerque Bezerra NR, Soares Bezerra MJ, Rodrigues Farias Filho J. The internet of things sensors technologies and their applications for complex engineering projects: A digital construction site framework. Brazilian Journal of Operations & Production Management. 2017;**14**(4):567-576

[83] Banks J, Carson JS II, Nelson BL, Nicol DM. Discrete-Event System Simulation. Prentice Hall: Upper Saddle River, USA; 2010. p. 564

[84] Mourtzis D, Doukas M, Bernidaki D. Simulation in manufacturing: Review and challenges. Procedia CIRP. 2014;**25**(C):213-229

[85] Rodič B. Industry 4.0 and the new simulation modelling paradigm. Organizacija. 2017;**50**(3):193-207

[86] Cavendish Group. The challenges of using new materials in shipbuilding [Internet]. Marine & Offshore Technology. 2017. Available from: http://www.marineoffshoretechnology.net/features-news/challenges-using-new-materials-shipbuilding

[87] Ferraris S, Volpone LM. Aluminum alloys in third millennium shipbuilding: Materials, technologies, persepectives. In: 5th International Forum of Aluminum Ships. Tokyo, Japan. 2005. pp. 1-10

[88] Koch PJ, van Amstel MK, Dębska P, Thormann MA, Tetzlaff AJ, Bøgh S, et al. A skill-based robot co-worker for industrial maintenance tasks. Procedia Manufacturing. 2017;**11**(June):83-90

[89] Samsung. The Future of the Global Shipbuilding and Offshore Industries [Internet]. Samsung Heavy Industries. 2017. Available from: http://www.samsungshi.com/eng/deFaulT.aspx

[90] Galindo PL, Morgado-Estévez A, Aparicio JL, Bárcena G, Soto-Núñez JA, Chavera P, et al. Development of a customized interface for a robotic welding application at navantia shipbuilding company. In: Third Iberian Robotics Conference (ROBOT 2017), Advances in Intelligent Systems and Computing. Vol. 694. Cham, Switzerland: Springer International Publishing; 2018. pp. 43-52

[91] Roldán JJ, Crespo E, Martín-Barrio A, Peña-Tapia E, Barrientos A. A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining. Robotics and Computer-Integrated Manufacturing. 2019;**59**(May):305-316

[92] Fraga-Lamas P, Fernandez-Carames TM, Blanco-Novoa O, Vilar-Montesinos MA. A review on industrial augmented reality systems for the Industry 4.0 shipyard. IEEE Access. 2018;**6**:13358-13375

[93] Yao X, Zhou J, Lin Y, Li Y, Yu H, Liu Y. Smart manufacturing based on cyber-physical systems and beyond. Journal of Intelligent Manufacturing. 2019;**30**(8):2805-2817

[94] Abramowski T. Application of artificial intelligence methods to preliminary design of ships and ship performance optimization. Naval Engineers Journal. 2013;**125**(3):101-112

[95] Dopico M, Gomez A, De la Fuente D, García N, Rosillo R, Puche J. A vision of industry 4.0 from an artificial intelligence point of view. In: International Conference on Artificial Intelligence (ICAI 2016—WORLDCOMP 2016). Las Vegas, USA: CSREA Press; 2016. pp. 407-413

[96] Wasserman S, Faust K. Social Network Analysis: Methods and Applications. New York, USA: Cambridge University Press; 1994. p. 857

[97] Havle CA, Ucler C. Enablers for Industry 4.0. In: 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). Ankara, Turkey: IEEE; 2018. pp. 1-6

[98] Agile Business Consortium. Chapter 10: MoSCoW Prioritisation [Internet]. DSDM Agile Project Framework Handbook. 2014. Available from: https://www.agilebusiness.org/page/ProjectFramework_10_MoSCoWPrioritisation

[99] Borgatti SP, Everett MG, Freeman LC. UCINET 6 for Windows: Software for Social Network Analysis. Cambridge, USA: Harvard Analytic Technologies; 2002

Chapter

Energy Infrastructure of the Factory as a Virtual Power Plant: Smart Energy Management

Eva M. Urbano and Víctor Martínez Viol

Abstract

Smart energy factories are crucial for the development of upcoming energy markets in which emissions, energy use and network congestions are to be decreased. The virtual power plant (VPP) can be implemented in an industrial site with the aim of minimizing costs, emissions and total energy usage. A VPP considers the future situation forecasting and the situation of all energy assets, including renewable energy generation units and energy storage systems, to optimize the total cost of the plant, considering the possibility to trade with the energy market. For a VPP to be constructed, a proper communication system is essential. The energy management system (EMS) enables the monitoring, management and control of the different energy devices and permits the transference of the decisions made by the VPP to the different energy assets. VPP concept is explained together with the methods used for forecasting the future situation and the energy flow inside the facility. To reach its benefits, the optimization of the VPP is assessed. After that, the communication technologies that enable the VPP implementation are also introduced, and the advantages/disadvantages regarding their deployment are stated. With the tools introduced, the VPP can face the challenges of energy markets efficiently.

Keywords: virtual power plant, smart grid, energy hub, ANFIS, communication technologies, energy management system

1. Introduction

Industry 4.0 is normally understood as smart factories where automation, digitalization, Internet of Things (IoT), cognitive computing and others are used. However, this does not stand without the use of energy. There is a settled relationship between energy consumption, energy prices and economic growth in different countries. For industries, the access to reliable and affordable energy is crucial to create greater economic and social prosperity. In the industry that is emerging nowadays, the physical processes are studied, modeled and monitored, and physical systems communicate and cooperate in a real-time scenario in order to optimize the behavior of the plant. The same can be done with energy. To reach the best efficiency of a manufacturing plant, the energy consumption processes have to be studied, modeled and monitored; the communication of the energy flows between equipment has to be known, and future situation prediction and real-time decisions have to be taken regarding energy purchasing, energy trading, generation and consumption.

There are several reasons for why the development of energy-smart factories is interesting. Policy is making an effort in order to achieve a reduction of greenhouse gas emissions, an increase in the share of renewable energy and an improvement in the energy efficiency. As an example, in Europe, the energy usage in the industrial sector accounts for more than 25% of total energy consumption, process heating having the most significant use with 66% followed by electricity with 26%. If energy efficiency measures are developed and incorporated in the industrial sector, the potential savings can be of more than 20% as shown in [1]. Regarding the increase in the share of renewable energies, it will be possible with the integration of smart energy systems. Some renewable energy sources such as solar and wind power generation are characterized by an intermittent nature. One of the fundamental properties of the electric grid is that the supply (generation) and the demand (consumption) must always be balanced. With the increase in the share of renewable power sources, the energy may not be generated in the best suited moment and with the exact amount of power dealing to grid instability and not assuring a security of supply. By defining, integrating and controlling the energy flow in order to optimize the consumption of energy hubs (EH) and, from there, exploit it in virtual power plants (VPP), the industrial sector the electricity usage can be optimized, allowing a greater efficiency and flexibility, improving the capacity factor of the installed renewable energy sources. Up to date, the EH concept has been presented by several studies in the industrial field, and its expansion into a VPP is a new research field in which the focus is the possibility of energy trading with the grid, as can be seen in [2, 3].

The constant monitoring of the energy flow combined with the integration of different energy generation sources will require management technologies capable of recognizing, predicting and acting in a way to guarantee quality, sustainability and efficiency, including costs, in energy consumption. Therefore, modern energy management systems should be able to monitor and exploit large volumes of data collected by various types of meters transmitted by digital channels mainly based on the IoT. The application of artificial intelligence techniques related with machine learning and big data will require thousands of meters collecting data at high resolution and high frequency (gigabytes per day), and, in order to assure the reliability and quality of this data, some aspects must be addressed such as the data model, the integration of information coming from several inputs or the data security.

The optimization of energy use will produce a direct reduction of costs and pollutants as the total energy consumption will be less. By increasing the share of renewable energy sources in the grid, the merit order will change. The merit order ranks the available energy sources from its operational cost, the cheaper ones being the first to meet the demand. Solar power generation and wind power generation are of the cheapest energy generation technologies, so if they are able to provide power, the operational cost of the last active power plant in order to meet demand will be less, allowing a more economic purchase of energy.

The path to reach a smart energy grid in the Industry 4.0 has already started. Development has been observed in the area of energy technologies, improving the efficiency of isolated systems. However, the overall energy efficiency can be greatly improved if multi-energy assets are analyzed and utilized in a more unified way. Energy assets can be interconnected physically in a plant, improving the energy usage in the plant and creating an EH. There is also the possibility to aggregate different plants physically or virtually, creating a digital entity of active prosumers that will be presented to the grid as a unique system that will be able to both consume and generate electricity.

This chapter is structured as follows. In Section 2, the VPP concept and tools are explained. First of all, its definition is exposed. This definition broadens the concept of EH and its functionality, creating a new entity able to perform an optimization considering internal and external factors. Secondly, the forecasting tools for predicting the situation at a stated horizon are presented. These tools include the forecast of renewable energy sources and demand and energy price from the grid. Third, the EH concept and method are developed for a general industry. Then, the optimization of the system is assessed, and resolution methods are proposed for obtaining high-quality results. In Section 3, some aspects related to the automation pyramid and the communication requirements of its levels are presented. Then some of the communication technologies and protocols are briefly introduced. Last of all, conclusions are drawn in Section 4.

2. Industry as a virtual power plant

One of the most important characteristics of the electrical grid is the constant balance between generation and consumption. With the rise of intermittent renewable energies, a degree of uncertainty is introduced. The discontinuity of this type of generation should not affect the fulfillment of the demand at every instant. With a proper management of energy assets and energy storage systems, renewable energy sources can be satisfactorily introduced without compromising the stability of the system. Once the balance between supply and demand is assured, there is leeway to generate an economic benefit from the energy transferred and stored inside a facility, such as a VPP. The VPP would be a power prosumer, meeting the local demand, and profit its own energy assets to trade energy with the external grid. Nowadays, the smart microgrid and prosumer concepts are being developed and tested in the tertiary sector, as can be seen in [4, 5]. Although the advancements are done, the presented ideas need further investigation. The prosumer smart grid approach can also be implemented in the industry, creating an energy-smart entity that will deal with the challenges and demands of the coming energy markets and will produce a profit from the exploitation of its own equipment against the external primary energy grids.

2.1 Virtual power plant concept

A VPP is a network of decentralized, medium-scale power-generating units as well as flexible power consumers and batteries. A VPP can be implemented in an industrial site, composed by all the controllable energy assets and the renewable energy generation units in the factory.

The VPP operates its energy assets efficiently taking into account the forecast of internal and external factors with the aim of maximizing the efficiency of the system in economic and environmental terms. As an example, internal factors can comprise coefficient of performance (COP) and efficiencies of energy equipment, energy storage capacity, energy generation at a given moment, cost of the different subsystems and reschedulable loads (both electrical and thermal). External factors may be constituted by electricity, natural gas and waste prices.

In **Figure 1** an example of a VPP is shown. It can be appreciated that the communication with the electrical grid is bidirectional, allowing to buy and sell electricity depending on the forecasted conditions. The working behavior lays in an energetic, economic and environmental evaluation that considers the forecasted input energy price, the forecast of available energy inside the VPP and the forecasted demand.

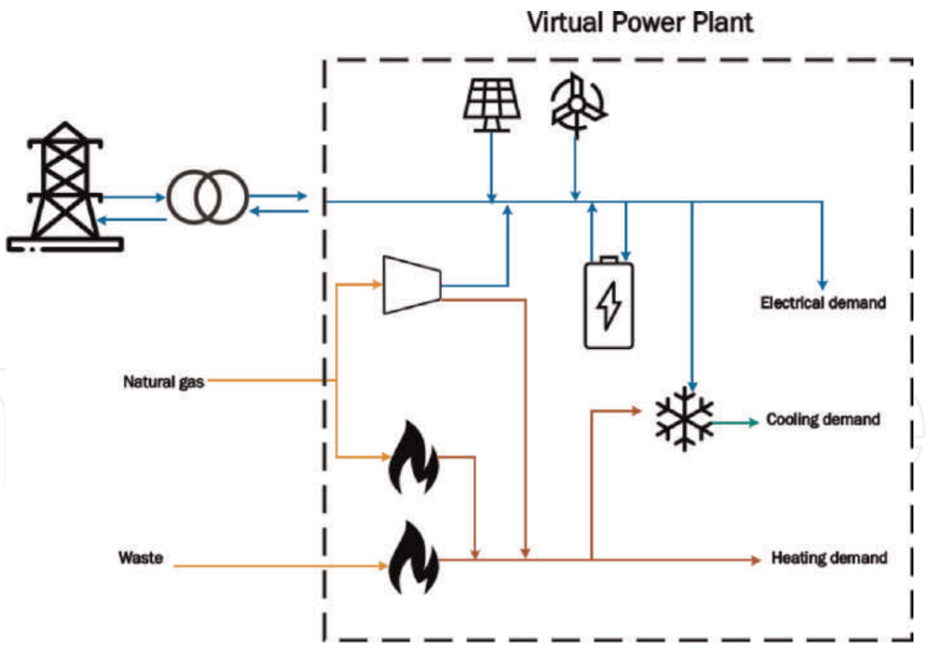

Figure 1.
Example schematic of a VPP where renewable energy sources (solar and wind) are present together with cogeneration system, boilers, absorption cooling and energy storage systems.

The benefits of implementing a VPP affect not only the industrial site itself but also the electrical grid through demand response (DR). The creation of a VPP out of an industrial facility will lead to:

- Integration of intermittent renewable energy, not only in the VPP but also in other points of the grid due to the electricity price response of the VPP. Also, expensive investments to expand the distribution network can be avoided if the generation is locally available.

- Integration of small electricity producers into the distribution network. The VPP itself is seen by the grid as a small electricity producer when the electricity cost is high, and thus there is a need to increase the generation at that moment.

- Optimization of energy use inside the VPP. The demand is analyzed, modeled and predicted using artificial intelligence method, and the optimal operation point of energy providers is computed.

- Optimization of the integration of electric vehicles (EV) for vehicle to grid (V2G) and grid to vehicle (G2V). The storage systems managing the surplus energy at the VPP can be combined with the EV batteries, which will work then as a part of the system. In this way not only the energy storage systems are improved, but also the EV-grid integration is made easier.

- Reduction of emissions. By integrating renewable energy sources and increasing the efficiency of the energy used, the emissions are directly reduced.

- Exploitation of energy assets. The systems present in a facility are nowadays not used in all its potential. With the implementation of a VPP, its working periods will be optimized according to internal and external factors and allowing an exploitation and efficient use of all energy carriers present in a system.

- Market opening. There are several facilities that will allow the creation of a VPP. However, their owners and operators are not aware of the possibilities and benefits it will produce. The introduction of a VPP in an industrial site will lead to a market opening that will encourage other similar facilities to take the same role, and thus the previous benefits will be amplified to the whole electrical grid.

- Autonomy and strong position of the owner of the facility in front of the operators of the electricity market that will allow a greater competitiveness market.

To implement the VPP features, the future energy status of the system should be continuously computed, which includes demand, generation of renewable sources and energy prices. This information leads to VPP operation including energy conversion and storage, which drives the EH, a crucial part of the VPP as it optimizes the path from energy input to demand. Once the forecast of the future situation and the model of the EH is obtained, the VPP is formed. The objective of the VPP is to fulfill local demand while, at the same time, exploiting its own energy assets to be able to trade electricity with the grid. During the modeling and the optimization of the VPP, the electricity exchange with the grid, the energy transfer with the energy storage system, the dispatch factors between the present transformers and the destination of power from the PV system are computed to assure an optimal operation from the economical, energetic and environmental points of view.

2.2 Future situation forecasting

Forecasting is the process of making predictions of the future based on past and present data analyzing the trends that appear. Forecasting can be qualitative or quantitative. For the application to a VPP, quantitative methods are more suitable, as they are based on past data to estimate future states and do not lay on subjective opinions. This approach extracts patterns of the available data and assumes that these are expected to continue in the future and are applied usually to short- and medium-term forecasts. There are several models used for forecast, and its suitability depends on the nature of the problem that is being studied. Examples of them are time series, causal and econometric forecasting and artificial intelligence. The forecast of several variables is needed to optimize the VPP. The demand, generation from renewable energy sources and electricity price from the grid are used in order to compute the optimal operation point of the VPP.

2.2.1 Renewable energy

The prediction of the renewable energy that is generated depends directly on the climatic conditions and the characteristics of the equipment. The prediction of weather conditions, i.e. sun irradiation and wind speed, can be obtained from the meteorology databases. Two types of renewable energy systems will be shown in this section: photovoltaics (PV) and wind power (WP) generation.

On the one hand, for a PV system, the most important factor in estimating its performance is solar radiation. The uncertainty in solar radiation is the largest source of error in the computation of the energy provided, as shown in [6]. The solar radiation depends on the orientation and the inclination of the area studied. Once this value is obtained, the theoretical energy output can be computed. However, the result should be corrected by adding a performance ratio that is influenced by factors such as shadows, dust, dirt, frost, snow, reflectance of the module surface, conversion efficiency, sunlight spectrum and temperature. As an example, in **Figure 2**, extracted from [7], the performance of different chemistries along temperature is shown. The value of the performance ratio (η) can be obtained statistically, and then the output power of the PV system will be:

$$P = P_{nom} \frac{G}{1000} \eta \tag{1}$$

where G is the received solar irradiance in W/m^2 and P_{nom} the peak power in kW.

On the other hand, for the case of wind turbines, there is a direct relationship between wind speed and energy output [8]. The extra parameter that has to be considered is air density, which can be computed using temperature and pressure and obtained from a meteorological database as with the wind speed. The output power can be computed with the data specified by using the wind turbine power curves provided by the manufacturer. These curves are obtained by the manufacturer by means of theoretical and statistical analysis of the performance of the turbine.

The previous methods are useful for a first assessment of the energy generated by the renewable sources. However, after the renewable energy sources equipment are installed and working on an industrial environment, the generation forecast can be improved by modeling specifically its behavior. A correlation of meteorological data with PV and WP output should be performed to assure high model accuracy and obtain the real efficiency and performance of the equipment. According to [9, 10], artificial neural networks (ANN) and support vector machine (SVM)-based forecasting methods are suitable for the modeling and prediction of the behavior of PV generation systems, while ANN, adaptive neuro-fuzzy inference systems (ANFIS) and autoregressive moving average (ARMA) perform well for WP generation.

2.2.2 Demand

The demand is the amount of load that the system has and the energy that is required to be fulfilled. Inside a VPP, this demand can be divided into two types: manageable and non-manageable. Non-manageable loads are those which run continuously or that cannot be controlled. Inside a VPP, the owner or end user can

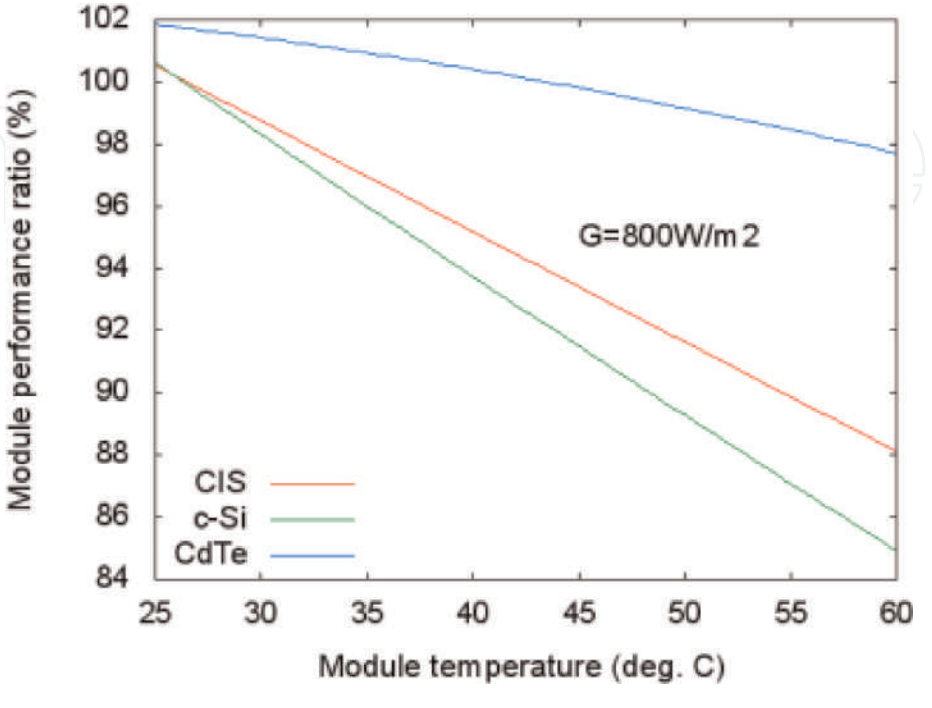

Figure 2.
Performance of PV modules with a solar radiation of 800 W/m².

decide which loads are manageable and which are not according to the business objective criteria. Manageable loads can be further divided into shiftable, interruptible and heating, ventilation and air conditioning (HVAC) loads. The forecasting of both types of demands follows a different way and will be now assessed.

2.2.2.1 Non-manageable loads

Classically, energy loads can be either electrical or thermal. The behavior of both types of demand lies in the same principles, so the prediction of them can be done using the same method. In recent times, the artificial intelligence methods that have been used for load forecasting (LF) include mainly neural networks, expert systems and support vector machines. Nowadays, the focus lays in the development of hybrid methods, combining different forecasting methodologies. For example, in [11] a LF method based on self-organized map and support vector machine is developed. The method is tested for prediction of the power consumption of a whole city. However, its suitability for an industrial site application has not been proven. In [12] an extreme learning machine with the Levenberg-Marquardt method is proposed, and in [13] the possibility to use artificial neural network to create a hybrid method with other techniques such as backpropagation, fuzzy logic, genetic algorithm and particle swarm optimization is shown. The industry is a sector where the demand can have an irregular and infrequent behavior depending on several conditions, and it is constantly under improvement processes. For this reason, a method that enables periodically auto-adjustment and high accuracy results is searched. ANFIS aim at mapping input to output for highly nonlinear processes such as energy management field. ANFIS was first introduced in [14] as a combination of two soft computing methods: artificial neural network and fuzzy logic. The ANFIS architecture is an adaptive network that uses supervised learning on learning algorithm, which has a function similar to the model of Takagi-Sugeno FIS [15]. This architecture is shown in **Figure 3**, extracted from [16].

In the first layer, the fuzzification of the inputs takes place. This is done by a membership function which can be a Gaussian membership function, a generalized bell membership function or other types of membership function. The parameters of this layer that define the membership function are called premise parameters. In the second layer, the fire strength of the rule is calculated. The output is the result of multiplying the signals coming into the node. In the third layer, a calculation of the ratio between the *i*th rule firing strength and the sum of all rules firing strength is

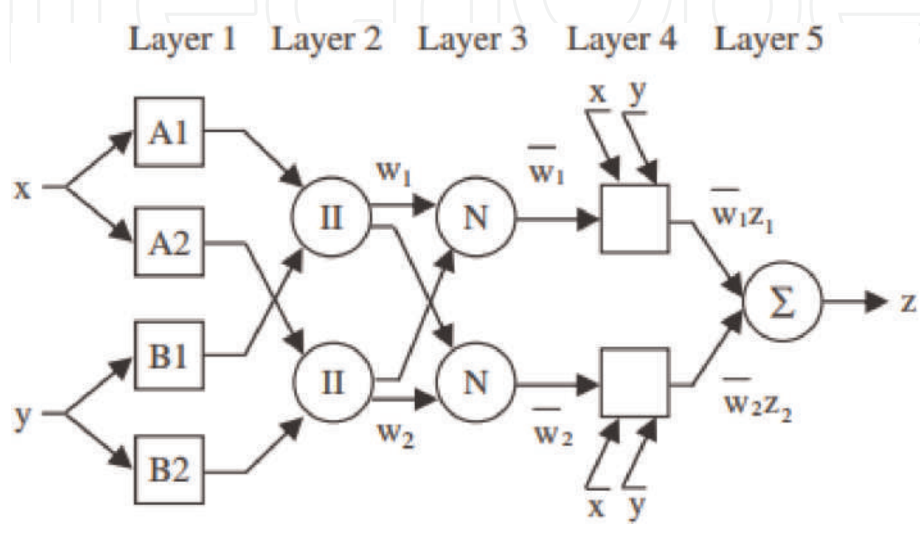

Figure 3.
ANFIS architecture.

done. The output is named the normalized firing strength. The fourth layer executes the Takagi-Sugeno fuzzy reasoning method. The parameters that appear here are the consequent parameters. Finally, in the last layer, the computation of the overall output as the summation of all incoming signals from previous nodes is done. It can be seen that the parameters that need to be trained are the premise and consequent parameters, present in layers 1 and 4. They can be obtained in the learning process by using the forward path and the backward path. During the forward path, the premise parameters are specified, while the consequent parameters change using a recursive least square estimation, and, during the backward path, the consequent parameters obtained remain fixed, while the error propagates to the first layer updating the premise parameter in a gradient descent way.

2.2.2.2 Manageable loads

According to [17], manageable loads can be divided into:

- Shiftable: Loads with predefined working cycles and load profiles. These loads appear between certain time limits which are specified by the end user. In an industry, these can be formed by noncritical processes with a variant energy consumption profile which can be rearranged on time depending on the production goals for the specific time interval.

- Interruptible: These loads are defined by its state, which can be either on or off. When its state is on the consumption remains constant. An example of a load of constant consumption is a water heater. The heating of water can be interrupted and restarted according to the time specification by the end user and the thermal inertia of the system.

- HVAC: Air conditioning and heating devices. Its consumption depends on parameters such as ambient conditions and comfort level specified by the end user.

The consumption of these loads depends on the situation on different factors regarding the state of the EH, the forecast of renewable energy input, the forecast of non-manageable demand and the price of energy from the distribution grids. The consumption of manageable loads is not forecasted but optimized inside a VPP according to restrictions specified by the end user with the objective of minimizing a utility function, which will be presented in the energy optimization section.

2.2.3 Energy price from the grid

In a future situation, demand side management (DSM) will be broadly implemented in the energy grids, specifically in the electrical grid. The price of the electricity is specified in the wholesale market with an anticipation of 24 h for each hour of consumption. In a situation where a VPP wants to interact with the market and obtain benefits from the exploitation of its energy assets, it is important to predict the price of the electricity in order to be able to optimize its energy carriers and offer or demand electricity from the grid.

In [18], two methods to predict next-day electricity demand and price daily curve are proposed given past curves: robust functional principal component analysis and nonparametric models with functional response and covariate. In [19], a hybrid methodology is proposed, combining autoregressive integrated moving average (ARIMA) with adaptive dynamic corrector lazy learning algorithm.

Although these methods were studied, due to the integration of renewable energies in the electricity market and the changes in the structure of the pricing that it supposes, during the last years, ANN have been the focus to forecast electricity prices. ANN models for short-term electricity modeling perform better than time series models such as ARIMA models, as shown in [20]. It is also verified that the performance of ANN depends on appropriate input parameters; clustering and data selection algorithms of k-nearest neighbor algorithm and mutual information methods were used. The problem of this model is the need to remove trend and seasonal components. In the electricity market, there are strong seasonal effects and other nonlinear patterns that harm ANN forecasting performance. In [21] a robust method to solve the seasonal problem with ANN is proposed and verified. The method is seasonal autoregressive neural network (SAR-NN) defined as a dynamic feedforward artificial neural network. In [16] a hybrid approach based on the combination of particle swarm optimization and ANFIS is proposed and demonstrated in a case study in Spain. The study shows that soft computing techniques such as neural networks can be much more efficient computationally and accurate if correct inputs are considered. To select the most suitable inputs, several methods can be used, and genetic algorithm (GA) is one of them. The combination of ANFIS with GA has been proved to solve market price prediction and other economic parameters, as shown in [22, 23].

2.3 Energy hub model

The energy conversion equipment of the VPP forms the EH. In order to develop the model and the optimization of the system to create a VPP, the EH should be modeled. An EH is a multi-carrier energy system consisting of multiple energy conversion, storage and/or network technologies and characterized by some degree of control. In **Figure 4** an example of a schematic of an EH can be seen. In the figure, it is possible to appreciate that the EH in this case is composed by the energy conversion equipment, excluding the storage system. The EH is nowadays understood as the set of energy drivers that allow energy management. However, with the implementation of the VPP concept, the energy management possibilities are expanded and can take place in a level above the EH. Thus, although in most cases energy storage is included inside the EH, when a VPP is implemented, the trading

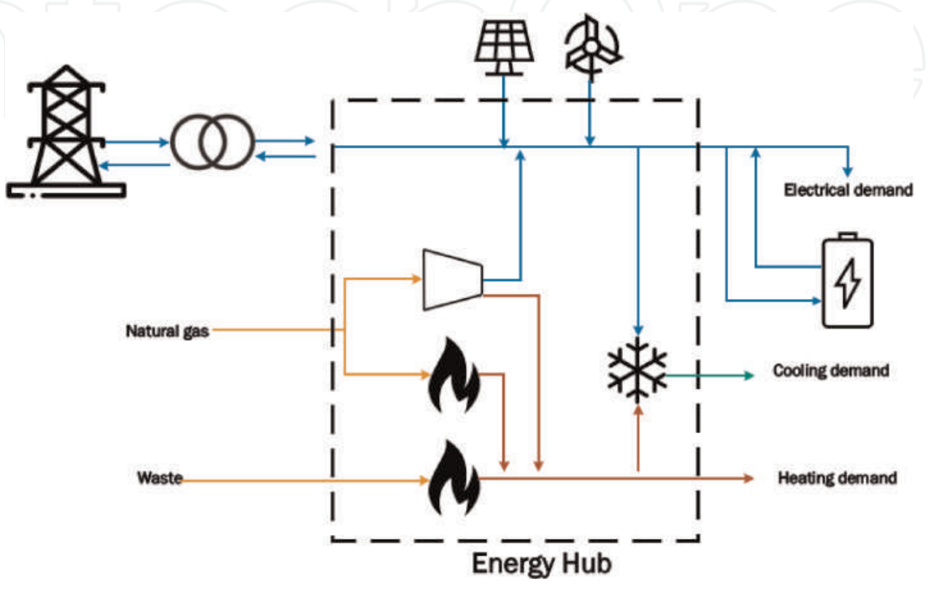

Figure 4.
Example schematic of an EH.

relationships are placed outside the EH, so it becomes coherent to also place the energy storage system outside the EH but inside the VPP.

In this section the formulation of an EH will be established from a generic perspective. According to [24], the relationship between input power and output power inside and EH is:

$$\begin{bmatrix} L_\alpha \\ L_\beta \\ \dots \\ L_\gamma \end{bmatrix} = \begin{bmatrix} \eta_{\alpha\alpha} & \eta_{\alpha\beta} & \dots & \eta_{\alpha\gamma} \\ \eta_{\beta\alpha} & \eta_{\beta\beta} & \dots & \eta_{\beta\gamma} \\ \dots & \dots & \dots & \dots \\ \eta_{\gamma\alpha} & \eta_{\gamma\beta} & \dots & \eta_{\gamma\gamma} \end{bmatrix} \begin{bmatrix} P_\alpha \\ P_\beta \\ \dots \\ P_\gamma \end{bmatrix} \tag{2}$$

where L represents the demand, P the power input and η the coupling matrix. It has to be observed that according to the example proposed, the energy coming from the electrical grid and the energy coming from the battery can be placed both in the demand and in the generation side.

The determination of the coupling matrix needs to be assessed taking into account the amount, characteristics and interconnections of the energy equipment. In the following paragraphs, an outline of relationships depending on different situations is carried out. These basic rules form the information needed to develop the model for more complex systems. With these, it will be possible to establish the coupling matrix that represents the EH and which relates the generation side with the demand side.

2.3.1 Energy converter with one input and one output

In this case an energy converter β with an input energy P_α has one only output: L_β. The power relationship between input and output is represented by:

$$L_\beta = P_\alpha \eta_{\beta\alpha} \tag{3}$$

where $\eta_{\beta\alpha}$ is the performance indicator of the converter, which can be the COP or the efficiency depending on the equipment considered. The COP can be constant or can be dependent on different parameters such as temperature or operating point.

2.3.2 Energy converter in series

This case represents the situation where all the output from one energy converter goes directly to another energy converter. This is called multistage energy conversion. The power output at the end of the last energy converted is computed by multiplying all the COPs in the chain. For the case with two energy converters:

$$L_\theta = P_\alpha \eta_{\beta\alpha} \eta_{\theta\beta} \tag{4}$$

2.3.3 Available energy in a converter

The power provided by an energy converter or energy source can be supplied to several energy converter or demand points. Power can be given to these systems simultaneously as long as there is energy available in the energy converter or generator. This can be represented mathematically as:

$$\sum_{i=1}^{n} P_{\alpha i} \leq P_{\alpha} \tag{5}$$

where:

$$P_{\alpha i} = P_{\alpha i} v_i \tag{6}$$

v_i being the dispatch factor to the different demands connected to the same source.

2.3.4 Upper and lower production limits

Every energy conversion equipment has a range within which it is possible to generate or convert electricity. It has to be assured that the energy that passes through the equipment falls between the specified thresholds. Mathematically it is expressed as:

$$lb_{\gamma} \leq P_{\alpha} \eta_{\gamma\alpha} \leq ub_{\gamma} \tag{7}$$

where lb_{γ} and ub_{γ} are the lower and upper limits, respectively.

The basic rules for the proper development of the coupling matrix have been explained. Their logic can be applied to any system composed by interconnected energy assets to develop the mathematical model of an EH.

2.4 Energy optimization

The optimization is an essential step for the successful implementation of a VPP. Once the model of the system has been developed, an evaluation of the state of the plant at a specified number of time instants has to be carried out to achieve all the benefits mentioned in this chapter. The optimization will allow to reach the best efficiency in the use of resources from an economical and environmental perspective as well as facilitate to the grid the integration of active prosumers, demand side management (DSM) and renewable energy sources.

An optimization is the selection of the best solution for a specified problem. The simplest optimization problems deal with the maximization or minimization of a variable. In mathematics, conventional optimization problems are usually stated in terms of minimization. A general manner to represent one of these is:

$$\text{Given}: f : A \rightarrow \mathfrak{R}$$

$$\text{Find}: x_0 \in A \text{ such that } f(x_0) \leq f(x) \text{ for all } x \in A$$

For the purpose here assessed, f can be considered as the energy of the system that is being considered, the operational and maintenance cost, the environmental impact or any other aspect related to the exploitation of energy assets. The function f is the *objective function* that wants to be minimized. A is a subset of the real space that is understood as a set of constraints that needs to be achieved or fulfilled. It is represented as group of equalities and inequalities that the solution should meet to be valid. In the energy frame, these equations deal with factors such as meeting the demand and comply with the operational bounds of the system. The domain A of f is called the *search space,* and the elements x in A are called *candidate solutions*. There are several types of optimization problems and possible solutions depending on the nature of the situation that is being studied. For a system where several energy assets are present and a time optimization has to be carried out, multi-period mixed-integer problems are the ones that represent the most of its operation, as can be seen in [25].

There are different purposes that lead to the decision of building a VPP, as, for example, total energy use, energy cost, production scheduling and emissions. All these factors have to be reflected in the objective function. The most used method to handle multi-criteria decisions is the weighted global criterion method. This method allows the interested party to adjust the preferences of the system. The objective function is obtained as:

$$f = \sum_{j=1}^{N} f_j^{trans} w_j \tag{8}$$

where f_j^{trans} is a normalized value of a single objective function and w_j the relative weight assigned to that objective function. f_j^{trans} is created in order to obtain the same range for the different objectives contemplated and has to be calculated as:

$$f_j^{trans} = \frac{f_j(x,y) - f_j^{\min}}{f_j^{max} - f_j^{\min}} \tag{9}$$

where f_j^{max} and $f_j^{\min}$ are maximum and minimum values of the objective function in question, respectively.

In order to obtain the optimal operation point of the VPP, the optimization process should be performed in two stages. The first stage deals with the decision of where to introduce or extract energy from the battery, decision of selling or buying energy from the electrical grid and the scheduling of manageable loads. The scheduling horizon of this optimization is normally one day, as this is the time interval at which the electricity price from the market is known. The scheduling horizon is divided into time slots; usually there are 96 time slots per day, one every 15 minutes. As shown in [17], the objective function in this optimization case is formed by three terms: energy cost, scheduling preferences and climatic comfort. For the case of the energy cost, it can be expressed as:

$$f_1^1 = B\sum_t P_{BE}C_{BE} + A\sum_t P_{CB}C_{CB} + (1-B)\sum_t P_{SE}C_{SE} + (1-A)\sum_t P_{DB}C_{DB} \tag{10}$$

where A and B are Booleans that designate if the VPP if selling/buying electricity from the grid and charging/discharging the battery. The other parameters refer to the following:

- P_{BE}: energy bought from the electrical grid
- C_{BE}: cost of the energy bought from the electrical grid
- P_{CB}: energy inserted in the battery
- C_{CB}: cost for inserting energy in the battery
- P_{SE}: energy sold to the electrical grid
- C_{SE}: cost of energy sold to the electrical grid. It has to be noted that this value is negative

- P_{DB}: energy extracted from the battery
- C_{DB}: cost of the energy extracted from the battery

The objective function related to the scheduling is expressed as:

$$f_2^1 = \frac{\sum_{SL} \sum_t \gamma}{N_{SL}} \tag{11}$$

where γ is a scheduling preference parameter and N_{SL} is the number of scheduling loads. Last of all, the objective function for the comfort is:

$$f_3^1 = g^{max} + \frac{\sum_{SL} \sum_t g_r}{RT} \tag{12}$$

where g^{max} is the maximum temperature gap allowable, g_r is the real temperature gap, R are the rooms considered and T are the time slots. For this first optimization stage, the restrictions should contain the fulfillment of non-manageable loads, the characteristics of manageable load (working cycles, minimum number of consecutive ON slots, maximum number of consecutive slots OFF, etc.) and power restriction on the energy input.

Once the energy input and output from the grid, batteries and loads are obtained, the second stage deals with the optimization of the energy flow inside the EH. In this case the objective functions are related to maximizing the efficiency and minimizing the energy cost and the total emissions. The function that represents the total energy use can be represented as:

$$f_1^2 = \sum_{\alpha} \sum_{t} P_t^{\alpha} \tag{13}$$

where P_t^{α} represents the energy generated or converted by α at the time instant t. It can also represent the energy input to the VPP such as the electricity from the grid and the natural gas. For the case of the cost of the system, the objective function is:

$$f_2^2 = \sum_{\alpha} \sum_{t} P_t^{\alpha} \lambda^{\alpha} \tag{14}$$

where λ^{α} represents the cost of the energy for a converter or energy input α. Last of all, for considering the emissions of the system:

$$f_3^2 = \sum_{\alpha} \sum_{t} P_t^{\alpha} e^{\alpha} \tag{15}$$

where parameter e^{α} represents the emission factor of the energy provided by α. For this stage, the restrictions should include the fulfillment of the demand and the power limitation of the different energy converters inside the EH.

3. Communication architecture and data management

As it has been mentioned in the previous section, forecasting techniques based on data-driven models are widely used when dealing with energy-related variables. This kind of models usually needs huge amounts of information to properly train or tune their inner structures, and once the models are generated, the central controller must be capable of sending the forecasted schedule decisions to each system's local controller. To do so, not only a sensor network has to be deployed in the facility, but also an efficient data communication system is needed.

Therefore, one of the key elements of the VPP concept is the communication systems. The existence of reliable, accurate, efficient and safe data exchange is crucial for a bidirectional, near real-time information flow. In addition, the current trend in the field is to make use of a service-oriented architecture (SOA), enabling an easy integration of the plant data in systems that can analyze and optimize not only the operation of the facility itself but also the global operation of the whole energy grid. To this extent, the cloud computing platforms such as Amazon Web Services, Microsoft Azure or Google Cloud.

The cost of implementing a communication system can be high, so it is vital to select a suitable data communication technology. There are several wired and wireless technologies available that can provide the required communication infrastructure. The selection of one (or more) of these communication technologies will depend on the quality of service (QoS), data range, reliability, latency, economic viability, etc. The capabilities offered by these technologies are also strongly related to the VPP grid structure. Looking it from the prosumer point of view, the main automation system is the energy management system (EMS) which is responsible for the management and optimization of the energy assets supervised in the VPP.

3.1 Energy management systems

The term energy management system (EMS) refers to an integrated system that enables the monitoring, management and control of several devices providing the necessary support for an effective operation of electrical generation and transmission facilities.

At a high level, the architecture of an EMS is divided into three layers which are management, automation and field levels [26] as depicted in **Figure 5**. The management (or supervisory) level comprises the human interface with the system by means of human machine interfaces (HMI) or SCADA-like software systems and contains most of the system logic and modules related with data analysis. The automation (or local) level provides the primary control devices connected via networked controllers and usually operating via BACnet, ZigBee, etc. protocols. The field (or plant) level represents the physical devices like energy meters, sensors and actuators installed to the plant equipment. These devices should be connected to local controllers by means of field-bus communications to allow control functionalities.

VPP supervision and control systems can be centralized or decentralized [27]. In the centralized control, all the knowledge about the devices in the VPP and the energy market is located in the central controller. Although this is a simple solution in most of the cases, when dealing with a large number of devices, the optimization of the control strategy can become computationally expensive for the central controller. In a distributed or decentralized control, the complexity is divided vertically within the VPP. Local controllers supervise and define the control strategy, and a higher-level controller coordinates their decisions in order to reach a global optimum state.

3.2 Communication requirements

The architecture defined above is organized in three hierarchical levels. Each of these communication layers has its requirements in terms of bandwidth, latency or cyber security. For example, at the field level, to have a large bandwidth is not a

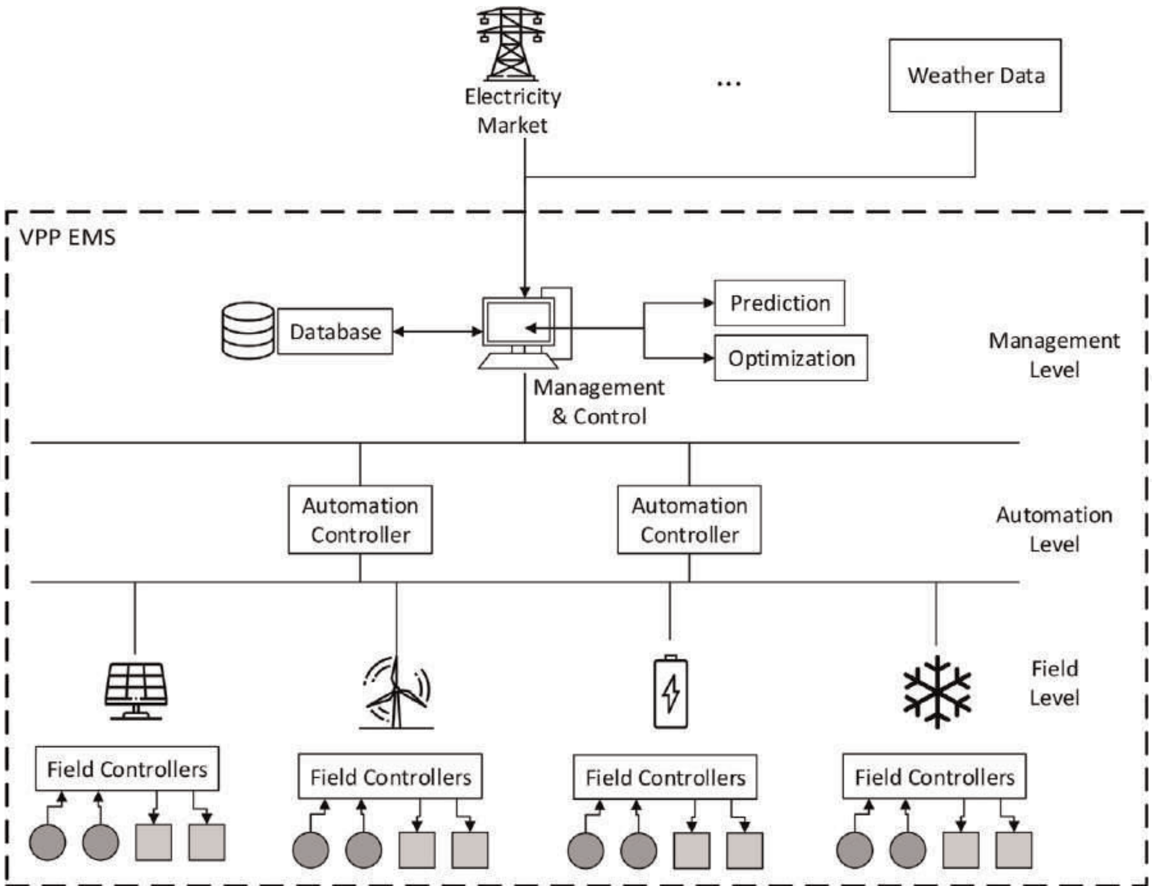

Figure 5.
EMS three-level architecture.

common requirement, but a short latency is mandatory given the near real-time control performed at this level.

3.2.1 Field level requirements

The total amount of data sent per node per transmission is typically less than a hundred bytes. That being the case, the communication bandwidth at this level is well within 100 kbps [28]. The sampling and transmission frequency are commonly between a range of 5 and 15 min. A simulation carried out in [29] showed that larger data collection frequencies fail at detecting short-term voltage anomalies. Besides, a time synchronization service is required to refer all the data gathered in the plant with respect to the UTC. A general-purpose time synchronization service like the network time protocol (NTP) is used given that the accuracy required does not exceed the order of seconds.

Typically, the sensors manage analogical data that is then is handled to an analog to digital converter (ADC) followed by an interface to a process control computer. The sensors can also have a digital communication module and contain embedded digital electronic processing systems. Actuators work in a reverse sense, converting electrical signals to the appropriate physical variable. However, as they have to amplify the energy level to produce the change in the real variables, actuators are high-power devices, while sensors are not.

3.2.2 Automation level requirements

At the automation level, the data from several local controllers is received; typically, the order of system it aggregates is in the order of tens. Hence, a bandwidth of more or less 1 Mbps is enough to fulfill its requirements [28]. The time synchronization and latency are also limited like in the field level.

The automation level is in charge of several tasks such as the monitoring of the variables to check the system or component failure, the management of the set points for the important process variables and the control reconfiguration and tuning of the control loops.

3.2.3 Management level requirements

The management level shares a large part of the requirements of the automation level. Typically, in this layer, the main limits for its requirements are represented by the capabilities of the already existing communication infrastructure.

Here, the information arrives as time series type of data; this data is characterized by having a timestamp associated with each value. In the management level, this data is collected and analyzed to perform some actions like process scheduling or maintenance management.

3.3 Communication technologies and protocols

When a message is transmitted onto a bus, it has to contain information like the identifier of the sending device, the message or data to transmit, the destination device address and some additional information (e.g. for error checking). After that, when the message reaches the destination device, this one has to know not only the message codification but also how to handle its reception using procedures to avoid collisions and prioritization.

These rules about connectivity and communication are defined by the communication system protocol. These protocols for VPP system must adhere to several criteria: efficient and reliable communication, interoperability with other systems and integration into the power system. For easier integration, it is usually desirable that the VPP system supports the communication protocols already in use by any other equipment. In addition to standardized protocols, there are many proprietary protocols like C-Bus or PROFIBUS.

Both wired and wireless technologies have been specified through standards. The advantages of wired technologies over wireless ones are the higher data transmission rate, security and reliability but at the expense of high installation cost. On the other hand, wireless technologies have fewer installation costs and can be easily deployed, but they exhibit low data transmission rates and signal interference problems. With the advent of ICT and IoT, more and more sensors and meters are needed to be integrated, monitored and controlled. In this situation, the lower deployment cost and better scalability of wireless technologies make them better candidates. In the below sections, some of the widely used communication technologies for metering and sensory purpose will be covered.

3.3.1 Power line carriers

In terms of wired technologies, PLC is the most widely used technology [30]. Power line carriers (PLCs) consist of introducing a modulated carrier signal over the existing electricity grid. No additional wiring is required; therefore, PCL can be considered as a cost-effective and straightforward solution. PLC can be classified into two major categories: narrowband PLC and broadband PLC.

The operating rate of the narrowband PLC is in a range of 3–500 kHz. It can be further classified as low data rate and high data rate narrowband PLC. The former is a single carrier technology with data rate up to 10 kbps and works on the recommendations of standards like LonWorks or KNX. The high data rate narrowband is a multi-carrier technology with a data rate below 1 Mbps. The broadband PLC technology has an operating range of 2–250 MHz with a data rate of hundreds of Mbps.

PLC technologies have been used since a long time ago for electric energy-related services in industrial automation like remote meter reading and remote load management. PCLs can be applied in any point of the VPP environment, and its main advantage is the low running costs, and that can be installed using current infrastructure. The security issues are solved like in the ZigBee technology, using the 128-bit AES encryption.

3.3.2 GSM and GPRS

Global System for Mobile Communications (GSM) is known as the world's most deployed cellular technology. It operates on the 1800 MHz and 900 MHz bands, and its data rate is up to 270 kbps. General Packet Radio Service (GPRS) data rate is much larger than GSM. Its main drawback is the reliability of Short Message Service (SMS) in case of network congestion.

The main application of GPRS and GSM is in smart metering solutions for remote billing and power consumption monitoring, usually applied in smart grids covering from the generation stage to the consumption one, including both the transmission and distribution.

3.3.3 WiFi

Wireless sensing technology has been gaining popularity in the last years given the fact that wireless sensors are easy to install and cheaper in price and, among all the wireless sensing technologies, WiFi is the most popular. Developed under the IEEE 802.11 standards family, it provides a robust performance even in noisy channels and supports a wide range of data rates. The local security issues are tackled by the WPA2 protocol based on the 128 bit AES encryption technique, and to ensure secure communication through public Internet access, virtual private networks (VPNs) are typically used [31].

WiFi is the most dominant wireless technology for the high speed it can offer but is more expensive than other technologies because of its higher consumption and device price. WiFi is mostly used for building automation, remote control, meter reading, etc. in the tertiary sector and has been used as a proxy for human occupancy in some HVAC actuation models.

3.3.4 Ethernet

Ethernet is a low-cost communication method and is widely used for communication between PLCs and SCADA systems. Ethernet is available like optical fiber, shielded twisted pairs or coaxial cables. Among these, optical fiber is more secure and popular due to the absence of electromagnetic interference and electrical current. Ethernet uses carrier-sense multiple access with collision detection (CSMA-CD) methods for sensing data. Ethernet is not suitable for real-time application because the a priori estimation of the data packet maximum transmission time is impossible.

The main disadvantage of Ethernet is its wired nature and the need of deploying a new cable network. However, it is robust and does not have running costs. The most common implementation of Ethernet in today's industrial automation field is to use an Ethernet/IP network, applying the capabilities of traditional Ethernet to connect different facilities in the same network via the Internet.

3.3.5 Modbus

Introduced by Modicon Corporation, it is widely used due to its simplicity and reliability. It includes a remote terminal unit (RTU), transmission control protocol (TCP) and ASCII mode of transmission and supports RS-232, R-422, RS-485 and Ethernet-based equipment. Because of its simplicity and open-source availability, it is popular for local communication building and also has become the standard for industrial SCADA systems.

The security issues are not addressed in Modbus. It does not support authentication nor encryption; thus, it is less secure and more vulnerable to cyberattacks.

3.3.6 OPC UA

The OPC UA is a machine-to-machine communication protocol for industrial automation developed by the OPC Foundation. It is the next generation of the original OPC which is applied in different technologies like building automation or process control. OPC UA was developed to tackle the emerging needs of industrial automation.

OPC UA was designed to be fully scalable and enable both the horizontal and vertical communications across all the layers. In addition, it uses a service-oriented architecture, and two transport protocols are defined: an optimized TCP for high performance and a HTTP/HTTPS web service with binary or XML-coded messages.

Table 1 shows a summary of the main characteristics of each of the communication technologies reviewed.

3.4 Selection of sensing solution

According to [32], the factors that influence the selection of sensing and metering solutions are the following:

- Accuracy: In Europe, the accuracy of meters is defined by directives such as the Measuring Instruments Directive (MID). A common feature in this kind of directives is to classify the meters by their percentage accuracy.

- Ease of deployment: The ease of deployment refers to the different installation and networking challenges that must be tackled. For example, wireless sensors have reduced installation costs and provide better flexibility than their wired counterpart. Other factors to consider are the interoperability, installation in an accessible location or safety regulations.

- Communication protocol: As it has been seen in the previous section, there is a wide range of communication technologies each with its advantages and disadvantages.

- Resolution: The resolution determines the possible level of analysis that can be performed. As aforementioned the typical data collection rate is within a range between 5 and 15 minutes.

- Cost: The cost of the equipment is always a driver when deciding the metering equipment. Both initial costs and operating costs must be considered. Usually,

Technology	Type of technology	Characteristics
PLC	Wired	• Low installation costs (no additional wiring is required) • Cost-effective, widely used solution • Narrowband PLC: up to 500 MHz with a data rate below 1 Mbps • Broadband PLC: up to 250 MHz with a data rate of hundreds of Mbps
GSM/GPRS	Wireless	• World's most deployed wireless technology • Operates on 900 and 1800 MHz bands • Rate up to 270 kbps • Low reliability in congested networks
WiFi	Wireless	• Most popular wireless technology • Robust even in noisy channels • Security issues tackled by the WPA2 protocol
Ethernet	Wired	• Low-cost solution • Not suitable for real-time sensing • Needs a new cable network
Modbus	Comm. protocol	• Simple and reliable • Open-source • Standard for SCADA systems • Vulnerable to cyberattacks
OPC UA	Comm. protocol	• Robustness • Scalable and platform independent • Standard transport and encoding protocols (TCP and HTTP)

Table 1.
Summary of characteristics of the technologies and protocols reviewed.

the number of sensors is limited to the minimum to provide adequate control and ensure compliance with regulations.

- Availability: The geographical availability of a particular manufacturer's sensing solution. It will affect to the delivery time and provisioning of technical support.

4. Conclusions

In this chapter the concept of VPP has been explained as the solution for the challenges of upcoming energy markets. The forecasting of future energy situation regarding demand, energy prices and renewable generation has been assessed, reaching the conclusion that artificial intelligence methods are best suited for the stated purpose. The internal energy assets have been modeled by means of an EH. By adding these factors, the VPP is constructed, and its optimization can be carried out. The optimal operation point is obtained by considering current and future energy prices from the market, renewable energy generation, manageable and non-manageable demands and costs and operation constraints of energy equipment. For it to be possible, the EMS and the communication technologies of the plant have to be studied and adapted. The high-level structure and requirements of the EMS have been explained together with the more common communication technologies and protocols. Its advantages and drawbacks have been presented and the important factors for the selection of the sensing technologies described. By incorporating all the exposed factors in an industrial plant, a VPP can be created which will satisfactorily help the energy grid to evolve and will also produce a benefit for the exploitation of its own energy equipment.

Author details

Eva M. Urbano and Víctor Martínez Viol*
MCIA Research Centre, Polytechnic University of Catalonia, Barcelona, Spain

*Address all correspondence to: victor.martinez.viol@upc.edu

References

[1] Chan Y, Kantamaneni R. Study on energy efficiency and energy saving potential in industry and on possible policy mechanisms. 2015; 1-461

[2] Davatgaran V, Saniei M, Mortazavi SS. Optimal bidding strategy for an energy hub in energy market. Energy. 2018;**148**:482-493

[3] Choi S, Min SW. Optimal scheduling and operation of the ESS for prosumer market environment in grid-connected industrial complex. In: 2017 IEEE Ind Appl Soc Annu Meet IAS 2017 2017. 2017. pp. 1-7

[4] Mengelkamp E, Gärttner J, Rock K, et al. Designing microgrid energy markets: A case study: The Brooklyn microgrid. Appl Energy. 2018;**210**: 870-880

[5] Barzegkar-ntovom GA, Kontis EO, Kryonidis GC, et al. Performance assessment of electrical storage on Prosumers via pilot case studies. In: 2019 1st Int Conf Energy Transit Mediterr Area (SyNERGY MED). 2019. pp. 1-6

[6] Huld T. Estimating solar radiation and photovoltaic system performance, the PVGIS approach. In: AFRETEP 1ST Reg Work; Kampala. 2011. pp. 1-84

[7] European Commission. Photovoltaic Geographical Information System (PVGIS). Available from: http://re.jrc.ec.europa.eu/pvgis.html [Accessed: 12 March 2019]

[8] Vestas Wind Systems A/S. General Specification V90–3.0 MW 60 Hz Variable Speed Turbine General Specification for V90–3.0 MW. 2004. pp. 1–31

[9] Chen Q, Folly KA. Wind power forecasting. IFAC-PapersOnLine. 2018; **51**:414-419

[10] Das UK, Tey KS, Seyedmahmoudian M, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews. 2018;**81**:912-928

[11] Shu F, Luonan C. Short-term load forecasting based on an adaptive hybrid method. IEEE Transactions on Power Apparatus and Systems. 2006;**21**: 392-401

[12] Li S, Wang P, Goel L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Transactions on Power Apparatus and Systems. 2016;**31**:1788-1798

[13] Baliyan A, Gaurav K, Kumar Mishra S. A review of short term load forecasting using artificial neural network models. Procedia Computer Science. 2015;**48**:121-125

[14] Jang JR. ANFIS: Adaptive-network-based fuzzy inference system. 1993;**23**. DOI: 10.1109/21.256541. [Epub ahead of print]

[15] Takagi T, Sugeno M. Derivation of fuzzy control rules from human operator's control actions. IFAC Proceedings. 1983;**16**:55-60

[16] Pousinho HMI, Mendes VMF, Catalão JPS. Short-term electricity prices forecasting in a competitive market by a hybrid PSO-ANFIS approach. International Journal of Electrical Power & Energy Systems. 2012;**39**:29-35

[17] Agnetis A, De Pascale G, Detti P, et al. Load scheduling for household energy consumption optimization. IEEE Transactions on Smart Grid. 2013;**4**: 2364-2373

[18] Aneiros G, Vilar J, Raña P. Short-term forecast of daily curves of electricity demand and price. International Journal of Electrical Power & Energy Systems. 2016;**80**:96-108

[19] Vaccaro A, El-Fouly THM, Canizares CA, et al. Local learning-ARIMA adaptive hybrid architecture for hourly electricity price forecasting. In: 2015 IEEE Eindhoven PowerTech, PowerTech 2015. 2015. DOI: 10.1109/PTC.2015.7232253. [Epub ahead of print]

[20] Keles D, Scelle J, Paraschiv F, et al. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Applied Energy. 2016;**162**:218-230

[21] Saâdaoui F. A seasonal feedforward neural network to forecast electricity prices. Neural Computing & Applications. 2017;**28**:835-847

[22] Svalina I, Galzina V, Lujić R, et al. An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices. Expert Systems with Applications. 2013;**40**:6055-6063

[23] Wei LY. A hybrid model based on ANFIS and adaptive expectation genetic algorithm to forecast TAIEX. Economic Modelling. 2013;**33**:893-899

[24] Kampouropoulos K, Andrade F, Sala E, et al. Multiobjective optimization of multi-carrier energy system using a combination of ANFIS and genetic algorithms. IEEE Transactions on Smart Grid. 2018;**9**:2276-2283

[25] Kampouropoulos K. Multi-Objective Optimization of an Energy Hub Using Artificial Intelligence. Universitat Politècnica de Catalunya; 2018

[26] Brooks DJ, Coole M, Haskell-Dowland P, et al. Building Automation & Control Systems: An Investigation into Vulnerabilities, Current Practice & Security Management Best Practice. Perth; 2017. Available from: https://goo.gl/RM7ukP

[27] Zia MF, Elbouchikhi E, Benbouzid M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Applied Energy. 2018;**222**:1033-1055

[28] Pasetti M, Rinaldi S, Manerba D. A virtual power plant architecture for the demand-side management of smart prosumers. Applied Sciences. 2018;**8**:432

[29] Wan Yen S, Morris S, Ezra MAG, et al. Effect of smart meter data collection frequency in an early detection of shorter-duration voltage anomalies in smart grids. International Journal of Electrical Power & Energy Systems. 2019;**109**:1-8

[30] Erlinghagen S, Lichtensteiger B, Markard J. Smart meter communication standards in Europe—A comparison. Renewable and Sustainable Energy Reviews. 2015;**43**:1249-1262

[31] Kolenc M, Nemček P, Gutschi C, et al. Performance evaluation of a virtual power plant communication system providing ancillary services. Electric Power Systems Research. 2017;**149**:46-54

[32] Ahmad MW, Mourshed M, Mundow D, et al. Building energy metering and environmental monitoring—A state-of-the-art review and directions for future research. Energy and Buildings. 2016;**120**:85-102

Chapter

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

Francisco Arellano Espitia and Lucia Ruiz Soto

Abstract

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

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

1. Introduction

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

Meanwhile, deep learning (DL), or also known as deep neural networks (DNN), has become an analytical tool that has attracted more and more attention from researchers in different areas of research in recent years. The main skill DNN has the ability to learn and extract useful patterns from the data. Therefore, there is currently a tendency to make use of this ability of DNNs to extract significant features from complex manufacturing systems, in order to find the characteristic patterns of faults and thus be able to diagnose anomalies in a timely manner.

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

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

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

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

2. Deep neural networks

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

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

2.1 Convolutional neural network

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

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

2.2 Auto encoders

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

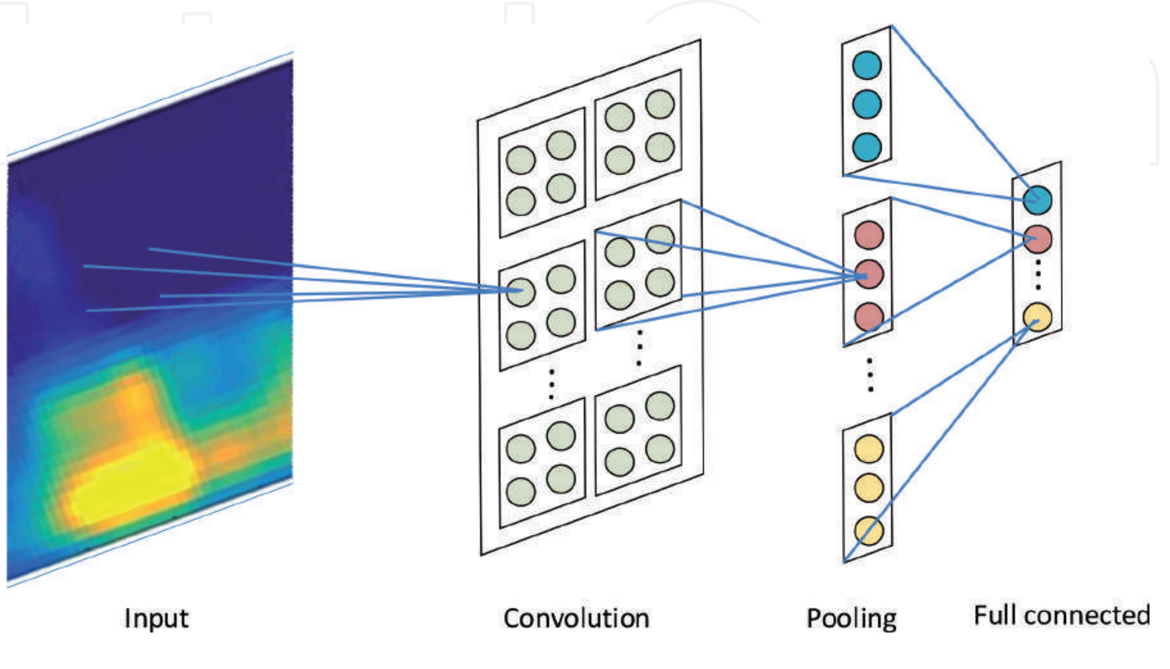

Figure 1.
Architecture of a CNN comprising commonly used layers as convolution and pooling.

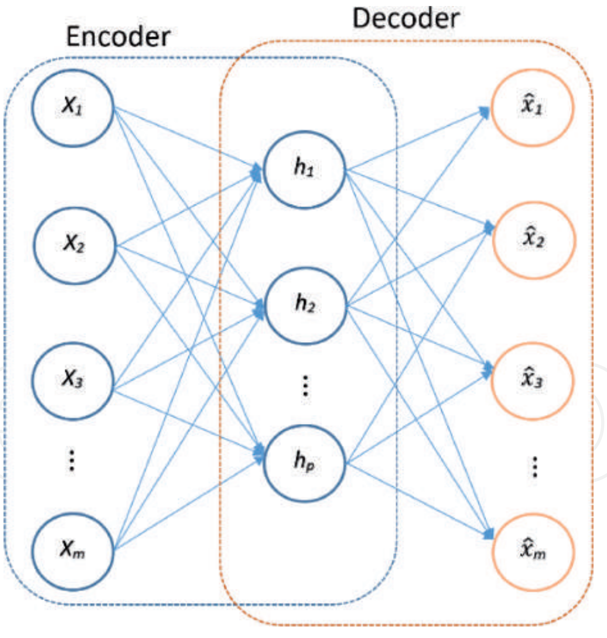

Figure 2.
Architecture of an auto-encoder.

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

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

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

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

2.3 Restricted Boltzmann machine

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

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

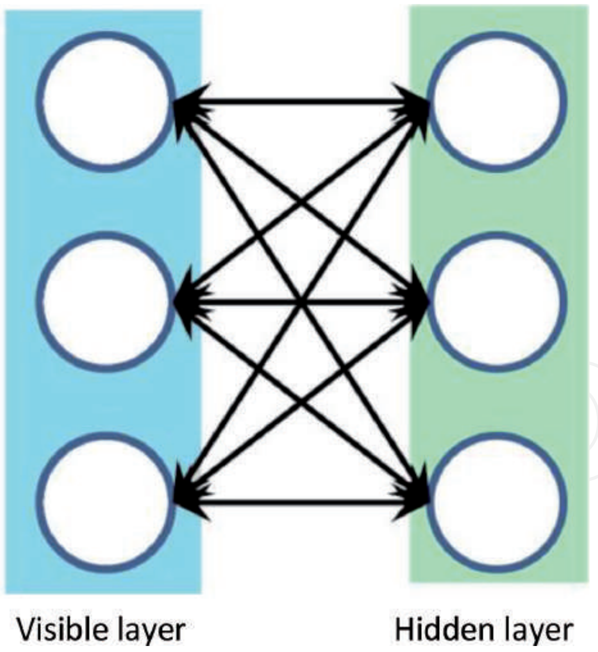

Figure 3.
Schematic illustration of restricted Boltzmann machine.

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

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

3. Applications of deep learning in condition-based monitoring

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

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

3.1 Feature extraction

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

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

3.2 Dimensionality reduction

Deep learning has attracted attention in several fields of study because it allows the extraction of features from complex signals and the processing of large data. Although the application of deep learning in the diagnosis of faults in industrial machines has concentrated on the automatic extraction of features, the utility of these tools goes further; a clear example is the application of DNN structures for the compression or reduction of dimensionality of data. As we have seen above, structures based on DNN are able to learn intrinsic relationships of the data; however, during this learning process, it is possible to generate a reduced representation of the data. A structure based on DNN capable of learning a coded and reduced representation is the so-called auto-encoder. Unlike linear dimensional reduction techniques, such as principal component analysis (PCA) and linear discriminant analysis (LDA), a structure of stacked auto-encoders can provide a non-linear representation that was learned from the data provided. Therefore, a reduction of dimensionality based on the auto-encoder can provide a better representation that helps to discriminate between the conditions of the machine. An example of the difference between the application of a linear technique for the reduction of dimensionality and one based on the auto-encoder is shown in **Figure 4(a)** and **(b)**, correspondingly.

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

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

3.3 Novelty detection

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

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

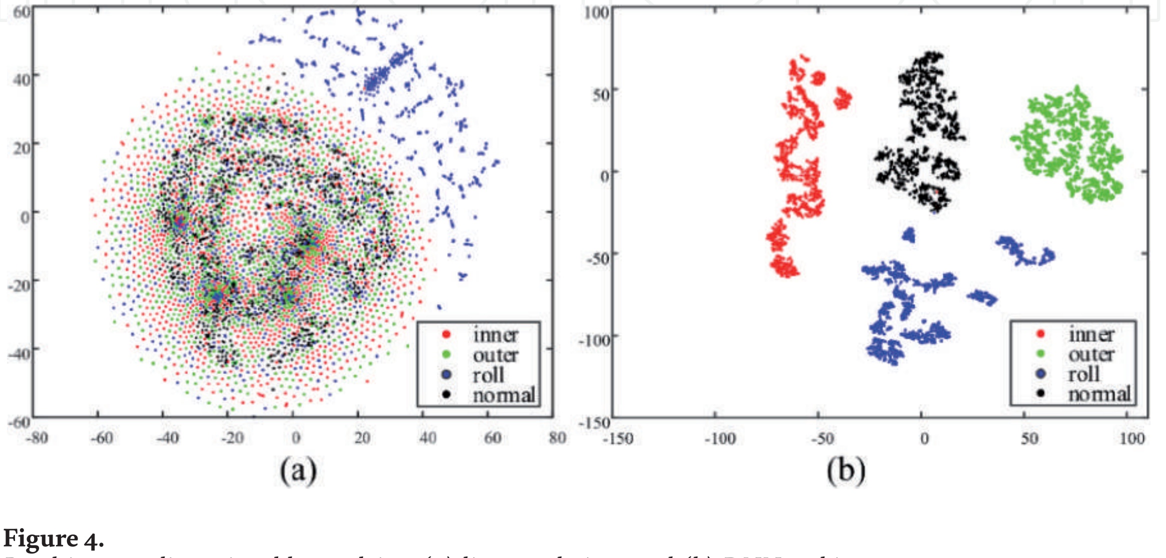

Figure 4.
Resulting two-dimensional by applying: (a) linear technique and (b) DNN architecture.

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

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

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

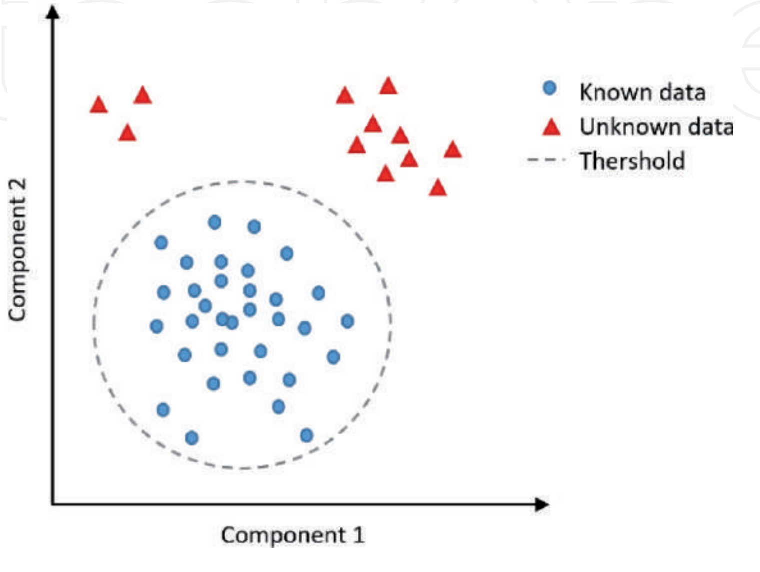

Figure 5.
Delimitation of a boundary by a novelty detection model.

3.4 Transfer learning

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

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

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

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

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

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

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

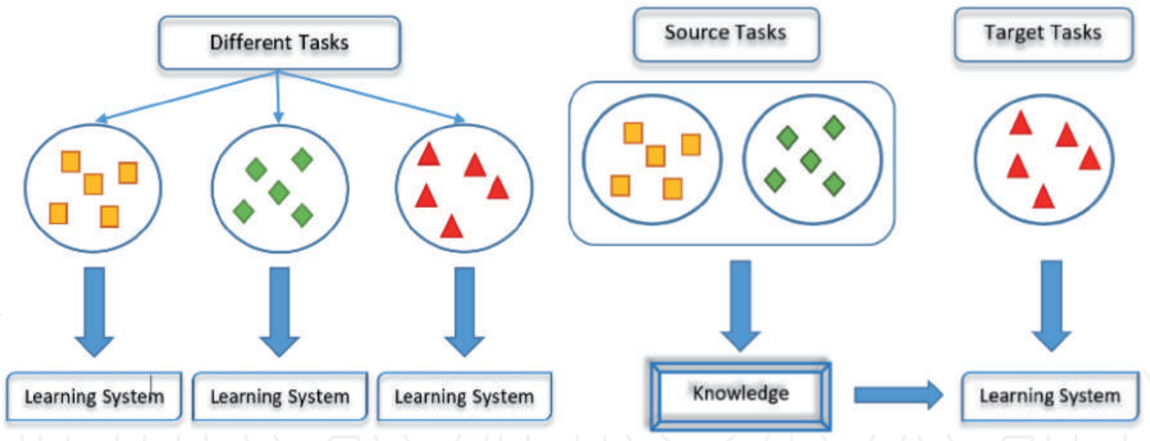

Figure 6.
Multi-task learning framework process and transfer learning process.

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

4. Experimental case of deep learning in CBM

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

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

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

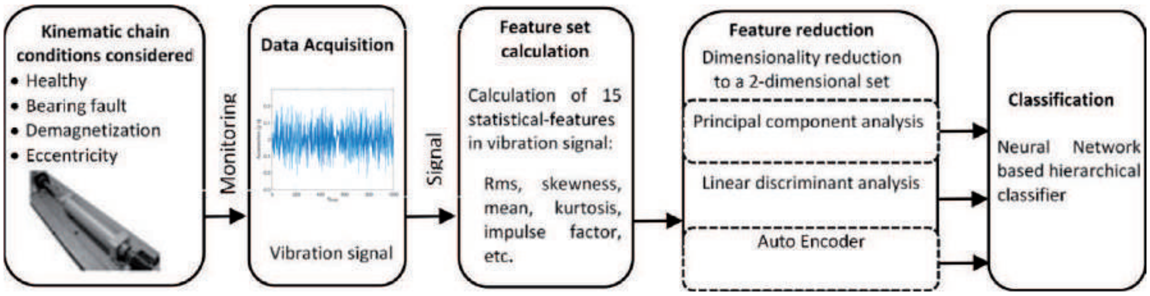

Figure 7.
Proposed scheme for multiple faults diagnosis.

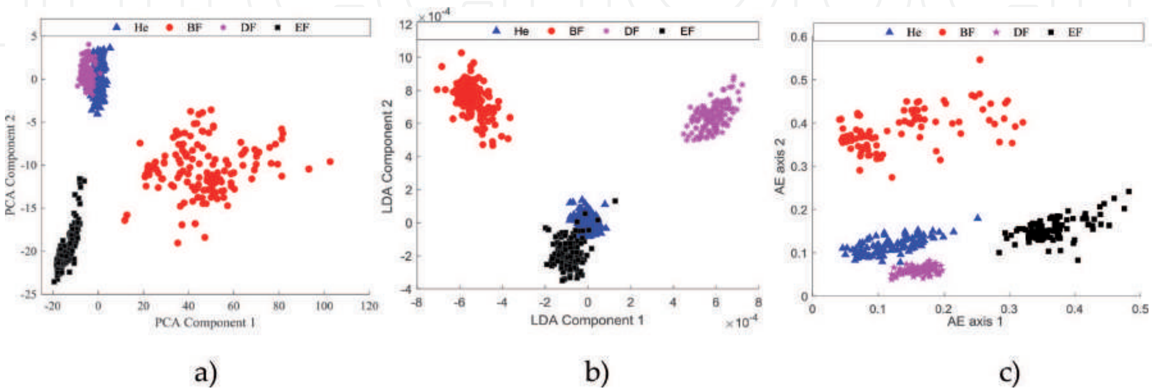

Figure 8.
Evaluation of resulting projections into a two-dimensional space. (a) PCA. (b) LDA. (c) Auto-encoder.

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

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

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

5. Conclusion and future challenges

In this chapter, a review of some of the current techniques based on deep learning and some of the functionalities that they may have within the environment of the diagnostic schemes of electromechanical systems is carried out. Having as reference the high complexity that is increasingly being found in the manufacturing processes, and the new challenges to face in the Industry 4.0 paradigm, it is necessary to improve the diagnostic capabilities of traditional schemes, which is why methodologies based on artificial intelligence and deep learning methods have increasingly called the attention of researchers. However, it remains to be discovered and identified the patterns that these deep neural networks learn, and specifically, within the industry environment, and electromechanical systems, what is the scope and benefits of applying these novel techniques.

Author details

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

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

References

[1] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain, Cornell Aeronautical Laboratory. Psychological Review. 1958;**65**(6):386-408

[2] Stuart G, Spruston N, Sakmann B, Häusser M. Action potential initiation and backpropagation in neurons of the mammalian CNS. Trends in Neurosciences. 1997;**20**(3):125-131. DOI: 10.1016/s0166-2236(96)10075-8

[3] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems. 2012;**25**(2):84-90. DOI: 10.1145/3065386

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

[5] Sun J, Yan C, Wen J. Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. IEEE Transactions on Instrumentation and Measurement. 2018;**67**(1):185-195

[6] Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996;**381**:607-609

[7] Ma X, Wang X. Convergence analysis of contrastive divergence algorithm based on gradient method with errors. Mathematical Problems in Engineering. 2015: Article ID 350102, 9 p

[8] Liu R, Yang Y, Zhao Z, Zhou J. A novel scheme for fault detection using data-driven gap metric technique. In: IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS); Enshi. 2018. pp. 1207-1212

[9] Zheng J, Cao X, Zhang B, Zhen X, Su X. Deep ensemble machine for video classification. IEEE Transactions on Neural Networks and Learning Systems. 2019;**30**(2):553-565

[10] Yuan Y, Mou L, Lu X. Scene recognition by manifold regularized deep learning architecture. IEEE Transactions on Neural Networks and Learning Systems. 2015;**26**(10):2222-2233

[11] Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Transactions on Medical Imaging. 2016;**35**(5):1313-1321

[12] Thirukovalluru R, Dixit S, Sevakula RK, Verma NK, Salour A. Generating feature sets for fault diagnosis using denoising stacked auto-encoder. In: 2016 IEEE Inter. Conf. on Progn. & Hlth. Manag. Ottawa, ON: ICPHM; 2016. pp. 1-7

[13] Li C, Zhang W, Peng G, Liu S. Bearing fault diagnosis using fully-connected winner-take-all autoencoder. IEEE Access. 2018;**6**:6103-6115

[14] Ding X, He Q. Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement. 2017;**66**(8):1926-1935

[15] Abolhasanzadeh B. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features. In: 2015 7th Conference on Information and Knowledge Technology. Urmia: IKT; 2015. pp. 1-5

[16] Zhang J, Chen Z, Du X, Xu X, Yu M. Application of stack marginalised sparse denoising auto-encoder in fault diagnosis of rolling bearing.

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

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

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

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

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

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

Chapter

Smart Monitoring Based on Novelty Detection and Artificial Intelligence Applied to the Condition Assessment of Rotating Machinery in the Industry 4.0

Saucedo-Dorantes Juan Jose, Jaen-Cuellar Arturo Yosimar and Elvira-Ortiz David Alejandro

Abstract

The application of condition monitoring strategies for detecting and assessing unexpected events during the operation of rotating machines is still nowadays the most important equipment used in industrial processes; thus, their appropriate working condition must be ensured, aiming to avoid unexpected breakdowns that could represent important economical loses. In this regard, smart monitoring approaches are currently playing an important role for the condition assessment of industrial machinery. Hence, in this work an application is presented based on a novelty detection approach and artificial intelligence techniques for monitoring and assessing the working condition of gearbox-based machinery used in processes of the Industry 4.0. The main contribution of this work lies in modeling the normal working condition of such gearbox-based industrial process and then identifying the occurrence of faulty conditions under a novelty detection framework.

Keywords: smart monitoring, condition assessment, novelty detection, artificial intelligence, Industry 4.0, rotating machinery

1. Introduction

Nowadays industrial applications are straightly involved with intelligent manufacturing processes, and the importance of this issue is reflected in many different activities of the human being, for example, in health, economy, and even comfort. Thus, it is possible to say that most of the daily activities carried out by humans have a direct relationship with those elements produced in the industry that facilitate its making. On the other hand, during the last years, industrial sites have been continuously subjected to several transformations, aiming to improve the effectiveness of its processes and to increase the production quality. Consequently, the integration of multiple technologies in the industry has been performed by the composition of actuators and sensors with cyber-physical systems and Internet of Things devices. Such integration leads to the Industry 4.0 that is the fourth phase of manufacturing and industrial sectors where the automated manufacturing and process monitoring have been enhanced [1]. Consequently, under the integration of such complex systems, it should be highlighted that it is important to ensure its safety and reliability by the implementation of condition-based monitoring approaches. Thereby, in order to guarantee the proper operation in manufacturing processes and aiming to avoid undesirable downtimes, the working condition of the machine components must be continuously assessed. Commonly, most of the industrial applications and processes are involved with the use of mechanical and electrical rotating machines, where electric motors and gearboxes represent the most used elements to perform specific manufacturing processes [2].

In fact, this statement is validated and justified because electric motors, gearboxes, couplings, and shafts represent approximately more than 90% of the elements that compose any industrial process [3]. Indeed, these elements that integrate the main operating system of industrial machinery are also considered, and also known, as the electromechanical machine system. In this sense, electric motors are considered as the most important element in electromechanical systems since its performed functions cannot be carried out and replaced by any other element; additionally, these elements play also an important role in most of the industrial applications because two-thirds of the total electricity is consumed by them. Therefore, these issues make suitable the application of condition monitoring approaches to avoid the occurrence of unexpected breakdowns; even more, it must be noted that under the appearance of a faulty condition, such damaged element can also have influence over the proper operation of the whole elements that are linked to the electromechanical system and crucial damages may be produced [4].

As it has been mentioned, industrial sites have been subjected to several transformations, and through the integration of multiple technologies, a significant improvement in the production efficiency has been obtained. Accordingly, complex electromechanical systems compose most of the industrial machinery that is used in different applications of modern industry. In this regard, several condition monitoring-based approaches have been developed aiming to guarantee the appropriate working condition of industrial machinery. Thus, data-driven condition monitoring strategies represent the most common and suitable approach for carrying out the condition assessment in electromechanical systems; this approach has been preferred since it only takes into account the use of information of available data; therefore, based on known and available information, an accurate diagnosis of the machine under inspection is obtained [2, 4, 5].

In this sense, most of the data-driven approaches mainly include the continuous monitoring of different physical magnitudes that contain significant information related to the machine working condition. Indeed, stator current signals, vibrations, temperatures, and operational rotating speeds, among others, are some of the most accepted and reliable magnitudes used in condition monitoring strategies. On the other hand, aiming to provide the condition assessment, such monitored signals are then analyzed by different signal processing techniques, where time-domain analysis, frequency-domain analysis, and time-frequency domain analysis have been commonly implemented in several condition monitoring strategies [5]. However, although there exist different signal processing, it has been demonstrated that statistical time-based domain features contain significant information that describes the behavior related to the rotating machine working condition. Thereby, the calculation of a high-performance set of features is achieved because statistical time domain-based features have advantages for describing changes and trends of time-domain signals [6].

On the other side, although other sophisticated techniques such as fast Fourier transform (FFT) and discrete wavelet transforms (DWT), among others, also lead to the calculation of features related to the machine condition, the implementation of such techniques considers additional knowledge and experience about the proper usage of the techniques and also complete information of the parameters of the machine operation. Accordingly, it should be highlighted that it is not totally true that sophisticated and complex signal processing may always lead to the estimation of the most representative set of features to describe the machine condition. In this regard, from a practical application viewpoint and based on practical experience, the simplest way to evaluate and identify the early occurrence of faults is by means of analyzing trends of physical magnitudes acquired during the continuous working operation of the machine. Thus, as aforementioned, the appropriate early detection of faults may help in the reduction of monetary losses caused by unscheduled maintenance task.

Certainly, the detection and identification of faulty operating modes involve a critical procedure in which the signal processing or feature calculation must be carefully performed. Another important issue to perform and improve the condition assessment is the consideration of artificial intelligence (AI) for carrying out the automatic fault diagnosis. Indeed, the use of AI in condition monitoring strategies has been rapidly increased, and its application to identify the occurrence of faults in rotating machinery is an adequate and coherent option to obtain high-performance results. Additionally, it has been shown that an appropriated application of AI in condition monitoring approaches provides a powerful capacity for detecting and classifying the appearance of single or multiple faults in electromechanical systems. This potential provided by AI is reached because the limitations of classical space-transform techniques, when nonlinearities characterize the analyzed system, are overcome [6].

Hence, several AI techniques have been addressed with the main purpose of being applied in monitoring tasks of industrial machinery, for instance, artificial neural networks, genetic algorithms, fuzzy logic, support vector machines, Bayesian networks, self-organizing maps (SOM), and case-based reasoning, among others, represent some of the most techniques used in condition monitoring approaches [7]. However, there are still great challenges for developing new condition monitoring strategies; indeed, the use of AI techniques has increased because the main challenge of the condition assessment in industrial sites is that nonlinearities are inherent to the working operation [8].

Thereby, the contribution of this chapter lies in the proposal of a condition monitoring strategy for detecting and assessing unexpected working conditions in rotating machines. Such proposal performs the condition assessment under a novelty detection approach based on self-organizing maps. Thus, the proposed condition monitoring method includes the estimation of a statistical time-based set of features from acquired vibration signals; then, the data modeling is carried out through SOM and then the evaluation of novelty detection events. Finally, if novelties are detected, a retraining and incremental learning procedure is considered by including a dimensionality reduction stage by means of the linear discriminant analysis. This proposal is validated and applied to a real laboratory gearbox-based electromechanical system.

2. Fault detection and identification

The condition monitoring assessment is involved with the behavior analysis of the machine working operation; thus, the consideration of stator currents or vibrations as informative physical magnitudes for condition monitoring represents the most preferred and accepted approaches in the related literature. Also, although different information fusion levels are considered, such as signal-level or decision-level, dealing with electromechanical condition monitoring, the feature-level represents the most appropriate, since many numerical fault indicators from aforementioned physical magnitudes have been proposed as suitable fault indexes in multiple studies [9, 10]. In this regard, time-domain, frequency domain, and time-frequency domain are the three feature estimation approaches widely applied during the physical magnitude characterization process. Although techniques based on frequency and time-frequency domain, such as classical Fourier transform or wavelet analysis have been widely applied, most of these techniques require a deep knowledge of the fault effects over the resulting frequency distributions of the physical magnitudes. Indeed, as stated by Zhang et al. in [11], dealing with complex electromechanical systems, where the resulting interaction among multiple parts is reflected in the acquired physical magnitudes, the consideration of statistical time-domain features represents a performing trade-off between computational simplicity and characterization capabilities of general patterns. Such feature-level fusion scheme needs to consider the processing of a high-dimensional set of numerical features estimated during the characterization of the available physical magnitudes that, although increases the fault detection and identification capabilities, inevitably contain redundant and nonsignificant information.

Dimensionality reduction procedures are applied in order to avoid low fault diagnosis performances and overfitting responses of the condition monitoring schemes. In this regard, classical dimensionality reduction techniques have been widely applied, as the principal component analysis (PCA). However, PCA aims to identify orthogonal components that maximize the preservation of the data variance. That is, PCA seeks for global data representation; thus, considering the unsupervised operation, a set of non-connected data clusters have a negative impact over the resulting representation. Other classical approaches, as linear discriminant analysis, overcome such data topology limitation by means of a supervised approach, as the LDA, where the resulting set of features is a mathematical combination of the original ones maximizing distances among classes [12, 13].

Finally, the classification algorithms, play an important role in data-driven condition monitoring schemes to perform the automatic and final diagnosis outcome. In this regard, neural networks and fuzzy inferred systems classically represent the most used classifiers, but also classifiers like decision trees and support vector machines have been widely applied [14–16]. The use of these techniques, however, is related with the maximization of the classification ratio by means of the feature set decomposition following supervised training schemes. According to Shannon's rate-distortion theory, mutual dependencies among various sources and between the input and output spaces contain the actual intrinsic dimension of the data and allow avoiding over-fitted responses. Thus, unsupervised learning approaches applied over the available feature space represent the most coherent processing procedure in order to maintain the underlying physical phenomenon of the system under monitoring. Concerning this problem, manifold learning methods have been applied in the last years to preserve the information in a lower dimensional space. Among them, the self-organizing map, SOM, is the most used, which is based on developing a neural network grid to preserve most of the original distances between feature vector representations in the original feature space [17]. Indeed, the SOM allows a high-dimensional input data mapping over a two-dimensional output layer while preserving as much as possible the structure of the input data. Although SOM leads to model the original data distribution following an unsupervised approach, each of the neuron units used during the original space characterization can be later associated with a class label; thus, through distance criteria, the diagnosis can be estimated during the assessment of a new measurement. Thus, both fault detection and identification tasks can be faced at the same time and, what is more important, considering the same criteria for both outcomes, that is, topological aspects of the data distribution.

3. Novelty detection and diagnosis methodology

The proposed condition monitoring strategy that is applied to the condition assessment of an electromechanical system under a novelty detection framework is composed of five important stages as depicted in **Figure 1**.

The first stage is based on the fact that initially the machine condition is known; in this sense, it is considered as an initial condition that only the available information belongs specifically to the behavior of the healthy condition of the electromechanical system under evaluation. This assumption is asserted and taken into account since all the machinery used in most of the industrial applications starts its life cycle from an initial healthy condition, which means that all elements work properly. Therefore, under this assumption, such available information is obtained from the continuous monitoring of one vibration signal that is monitored during the working operation of the electromechanical system.

In the second stage, the characterization of the machinery behavior is performed; thus, the available vibration signal is processed and analyzed, aiming to carry out a characterization of the machine working condition and also with the aim of highlighting those representative features that can represent the occurrence of abnormal operations. Precisely, the calculation of a representative set of eight statistical time-based domain features is estimated from the acquired vibration signal; this proposed set of features consists of some well-known statistical features such as mean, rms, standard deviation, variance, shape factor, crest factor, skewness, and kurtosis. Indeed, and as it has been mentioned, statistical time-based domain features provide meaningful information leading to the estimation of high-performance feature characterization due to its capability of describing trends and changes in signals; additionally, this proposed set of statistical features has been included in several condition monitoring approaches to perform the assessment of the operating working condition of electromechanical systems used in industrial application [3, 11, 16]. The corresponding mathematical equations of such numerical features are shown in **Table 1**.

Subsequently, in the third stage, the set of statistical features estimated from vibrations is modeled through SOM; the data modeling is performed by SOMs since

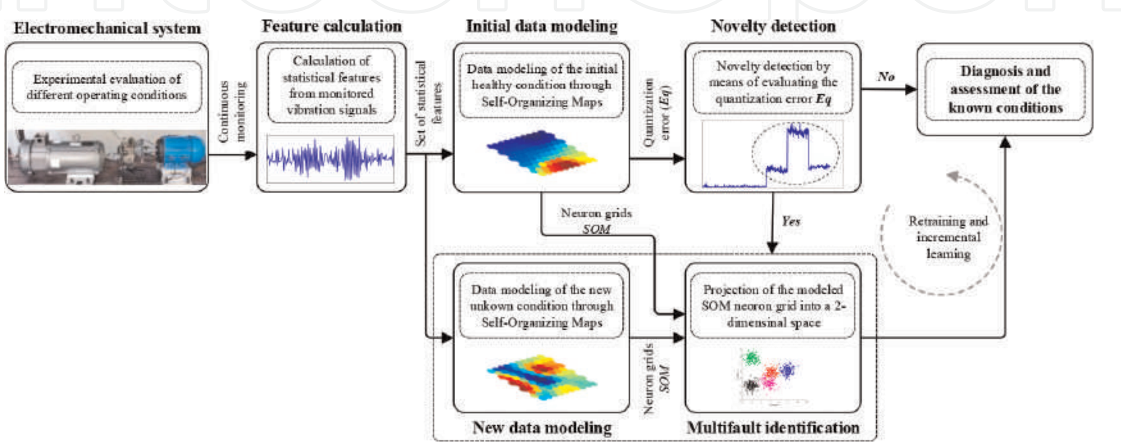

Figure 1.
Rotating machinery-based electromechanical system used to demonstrate the practical implementation of the proposed method.

Mean	$\overline{x} = \frac{1}{n} \cdot \sum_{k=1}^{n} \lvert x_k \rvert$	(1)
Root mean square	$RMS = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} (x_k)^2}$	(2)
Standard deviation	$\sigma = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2}$	(3)
Variance	$\sigma^2 = \frac{1}{n} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2$	(4)
Shape factor	$SF_{RMS} = \frac{RMS}{\frac{1}{n} \cdot \sum_{k=1}^{n} \lvert x_k \rvert}$	(5)
Crest factor	$CF = \frac{\hat{x}}{RMS}$	(6)
Skewness	$S_k = \frac{E\left[(x_k - \overline{x})^3\right]}{\sigma^3}$	(7)
Kurtosis	$k = \frac{E\left[(x_k - \overline{x})^4\right]}{\sigma^4}$	(8)

Table 1.
Set of statistical time-domain features.

this approach allows to preserve the data topology. Due to the proposed condition, monitoring strategy is based under a novelty detection framework, and the initial and available information is modeled, aiming to represent the initial known condition which is the healthy condition. As a result, a pre-defined neuron SOM grid model is first obtained to characterize the healthy condition of the electromechanical system. Then, in case additional conditions appear, the data modeling is also performed by a specific neuron SOM grid model for each one of the additional operating condition.

Afterward, the novelty detection is performed in the fourth stage; in this sense, there exist different approaches for carrying out the detection of novel events. Classic novelty detection approaches are based on the evaluation of numerical threshold values, and the definition of such values depends on different criteria. Thereby, for this proposal, the novelty detection is performed by evaluating the average quantization error, *Eq*, obtained during the data modeling through SOMs; indeed, the novelty detection based on the *Eq* is a coherent option according to the data modeling to detect whether the electromechanical system condition is known or unknown. Certainly, because the healthy condition is initially the unique known and available condition, the evaluation of any other new measurement that does not belong to the known condition will exhibit a different *Eq* value. Thus, any change presented in the *Eq* value should be analyzed because this value is an important measurement related to the occurrence of unexpected and unknown events which results in the novelty detection. Otherwise, the diagnosis and condition assessment of the known conditions is carried out if any change is presented in the *Eq* value.

Finally, the last stage is carried out in case of novelty detection; thus, this stage considers a retraining process where an incremental learning is performed with the aim of updating the available information with new data that belongs to new operating conditions. In this sense, during the detection of a novelty event, the available information that describes such novel condition is also processed, and from the acquired vibration signal, the statistical time-domain features are also estimated. Then, such new available information represented by the estimated statistical features is modeled through SOMs, and a new neuron SOM grid represents the new condition. Accordingly, as aforementioned, each new operating condition detected under this novelty detection approach has to be modeled by a specific neuron SOM model. Finally, when novelty detection occurs, such neuron SOM grids are subjected to a dimensionality reduction procedure by means of the linear discriminant analysis in order to obtain a maximum linear separation between the considered conditions and also with the aim of obtaining a visual representation the assessed conditions.

4. Case study

In order to demonstrate the practical implementation of the proposed smart monitoring based on novelty detection in an industrial application, a case of study is proposed next.

A rotating machinery-based electromechanical system has been considered; such electromechanical system includes a three-phase IM of 1492-W (model WEG00236ET3E145T-W), a gearbox with 4:1 ratio (model BALDOR GCF4X01AA), and a DC used as a mechanical load (model BALDOR CDP3604). The IM is coupled shaft to shaft to the gearbox, and the gearbox is also coupled shaft to shaft to the DC generator, and a VFD (model WEGCFW08) is also used to feed and control the different operating frequencies of the IM. Besides, the DC generator is used as a non-controlled mechanical load representing around 20% of the nominal load. A picture of the second electromechanical system based on a gearbox is shown in **Figure 2**.

Aiming to detect and assess the appearance of unexpected conditions, a database of different experiments is generated. The data acquisition is carried out by means of a data acquisition system (DAS) that is a proprietary low-cost design based on a field programmable gate array; such DAS uses two 12-bit 4-channel serial-output sampling analog-to-digital converters, model ADS7841 from Texas Instruments. Different physical magnitudes have been acquired during the experiments; that is,

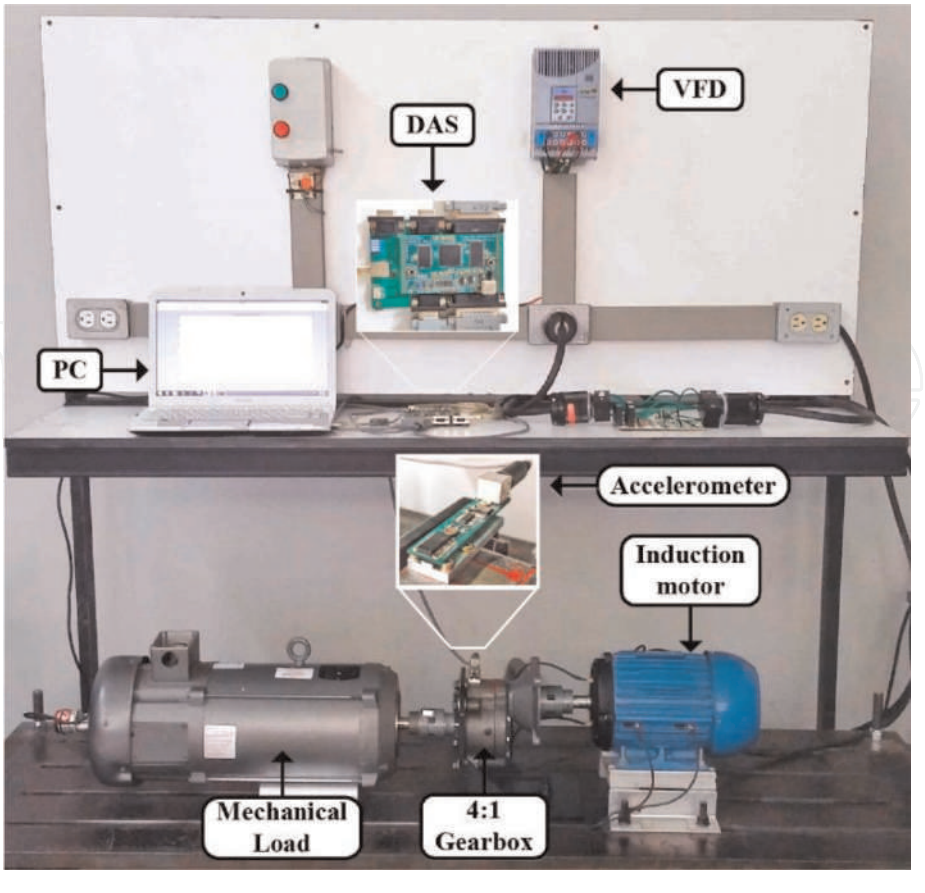

Figure 2.
Rotating machinery-based electromechanical system used to demonstrate the practical implementation of the proposed method.

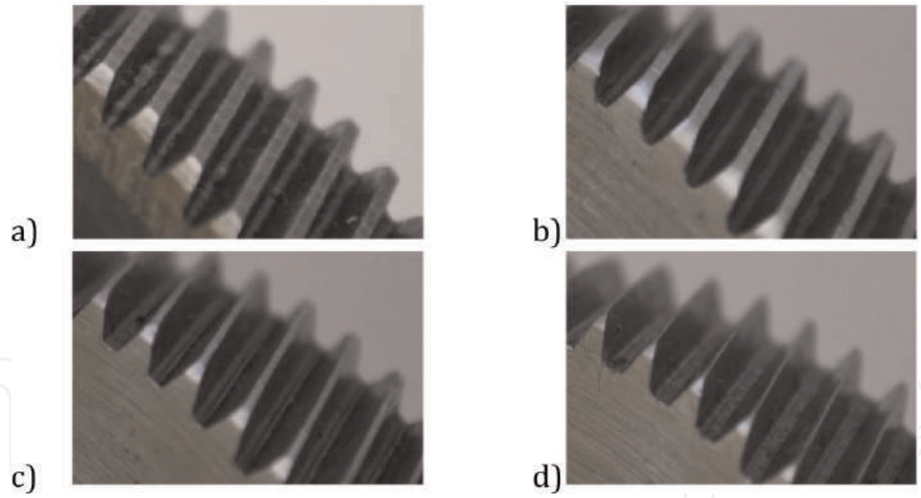

Figure 3.
Set of the faulty conditions evaluated in the gearbox-based electromechanical system: (a) healthy gear, (b) 25% of uniform wear, (c) 50% of uniform wear, and (d) 75% of uniform wear.

the appearance of mechanical vibrations is acquired by means of a triaxial accelerometer (LIS3L02AS4). In this regard, the accelerometer sensor is fixed on the top of the gearbox. For this proposed work, the occurrence of vibrations is analyzed because they are inherent to the rotating condition of the rotating elements that compose the electromechanical system, i.e., electric motors, gearboxes, and bearings, among others [2].

The accelerometer sensor is individually mounted on a board with its corresponding signal conditioning and anti-alias filtering. During the acquisition of vibration signals, the sampling frequency is set to 3 kHz; as a result, 270 kS are stored during 90 s of continuous sampling of the working condition, in the steady-state regime, of the electromechanical system are stored. Furthermore, the IM of the experimental test bench is driven at different operating frequencies during the experimentations; specifically, the operating frequencies are set at 5, 15, and 50 Hz.

During the experimentation, four different operating conditions are also evaluated: healthy (HLT), 25% of uniform wear in the gearbox (W25), 50% of uniform wear in the gearbox (W50), and 75% of uniform wear in the gearbox (W75). In this regard, the gearbox with 4:1 ratio is composed of two gears, the driver gear and the driven gear which has 18 and 72 teeth, respectively. The wear was artificially induced uniformly in all teeth of three similar driven gears: from **Figure 3a–d**, the set of gears tested in the gearbox-based electromechanical system. The experiments are performed by replacing iteratively the healthy gear with the damaged ones.

5. Competency of the method/results

The proposed condition monitoring strategy is based on a novelty detection approach; the implementation of such proposal has been done in Matlab that is a sophisticated software used in several engineering applications. Indeed, the use of Matlab facilitates the signal processing for carrying out the condition assessment of the electromechanical system. Thus, the available vibration signal is first continuously monitored and acquired during the operating condition of the electromechanical systems, and then the statistical set of features is estimated from the vibration signal.

As aforementioned, the initial condition belongs to the healthy condition; in this sense, the data modeling is carried out aiming to obtain a neuron SOM grid model that represents such initial condition. As a result of the data modeling, the first

SOM_1 model obtained and this SOM model only characterize the healthy condition of the electromechanical system. During the data modeling, an average Eq error of 0.4932 has been obtained during the training procedure, and during the evaluation the Eq error reaches a value of 19.4419. It should be noted that during the evaluation different data information has been used; indeed, the evaluated data belong to a faulty condition tested in the gearbox. In **Figure 4**, a visual representation of the novelty detection achieved by the first modeled neuron SOM_1 grid is shown.

After the first novelty detection, the process and incremental learning is carried out; in this regard, the available data that belong to the first faulty condition (25% of uniform wear) is modeled by a second neuron SOM_2 grid. Thus, the data information related to the working condition of the machine consist of two known conditions which are healthy and 25% of uniform wear. Indeed, during the training of the second SOM model, a Eq of 0.8997 is achieved during the training and during the evaluation with available data, which belongs to another unknown condition; the Eq error was 7.0773; thus, such significant increase in the Eq error depicts that an anormal condition is detected by the novelty detection approach. The visual representation of the Eq error is shown in **Figure 5** where it is possible to appreciate the abrupt change due to the occurrence of the unexpected faulty condition.

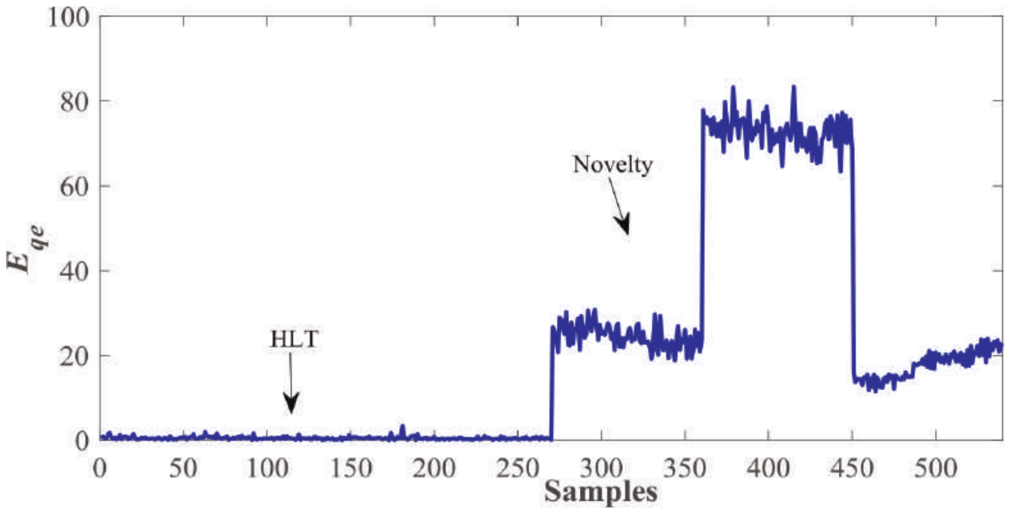

Figure 4.
Novelty detection performed by SOM_1 during the evaluation of the first faulty condition tested in the electromechanical system, 25% of uniform wear in the gearbox.

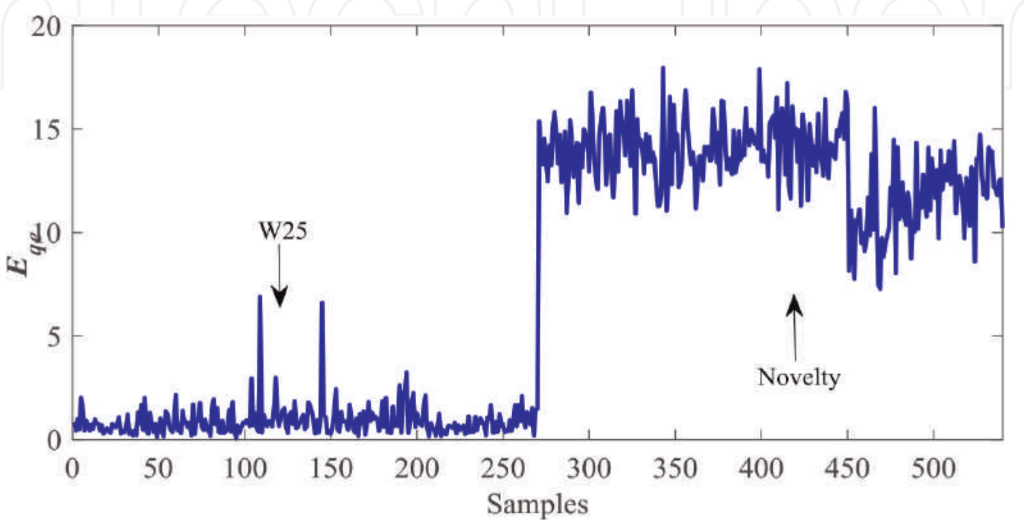

Figure 5.
Novelty detection performed by SOM_2 during the assessment of the second faulty condition, 50% of uniform wear in the gearbox.

The data information to the electromechanical system condition is currently composed of three different conditions, healthy, 25%, and 50% of uniform wear in the gearbox. Later, available information related to another faulty unknown condition is evaluated after performing the retraining process and incremental learning. In this regard, during the training procedure of the third neuron SOM_3 grid, the obtained Eq value was around 0.7077, and during the evaluation of the last faulty condition the achieved Eq was around 6.4367. Thus, the SOM_3 model represents the available information to the third faulty condition that is 50% of uniform wear. In **Figure 6**, the visual representation of the novelty detection performed is shown during the evaluation of the SOM_3 model.

Subsequently, after the last retraining and incremental learning, the available information related to the faulty condition of 75% of uniform wear is also modeled

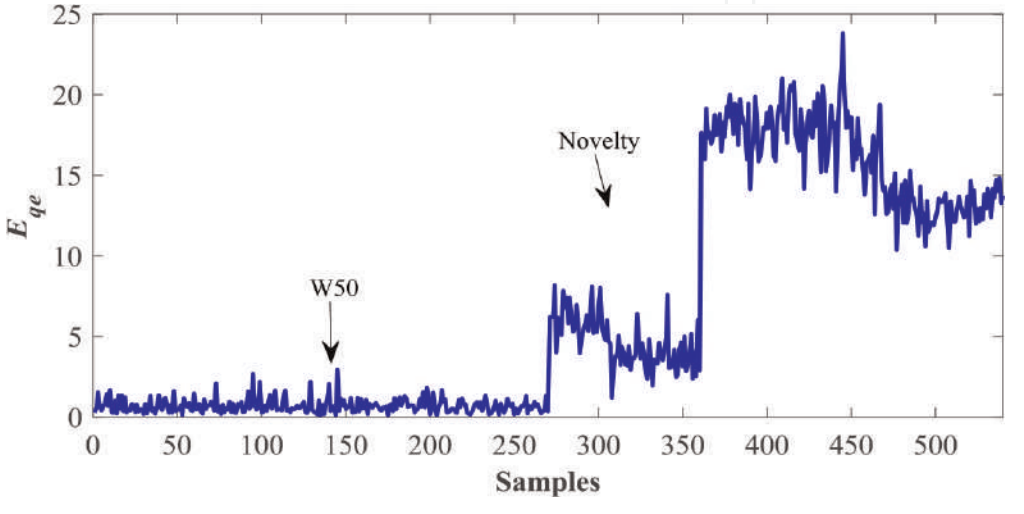

Figure 6.
Novelty detection carried out by SOM_3 obtained for the evaluation of the third faulty condition, 75% of uniform wear in the gearbox.

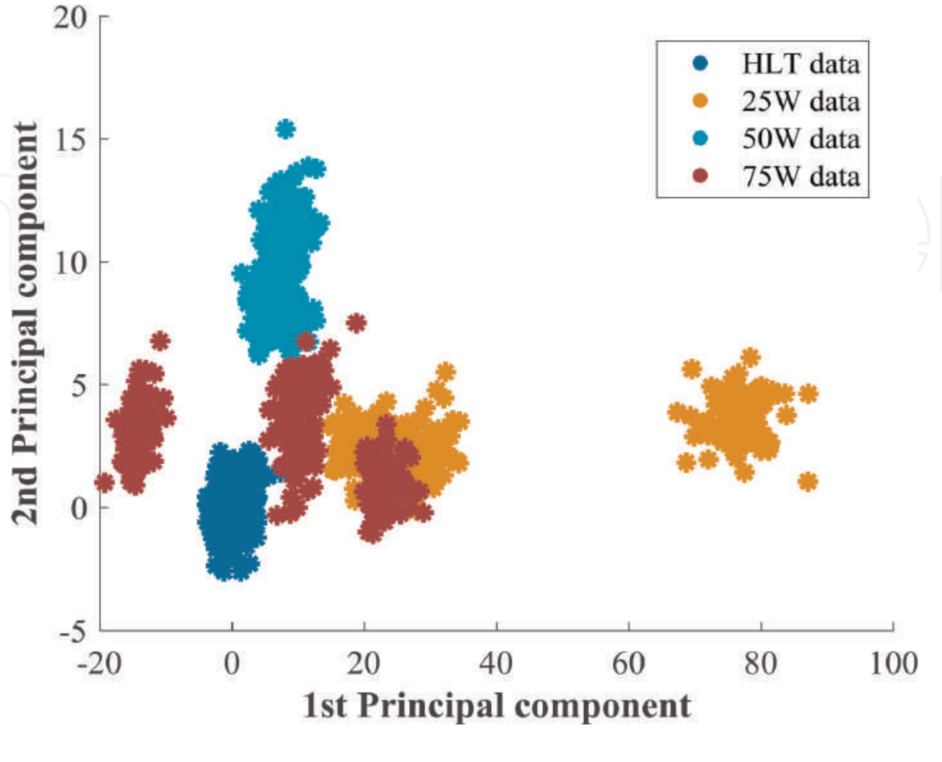

Figure 7.
Resulting two-dimensional projection obtained by considering the four neuron SOM grids modeled for each one of the detected conditions.

by a fourth SOM model, such model is the neuron SOM_4 model, and the E_q error achieved during the training was 0.7700. Because four different operating conditions are detected during the operating condition of the electromechanical system, the final available information stored by the proposed novelty detection approach consist of information capable of detecting four different operating conditions. In case of more novelty detections, the retraining process and incremental learning are again performed, and the information related to the different operating conditions is updated.

Finally, a visual representation of the operating conditions detected during the application of the proposed diagnosis methodology is obtained by means of applying a dimensionality reduction technique, PCA. In this sense, in **Figure 7**, a visual representation of the data distribution of all detected conditions is shown; in this visual representation, it is appreciated that different operating conditions appears. Indeed, different clusters appear for each detected condition because different operating frequencies were considered during the experimental evaluation of the considered conditions.

6. Conclusions

Modern industrial production is characterized by the consideration of machine learning data-based models to support the main aspects of the manufacturing process. In this regard, two main data science challenges related with condition monitoring of electromechanical assets in the Industry 4.0 framework are (i) the premise that only information of the healthy condition is initially available and (ii) the adaptation of the fault detection and identification scheme in order to incorporate new operating conditions. Thus, this paper proposes a new methodology for multi-fault detection and identification based on incremental learning applied to novel fault detection on electromechanical systems by analyzing vibrations and stator current signatures of the electric motor drive.

Moreover, the proposed condition monitoring strategy based on a novelty detection approach is capable of being applied to other electromechanical systems, and also the consideration of other different physical magnitudes can be also included in such proposal.

Acknowledgements

This research work has been partially supported by the FOFIUAQ2018 under the registered project FIN201811.

Conflict of interest

The authors declare no conflict of interest.

Author details

Saucedo-Dorantes Juan Jose*, Jaen-Cuellar Arturo Yosimar
and Elvira-Ortiz David Alejandro
HSPdigital CA-Mecatronica, Engineering Faculty, Autonomous University of Queretaro, Mexico

*Address all correspondence to: jsaucedo@hspdigital.org

References

[1] Frank AG, Dalenogare LS, Ayala NF. Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics. 2019; **210**:15-26

[2] Contreras-Hernandez JL, Almanza-Ojeda DL, Ledesma S, Ibarra-Manzano MA. Motor fault detection using quaternion signal analysis on FPGA. Measurement. 2019;**138**:416-424

[3] Saucedo-Dorantes JJ, Osornio-Rios RA, Delgado-Prieto M, Romero-Troncoso RJ. Diagnosis methodology based on statistical-time features and linear discriminant analysis applied to induction motors. IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED). 2017;**2017**:517-523. DOI: 10.1109/DEMPED.2017.8062404

[4] Ribeiro DA, Ferreira DD, Pereira DA, Braga Junior RA, Nunes RD. Mechanical fault detection in electric motors measured by a digital signal processing device in an optical mouse. Measurement. 2019;**138**:350-355

[5] Frini M, Soualhi A, El Badaoui M. Gear faults diagnosis based on the geometric indicators of electrical signals in three-phase induction motors. Mechanism and Machine Theory. 2019; **138**:1-15

[6] Zhao Y, Li T, Zhang X, Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. Renewable and Sustainable Energy Reviews. 2019; **109**:85-101. DOI: 10.1016/j.rser. 2019.04.021

[7] Nasiri S, Khosravani MR, Weinberg K. Fracture mechanics and mechanical fault detection by artificial intelligence methods: A review. Engineering Failure Analysis. 2017;**81**: 270-293. DOI: 10.1016/j.engfailanal. 2017.07.011

[8] Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing. 2018;**108**:33-47. DOI: 10.1016/j.ymssp.2018.02.016

[9] Kang Md. M, Islam R, Kim J, Kim JM, Pecht M. Hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. IEEE Transactions on Industrial Electronics. 2016;**63**(5):3299-3310

[10] Wang S, Yub J, Lapirac E, Lee J. A modified support vector data description based novelty detection approach for machinery components. Applied Soft Computing. 2013;**13**: 1193-1205. DOI: 10.1016/j.asoc. 2012.11.005

[11] Zhang X, Chei W, Wang B, Chen X. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. Neurocomputing. 2015;**167**:260-279

[12] He Q, Ding X, Pan Y. Machine fault classification based on local discriminant bases and locality preserving projections. Mathematical Problems in Engineering. 2014;**2014**: 1-12. Article ID: 923424

[13] Prieto-Moreno A, Llantes-Santiago o, Garcia-Moreno E. Principal components selection for dimensionality reduction using discriminant information applied to fault diagnosis. Journal of Process Control. 2015;**33**:12-24

[14] Cunha-Palacios RH, Goedtel A, Godoy WF, Fabri JA. Fault identification in the stator winding of induction motors using PCA with artificial neural networks. Journal of Control, Automation and Electrical Systems. 2016;**27**(4):406-418

[15] Manimala K, Getzy David DI, Selvi K. A novel data selection technique using fuzzy c-means clustering to enhance SVM-based power quality classification. Soft Computing. 2015; **19**(5):3123-3144

[16] Sadooghi MS, Khadem SE. Improving one class support vector machine novelty detection scheme using nonlinear features. Pattern Recognition. 2018;**83**:14-33

[17] Li W, Zhang S, He G. Semisupervised distance-preserving self-organizing map for machine-defect detection and classification. IEEE Transactions on Instrumentation and Measurement. 2013;**62**(5):869-879

Chapter

AI for Improving the Overall Equipment Efficiency in Manufacturing Industry

Francesc Bonada, Lluís Echeverria, Xavier Domingo and Gabriel Anzaldi

Abstract

Industry 4.0 has emerged as the perfect scenario for boosting the application of novel artificial intelligence (AI) and machine learning (ML) solutions to industrial process monitoring and optimization. One of the key elements on this new industrial revolution is the hatching of massive process monitoring data, enabled by the cyber-physical systems (CPS) distributed along the manufacturing processes, the proliferation of hybrid Internet of Things (IoT) architectures supported by polyglot data repositories, and big (small) data analytics capabilities. Industry 4.0 paradigm is data-driven, where the smart exploitation of data is providing a large set of competitive advantages impacting productivity, quality, and efficiency key performance indicators (KPIs). Overall equipment efficiency (OEE) has emerged as the target KPI for most manufacturing industries due to the fact that considers three key indicators: availability, quality, and performance. This chapter describes how different AI and ML solutions can enable a big step forward in industrial process control, focusing on OEE impact illustrated by means of real use cases and research project results.

Keywords: machine learning, supervised learning, unsupervised learning, classification, regression, ensembles, artificial intelligence, data mining, data-driven, industry 4.0, smart manufacturing, cyber-physical systems, predictive analytics

1. Introduction

Industry 4.0 has emerged as the perfect scenario for boosting the application of novel artificial intelligence (AI) and machine learning (ML) approaches to industrial process monitoring and optimization. Artificial intelligence is a set of techniques and methodologies aimed at allowing machines, especially computer systems, to simulate human intelligence processes. Machine learning is a subset of artificial intelligence, which provides a set of methodologies and strategies to allow systems for improvement. ML relies in automatic learning procedures, which generate knowledge from previous experiences (data).

One of the key elements on this new industrial revolution, aligned with the disruptive capabilities that AI and ML provide, is the hatching of massive process monitoring data, enabled by the cyber-physical systems (CPS) distributed along the manufacturing processes, the proliferation of hybrid IoT architectures supported by polyglot data repositories, and big (small) data analytics capabilities. Industry 4.0 paradigm is data-driven, and the smart exploitation of this data can provide a large set of competitive advantages impacting productivity, quality, and efficiency key performance indicators (KPIs), which are of utmost importance in the current competitive scenario. Moreover, the manufacturing companies are evolving to low volume with high personalization manufacturing environments [1, 2], where their competitiveness depends on the industries' facilities, considering asset and resource availability, but also in the optimal execution of production processes [3].

Therefore, there is an opportunity on improving the performance of manufacturing processes taking as input those new streams of information; going through analytical processes; creating new supporting models, tools, and services; and benchmarking their recommendations and outcomes against classical approaches. To that end, the overall equipment effectiveness (OEE) is aimed at measuring types of production losses and indicating areas of process improvement [4, 5], ideal to be used as a benchmarking KPI, and one of the main indicators used in manufacturing execution systems (MES) [6, 7].

In the recent years, research projects are aiming to develop novel stand-alone solutions covering the entire monitoring and control value chain: from the CPS for retrieving the data, to wireless communication protocols, big data storage for traceability and advanced artificial intelligence techniques for production control, optimization, and maintenance.

The use of artificial intelligence algorithms is enabling a big step forward in industrial process control and monitoring: from statistical process control (SPC) and statistical quality control (SQC) methodologies, which require a high prior knowledge of the process, to AI optimized process boundaries that provide valuable insights of the monitored process. Industrial applications of AI have its particular requirements. Not only prediction and forecasting capabilities are desired but also increasing the process knowledge with the right selection of AI algorithms, providing a competitive edge over traditional approaches.

AI provides the right set of tools for automatic quality prediction and full part traceability, process optimization, and preventive maintenance. These sets of benefits are directly impacting into productivity KPIs such as OEE and breakdowns, among others.

This chapter will describe the application of different AI and ML algorithms, including classifiers, regressors, or ensembles such as random forest trees, gradient boosting, or support vector machines, to some real-case industrial scenarios, such as quality prediction or process characterization for plastic injection molding or iron foundry, predictive maintenance for industrial water treatment processes, and means of leveraging production data (quality control, time series, batch data, etc.) at different granularity levels and its impact to OEE: from soft real-time to batch analysis and how this can be translated to valuable production insights.

2. Overall equipment effectiveness as KPI

As introduced before, the current scenario for manufacturing industries can be summarized as high demanding, very competitive, with dynamic market demand, and last but not least, hyperconnected and digital. Low-volume and more personalized parts or product work orders are replacing old high-volume ones without personalization, and this implies that effectiveness may not only focus on specific process optimization but also, for example, on improving changeover setup times, reducing scrap, or improving quality. Therefore, there is a clear need on improving and optimizing all manufacturing processes to overcome this demanding situation with effective response, also considering the efficient adaptation and usage of production lines. Traditional approaches tended to focus on throughput and utilization rate, but nowadays this is insufficient. The main reason relies on the importance of unconsidered context information, or even small details, which are making a difference.

The overall equipment effectiveness indicates how good the equipment is being used. OEE has emerged as the target KPI for most manufacturing industries due to the fact that considers three key indicators:

- Availability: Percentage of time that an equipment can operate
- Quality: Percentage of good produced parts
- Performance: Percentage of maximum operation speed used

But before going deep into OEE calculation, we must first understand in which phases of the manufacturing process AI can impact, so that we can relate all together. To that end, please refer to **Figure 1**, where OEE components are summarized, and **Figure 2**, where a standard manufacturing process is compared with an AI-powered one.

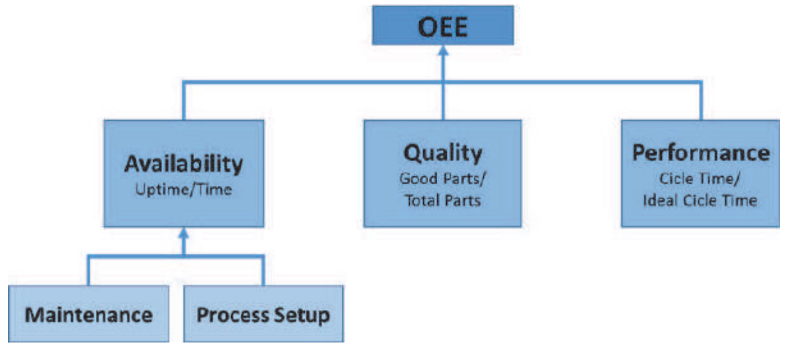

Figure 1.
OEE components and focus.

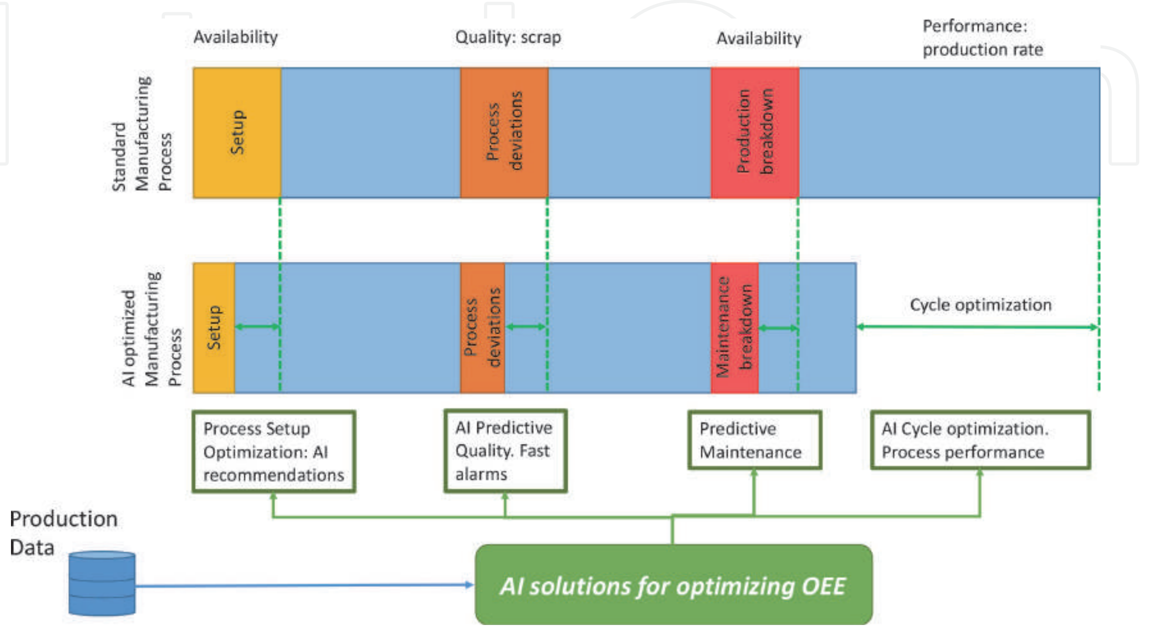

Figure 2.
OEE optimization using AI.

Focusing on **Figure 2**, let us introduce some simple examples of how AI impacts in the manufacturing process:

- Setup: We can improve the time needed to set up or adapt the environment, lines, and tools when a new incoming work order arrives, considering results from previous similar experiences. As we are able to do it in less time, and in a more effective way, we are impacting to the availability of the assets, and consequently, improving the OEE.

- Process deviations: In a similar way, AI allows for quality prediction relying on process parameters, which combined with real-time tuning of execution parameters, results in better quality outcomes, and scrap reduction, again, improving OEE.

- Maintenance: Predictive maintenance allows us to plan and provision with the needed spare parts so that impact in production is minimized. With this management we improve availability, and therefore, OEE is also improved.

In the text below, we define how the literature calculates the OEE, while in the following sections, we'll provide some real examples in which the OEE performance indicator has improved thanks to AI.

According to [8], the overall equipment effectiveness can be calculated as follows:

$$OEE = Availability * Performance\ rate * Quality\ rate. \tag{1}$$

where

- Availability

$$Availability = \frac{(available\ time - unplanned\ downtime)}{available\ time} \tag{2}$$

$$Availability\ time = \text{total available time–planned downtime} \tag{3}$$

Planned downtime: excess capacity, planned breaks, planned maintenance, communication break, and team meetings

- Unplanned downtime: breakdowns, setup and adjustment, late material delivery, operator availability

- Quality rate

$$Quality\ rate = \frac{(total\ produced\ parts - defective\ parts)}{total\ produced\ parts} \tag{4}$$

- Performance

$$Performance = \frac{(total\ production\ parts/operating\ time)}{idle\ run\ rate} \tag{5}$$

$$Operating\ time = Available\ time - unplanned\ downtime. \tag{6}$$

$$Idle\ run\ rate = number\ of\ parts\ per\ minute. \tag{7}$$

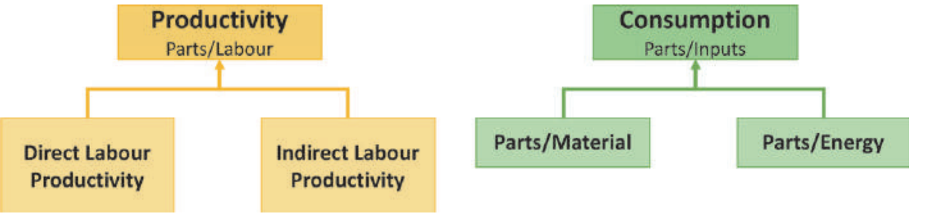

Figure 3.
Productivity indicators.

Other productivity indicators can also be very helpful when evaluating a manufacturing process and benchmarking how AI and ML solutions can provide tangible benefits (**Figure 3**).

Productivity indicators:

- Good produced parts/operator
- Good produced parts/total produced parts (scrap, setup, testing, etc.)

Consumption indicators:

- Material consumption (MC): weight of material consumed per time unit

$$MC\left(\frac{kg}{h}\right) = \frac{part\ weight\ g}{cycle\ time\ sec} * \frac{3600\ sec}{1\ hr} * \frac{1\ kg}{1000\ g} \tag{8}$$

- Specific energy consumption

The specific energy consumption (SEC) can be defined in terms of the amount of power (P) input into the system, divided by the process rate ($\dot{m}$):

$$SEC = \frac{P}{\dot{m}} \tag{9}$$

3. Artificial intelligence for availability

While guarantying high OEE, availability is key. OEE considers availability loss, which considers any event that stops the production plan for a significant amount of time, including unplanned and planned stops. An availability of 100% means the process is always running during planned production time.

There are other considerations which should be included in the availability computation, such as the changeover times. Changeovers are a source of setup and adjustment time, which is one of the main time loss reasons, and thus represent a valuable opportunity for improvement. Changeover times are most commonly improved (reduced) through the application of single-minute exchange of dies (SMED), which relies on performing as many changeover steps as possible while the equipment is running. In fact, these days equipment manufacturers tend to provide an availability rate in the specifications of their equipment, considering, among others, these changeovers.

But what can AI do for us? If we think in data processing and analytical capacities that can be run over information coming from equipment, we rapidly think in predictive maintenance to anticipate problems or virtual sensors to simulate, when feasible, some defective or malfunctioning sensors. Let us see some examples of this in the following subsections.

3.1 Virtual sensors

Virtual sensors (VS) are implemented with software to emulate real-world or even newly artificially defined sensors and are commonly used to (i) compute extra parameters derived from real sensors that are impossible to be measured, contributing to a better understanding of the whole environment, and (ii) simulate real sensor outputs. In the scope of this chapter, the second functionality becomes useful to mitigate system stops due to equipment failure or even planned maintenance, increase the availability of complex systems, and therefore improve the OEE.

For example, in a water treatment facility, where a lot of processes are continuously and simultaneously working to improve the quality of water, the decisions taken to manage the global system depends directly on the observations obtained by the sensors that are deployed along the premises. When any of those sensors is not working, the system cannot operate correctly because sometimes those input values are of utmost importance to determine which decision is correct.

In this case, a VS can be used to simulate and replace that lost sensor during the downtime. For this purpose, the VS is implemented through machine learning algorithms and is based on different inputs or sensors that are operating in the different parts of the water treatment cycle in the system.

Following this procedure, we showcase a VS simulating a measurement of one of the water quality parameters in a water treatment facility. In this case, this measurement is of utmost importance in the system because, depending on its observations, the processes adapt their execution parameters to fit the required quality requirements.

Therefore, we must overcome three main challenges, the combination of which increases considerably the complexity of the problem to be solved using AI/ML algorithms:

- The complexity of the processes: In water treatment facilities, physical and biological processes are combined to clean the water and achieve the expected levels of quality.

- The delayed responses: The water flow may be slow, so a change in the input state will not be immediately reflected in the rest of the system.

- The bad quality of the signals: In this kind of environments, where the sensors are in direct contact with dirty water, the observations usually contain anomalous values.

We start implementing the needed filters and preprocessing steps to clean and improve the data, but usually this is not enough, and ML algorithms cannot achieve the desired performance. Consequently, extra efforts are needed to obtain better models.

This example is a regression problem, where the target is a continuous value, and the predictors are composed of current and past values from other sensors which are part of the same process.

During the first iterations of the analysis, one of the main tasks was to select the optimal past values of each observation/sensor to be used as predictors. This process was done through the analysis of the importance of the variables once a model has been trained, selecting the N last values with the most importance. Also different frequencies of lags were tested using the same approach.

It is important to note that the target variable was not used to make next predictions, avoiding accumulated errors and allowing an infinite horizon of predictions, since the only requirement was the observations of the other sensors.

Different ML algorithms were tested and compared, and **Figure 4** showcases the three ML models that have better performance:

- XGBoost: Extreme gradient boosting. Optimized distributed gradient boosting library. Gradient boosting is a ML technique which produces a prediction model in the form of an ensemble of weak prediction models. It builds the model in a stage-wise fashion, training the weak models sequentially, each trying to correct its predecessor, and it generalizes them by allowing optimization of an arbitrary differentiable loss function [9].

- KNN: K-nearest neighbors. Nonparametric algorithm. Predictions are computed based on the mean of the labels of its nearest neighbors [10].

- RF: Random forests. Ensemble of decision trees, where each tree is usually built from a sample drawn with replacement (bagging method) from the training set. If the sample is obtained without resampling, the method is called pasting. When splitting each node during the construction of a tree, the best split is found either from all input features or a random subset of size *max_features* (RF algorithm hyperparameter) [11].

In order to compare the performance between model results, we are using the following metrics:

- Mean squared error (MSE) measures average squared error of our predictions, calculating the square difference between the predictions and the target and then the average of those values:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{10}$$

- Mean absolute error (MAE) is calculated as an average of absolute differences between the target values and the predictions:

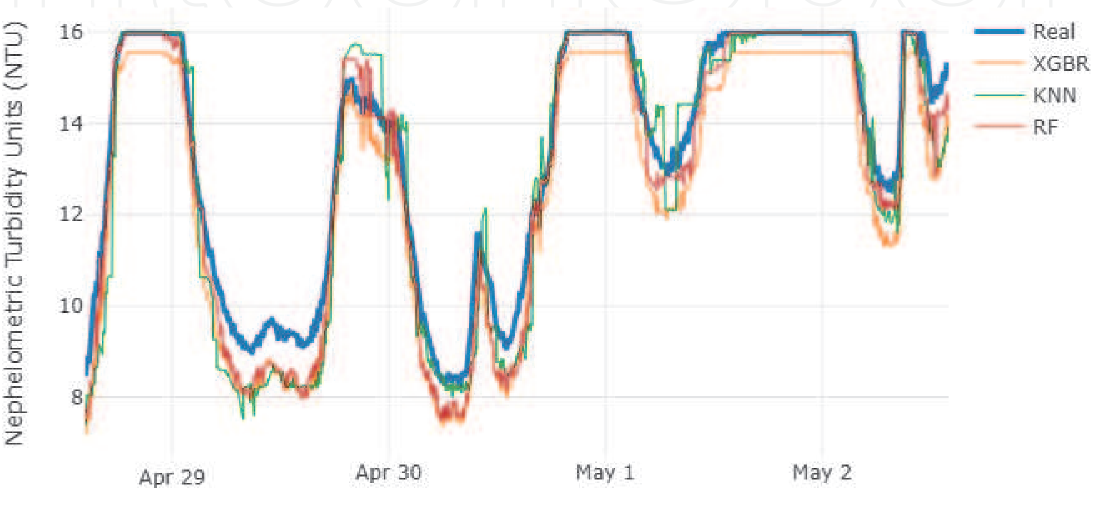

Figure 4.
Initial predictions.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{11}$$

- Explained variance score (EVS) measures the proportion to which a mathematical model accounts for the variation (represented as σ^2, s^2, or Var (X)) of a given data set:

$$EVS = 1 - \frac{Var\{y_i - \hat{y}_i\}}{Var\{y_i\}} \tag{12}$$

The best performance is achieved by random forests followed by KNN (MSE, 0.69; MAE, 0.43; EVS, 0.86) and XGBoost (MSE, 0.81; MAE, 0.85; EVS, 0.98). In all the cases, grid search [12] has been used to tune the hyperparameters.

The scorings seem to be acceptable, but analyzing one by one the predicted values (**Figure 5**), unusual behaviors appear in the predictions. So, in order to try to improve the outputs, an ensemble model is implemented combining the previous algorithms and following the stacking methodology (**Figure 6**, [13]), where a new ML algorithm (called blender or meta learner), in this case a ridge regressor [14], takes the previous predictions as inputs and makes the final prediction, usually better. The blender has been trained following the hold-out set approach.

Basically, the main idea is to, instead of taking the best model and use it to make predictions, try to combine the predictions of completely different ML algorithms, which are based on really different approaches and are good to operate in specific conditions, into a new ensemble which combines the best of each one, is able to operate in all the cases, and reduces the global error.

This process improves significantly the predictions (**Figures 7** and **8**), achieving the following scores, MSE, 0.27; MAE, 0.40; EVS, 0.98, and resulting in a ML model that is able to simulate the real sensor during downtimes, allowing the system to continue working normally.

3.2 Maintenance

We define predictive maintenance as the set of techniques used to determine the condition of equipment, allowing for a better and more personalized maintenance

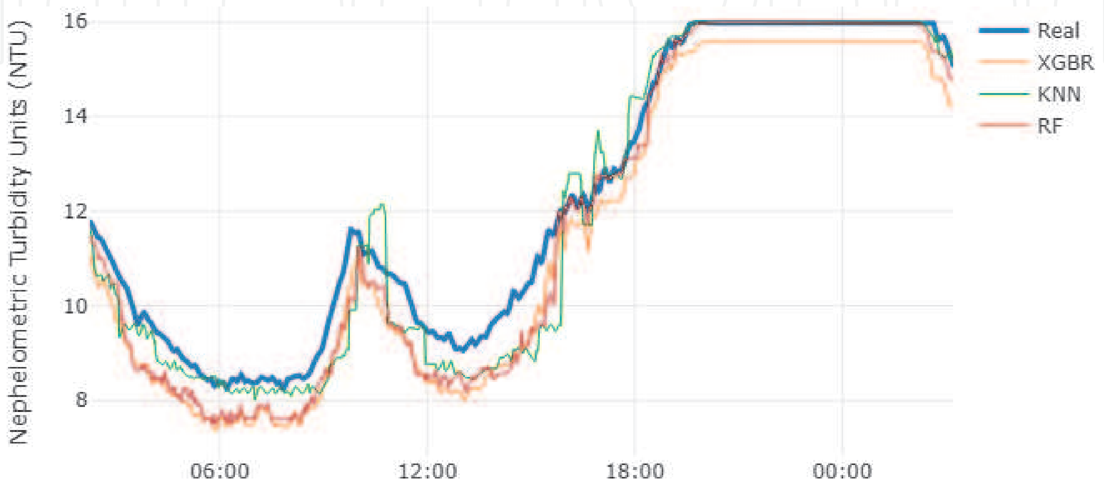

Figure 5.
Initial predictions detail.

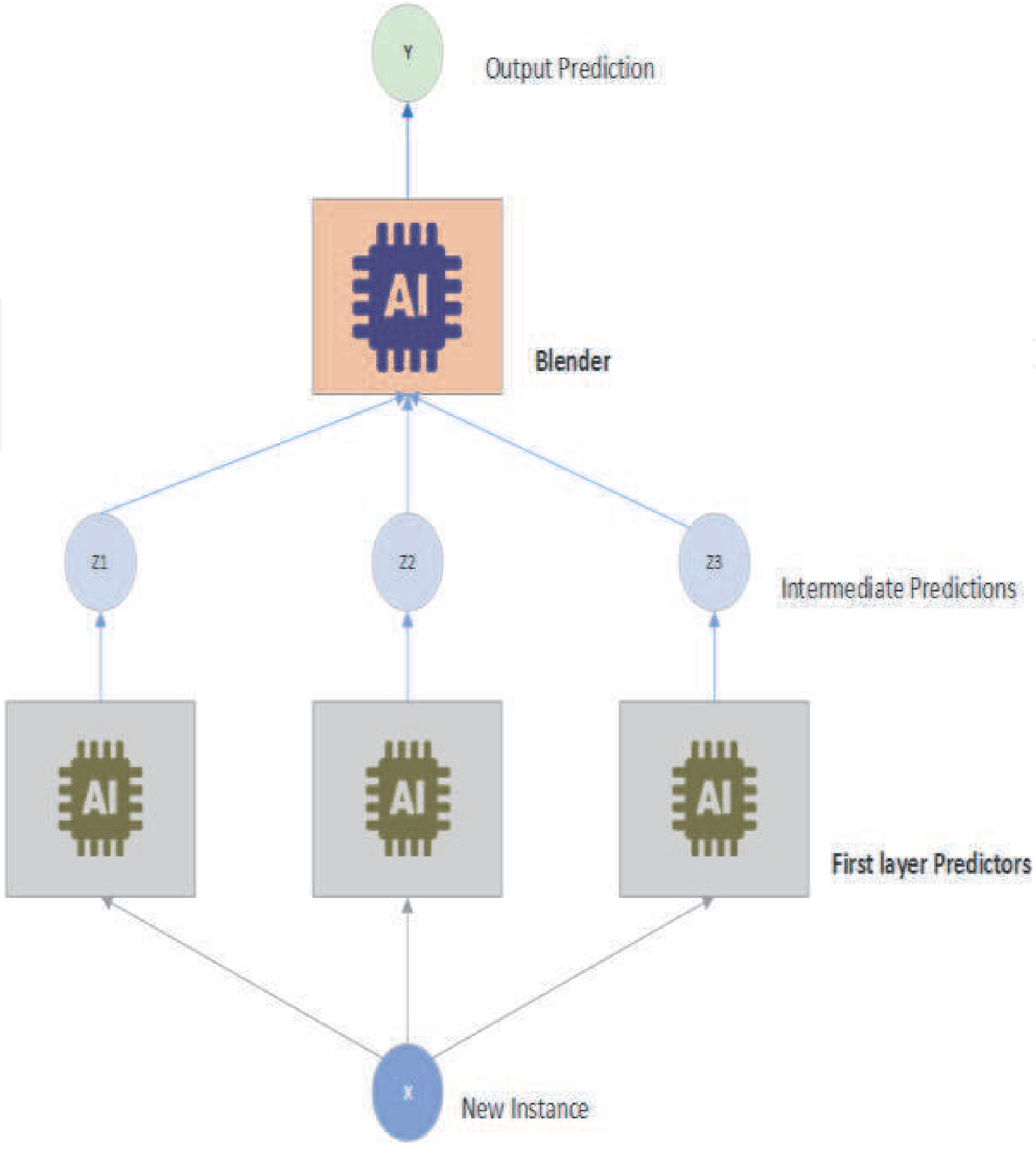

Figure 6.
Blending predictor schema.

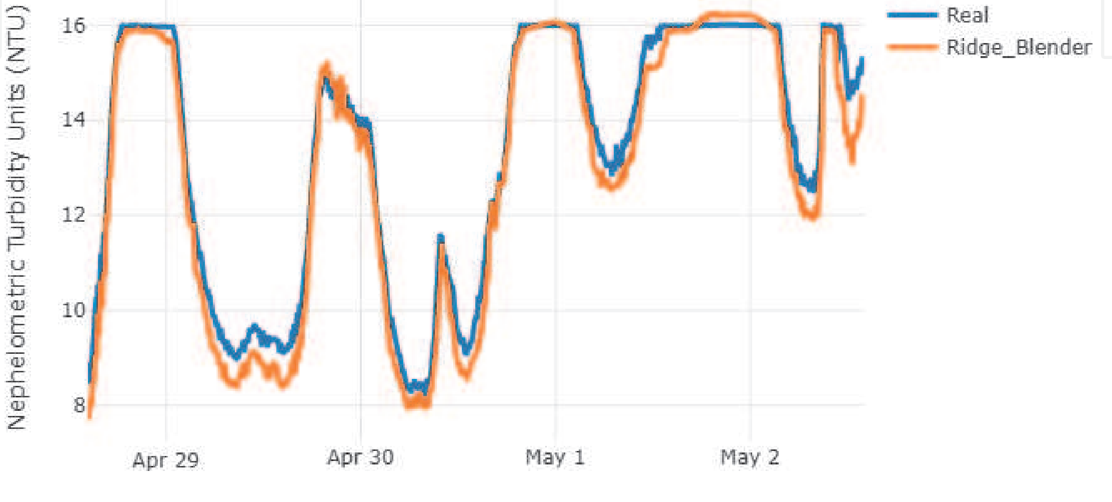

Figure 7.
Final virtual sensor predictions.

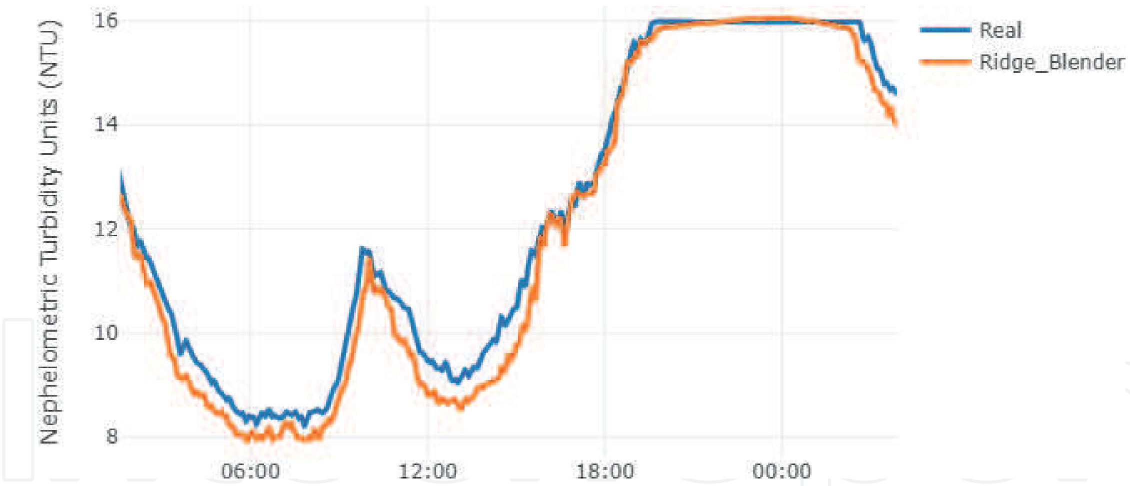

Figure 8.
Final virtual sensor predictions' detail.

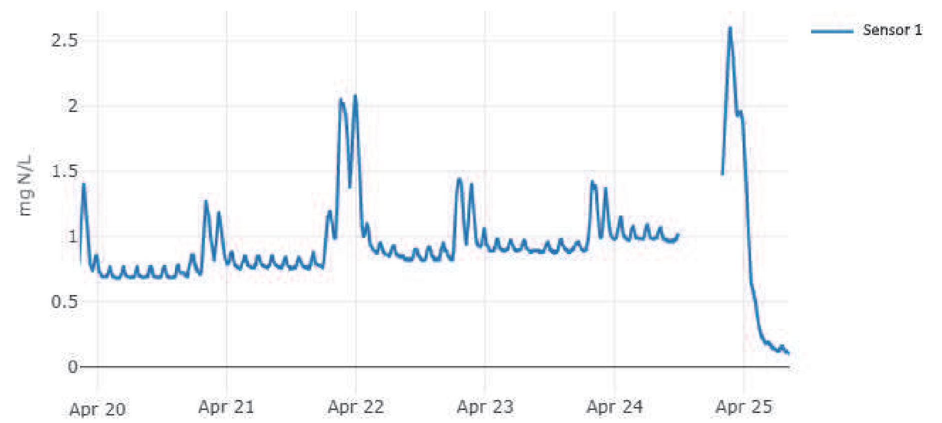

Figure 9.
Drift over the time.

plan. This plan depends on the performance, among other indicators, of the specific equipment (actual condition), instead of only relying on periodic maintenance routines, and this enables spare parts optimization, better maintenance actuations planning, and of course, OEE improvement due to its impact in availability, performance, or even quality.

Continuing on the water treatment facilities case introduced previously, one of the main problems faced in this environment is related to those sensors that are in direct contact with dirty water.

Over the time, a continuous and incremental drift appears in the observations of the sensors, thereby generating incorrect measurements. Since these measurements are the base of the system which takes operational decisions, the sequent of the taken actions will be incorrect, resulting in an unnecessary waste of resources or, even worse, an immediate stop of the system to repair and calibrate the sensors.

This pattern can be easily identified in **Figure 9**, having an incremental drift over the time until day 25, when the sensor was stopped during some hours for maintenance. Once the sensor is turned on again, the real value of the observations is shown, approximately 0.

Before the proposed approach, trying to prevent these problems, a set of preventive maintenances was defined, which consisted of manually taking measurements to compare them in the laboratory with the values of the sensors. Despite this, these actions were not enough, and the drift usually appeared before the scheduled maintenance, making necessary a better approach: a predictive maintenance-based approach.

There are different ways to implement a ML predictive maintenance solution. For example, it is possible to predict the remaining useful life of an equipment, which is a regression problem. But in this case, it has been defined as a binary classification problem, where the goal is to, given an observation (and the previous values), predict if there will be an anomaly in the following 24 hours (estimated minimum range of time to define a maintenance).

In the presented problem, the term anomaly refers to a sensor deviation or a drift in the observations measured by it, due to the contact with dirty water, making necessary a maintenance action in this specific sensor to clean or even replace it if it is necessary.

As in other classification problems, the basic requirement is labeled data, in this case, labeled anomalies. This was the main problem, there was a lot of historical data, but the anomalies were not labeled so the first step consisted of an anomaly detection problem.

Through unsupervised anomaly detection algorithms, such as:

- Isolation forest: Isolation forest algorithm isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. Since recursive partitioning can be represented by a tree structure, the number of splittings required to isolate a sample is equivalent to the path length from the root node to the terminating node. This path length, averaged over a forest of such random trees, is a measure of normality [15].

- Local outlier factor: Local outlier factor algorithm computes a score reflecting the degree of abnormality of the observations. It measures the local density deviation of a given data point with respect to its neighbors. The idea is to detect the samples that have a substantially lower density than their neighbors [16].

And thanks to an intensive data preprocessing steps such as data segmentation or feature engineering (which made the task easier to detect this specific anomaly), the historical dataset was labeled. Finally, a simple clustering algorithm was run to discard different anomalies.

The result of the anomaly detection analysis is shown in **Figure 10**, where sensor 1 is measuring a value different than 0 (anomaly), and therefore the system tries to force a response increasing excessively the resource measured by sensor 2.

Finally, we face the predictive maintenance classification problem, where the key was the definition of the target variable: a binary column indicating whether in

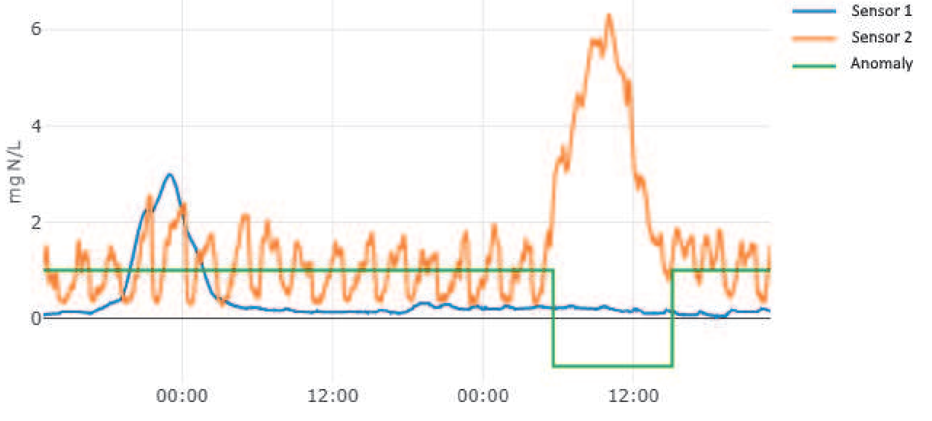

Figure 10.
Anomaly detection results.

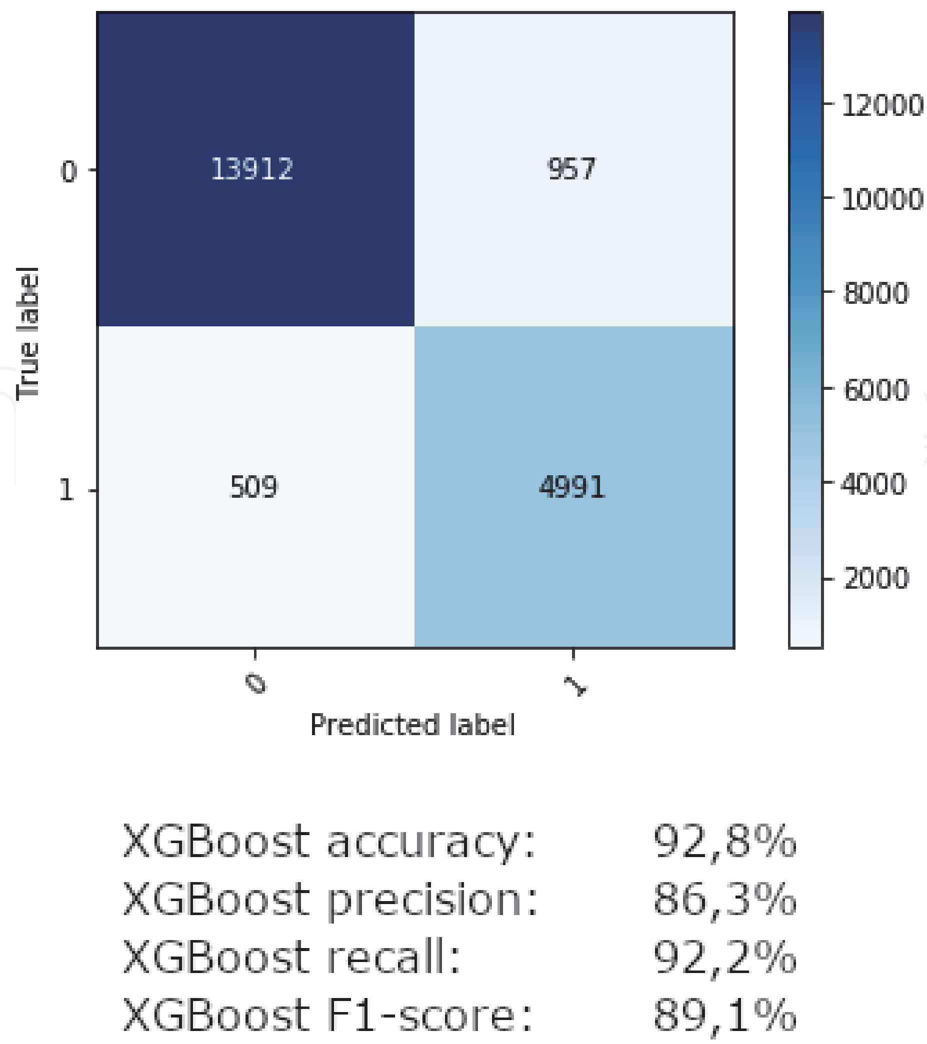

Figure 11.
Confusion matrix.

the next 24 hours an anomaly is detected or not. At this point, different ML classification algorithms were tested, and the best performance was achieved by XGBoost, obtaining the following classification results (**Figure 11**) in a test set.

In order to measure the algorithm performance in the classification task, we are using confusion matrix and the following metrics:

- Confusion matrix: In a ML classification problem, a confusion matrix is a specific table that simplifies the analysis of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instance's real class (or vice versa) [17].

- Accuracy: Classification metric that computes the fraction of correct predictions:

$$\mathrm{accuracy}(y, y') = \frac{1}{n_{\mathrm{samples}}} \sum_{i=0}^{n_{\mathrm{samples}}-1} 1(y'_i = y_i) \tag{13}$$

- Precision: Classification metric that computes the fraction of relevant instances among the retrieved instances. It is also called positive predictive value:

$$\mathrm{precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{14}$$

- Recall: Classification metric that computes the fraction of relevant instances that have been retrieved over the total amount of relevant instances. It is also called positive predictive value:

$$\text{recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{15}$$

- F1-score: Classification metric that computes a weighted harmonic mean of the precision and recall. F1 score reaches its best value at 1 and worst score at 0.

$$\text{F1 score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \tag{16}$$

As depicted in **Figure 11**, the final version of the model provides good results while predicting anomalies with time enough to articulate the needed preventive actions. Not only the accuracy is important, but we would also like to remark that the false negative rate is low, that is, the algorithm performs very well in detecting anomalies, and only a few of them are undetected.

3.3 Process setup

Process setup, especially during changeover operations, can affect the availability indicator and thus represents an opportunity for manufacturing AI and ML based solutions. New production trends based on a high degree of flexibility, customization, and small batches require for an extra effort in terms of process setup and scheduling. For instance, in plastic injection molding quite often due to production flexibility and scheduling, a mold needs to be re-installed and set up for production again in order to deliver a new production batch to the final customer. This situation requires for a new tuning process involving an important waste of time, material, and energy. This situation opens the opportunity for developing supervised ML models to compare past production data with real-time data for recommending tuning parameters and reach in a shorter time frame the optimal process operation.

To this end, the real-time evolution of a key process parameter can be used as training of the manufacturing process setup or configuration. By comparing the actual real-time evolution within the manufacturing cycle versus the known optimal (acquired from previous production runs), a set of recommendations can be provided. This strategy can boost the process setup, providing recommendations to reach the optimal targeted key parameter cycle evolution, following an iterative method as depicted in **Figure 12**.

Following the plastic injection molding example, within the PREVIEW project [18], a set of experimental trials were performed in order to create the historical database that supports the AI system in charge of providing process tuning recommendations. Within the AI solution, different algorithms were tested for comparing new sensor data versus historical data to provide tuning recommendations. **Figure 13** shows a PREVIEW project result using random forest trees [19] to provide tuning recommendations when the injection speed parameter was changed to different operational points. As can be seen, for lower than optimal injection speeds, the AI system based on RF recommends increasing the injection speed, while for higher injection speeds recommends a reduction, driving always the parameter toward the optimal operational window that leads to optimal cavity pressure evolution within the manufacturing cycle.

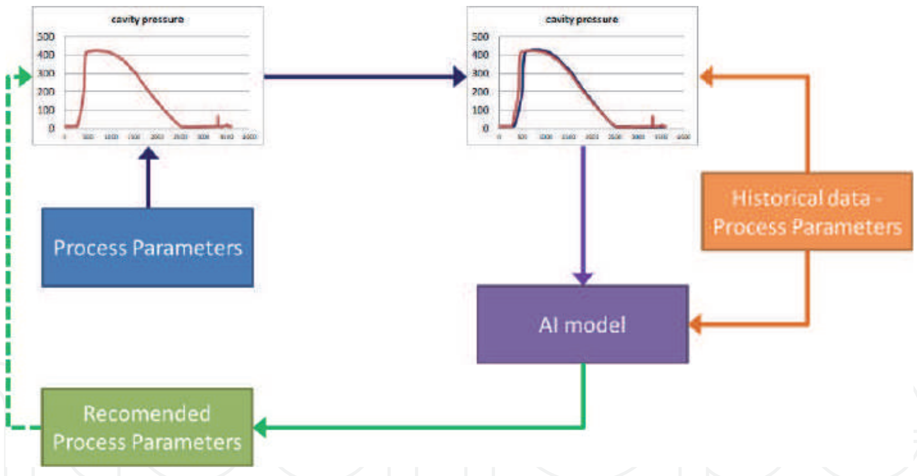

Figure 12.
Iterative comparison to optimized production setup comparing known optimal process parameters versus new acquired ones.

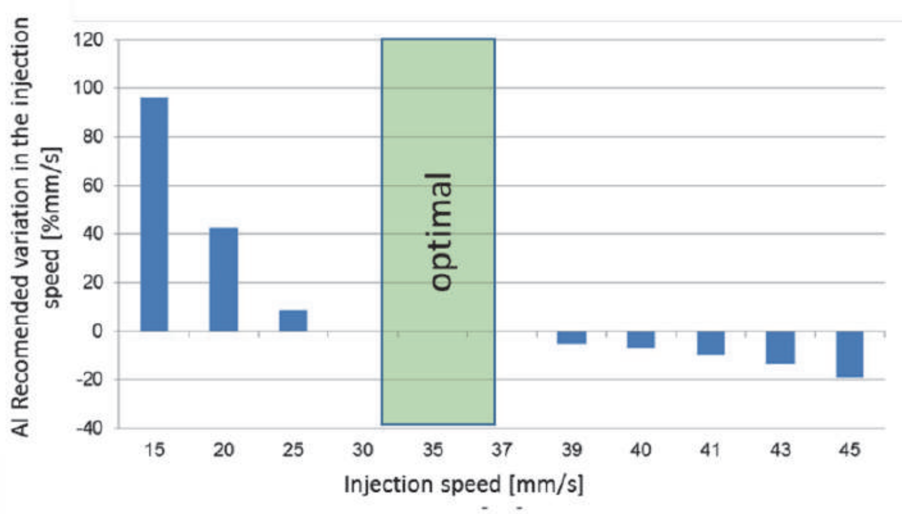

Figure 13.
Process optimization recommendation. The PREVIEW project result.

4. Artificial intelligence for quality

It is a well-known problem that high added-value industrial and manufacturing processes combining several operations (welding, milling, etc.) and thus heterogeneous data sources do not always reach their maximum performance potential due to the lack of powerful and tailored solutions for data analysis toward the zero defects manufacturing paradigm. Today's artificial intelligence and machine learning based solutions are mature enough to boost production processes by means of exploiting the process data generated thanks to the in-line sensors, workers' feedback, reports, quality control, etc. Thus, developing a tailored predictive quality solution based on artificial intelligence and machine learning has become a crucial key element for impacting OEE to prevent the manufacturing of non-quality parts and its exportation to the final client. Several research works have been carried out for different manufacturing processes, including plastic injection molding, foundry, milling, welding, etc. (e.g., see [20–24]) showing the potential benefits of applying AI and ML to exploit process data.

Continuous quality estimation at each step of the manufacturing process by means of machine learning and artificial intelligence, applied on the in-line acquired data, enables predictive warnings and alarms even before the target quality is affected and thus quality indicator of OEE is degraded. Two different approaches can be implemented when developing AI predictive quality tools: supervised versus unsupervised solutions. Supervised solutions can provide a better accuracy when predicting undesired quality deviations, but a properly tagged dataset is required. Unsupervised methods have the benefit of not requiring the tagged dataset and are typically used for anomaly detection, meaning strong quality deviations. Moreover, supervised system results can be tracked down and analyzed to provide process insights which can lead to knowledge discovery [25] solutions that help to address the root cause of the undesired quality deviation and thus improve quality and, therefore, OEE.

Focusing on supervised solutions, a proper dataset labeling is a key element. It is highly recommended to perform a Design of Experiments (DOE) where quality deviations are forced in order to obtain a more balanced dataset compared to the typical production dataset where non-quality parts are rare. In the case of qualitative quality labels (e.g., good, bad, type of defect, etc.), a classifier will be preferred, while for quantitative quality indicators (e.g., weigh, tensile strength, etc.), a regressor will be implemented.

Let us consider as illustrating example a plastic injection molding quality prediction problem. The four-cavity mold used for the experimental trials can be seen in **Figure 14**. Only one cavity was sensorized to obtain the pressure and temperature evolution of the melt during the production cycle. The machine pressure and screw position were also acquired for each one of the 199 injected parts of the trial. The injection cycle was 7.2 seconds and was sampled at 500 Hz. Thus, the dataset is the time series evolution of the key parameters of the process.

The DOE was designed in order to obtain seven different part qualities: good, short shot, shrinkage, flash, jetting, over-compaction, and flow lines. The different qualities were obtained by means of varying the injection machine configuration. A total of 199 parts were produced (**Figure 15**).

Depending on the data granularity (continuous, cycle, or batch), different preprocessing techniques can be implemented to boost the performance of the later machine learning classifier. For instance, entropy analysis and complexity reduction algorithms such as principal component analysis (PCA) [27] can provide a substantial advantage as seen in **Figure 16**, where the PCA projection of the screw position sensor is plotted.

In order to compare the performance of different machine learning classifiers, a benchmark based on cross-validation techniques was implemented, using a stratified shuffle split [28] strategy to preserve the percentage of samples of each class

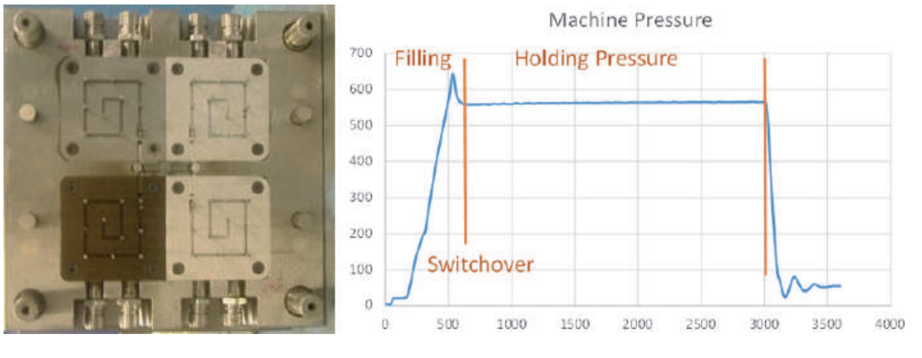

Figure 14.
Mold cavity picture and example of acquired machine pressure. Experimental data provided by EURECAT [26].

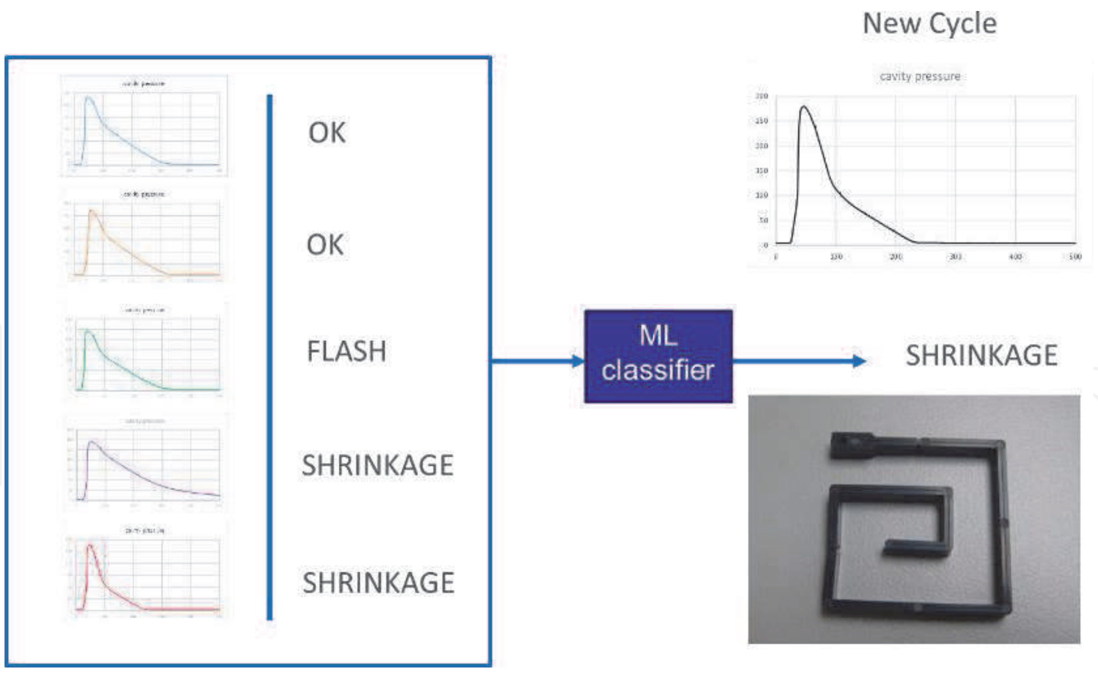

Figure 15.
Supervised approach for quality prediction.

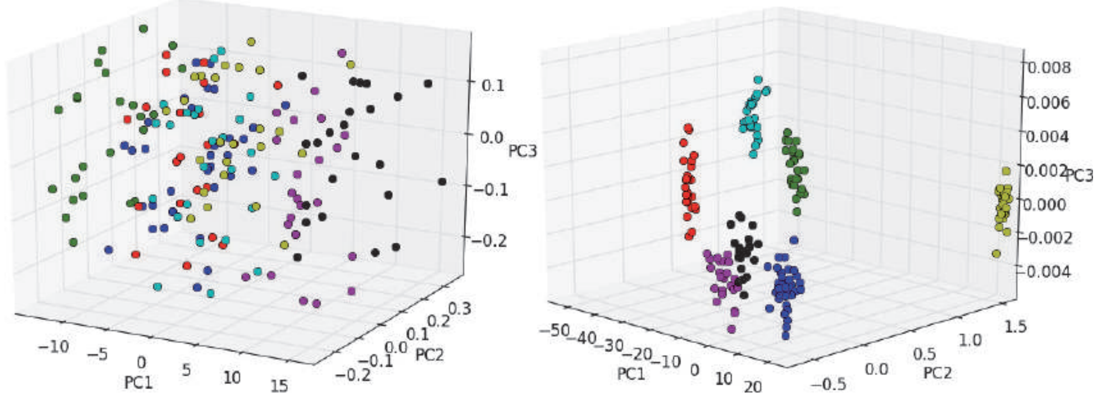

Figure 16.
3D PCA data projection using the raw data or a preprocessed data where the 10 time stamps with higher entropy are selected. Each color represents a different part quality or defect.

(quality). This test can be run for each sensor or by applying data fusion and combining all sensors in a single dataset (**Figures 17** and **18**).

As can be seen in **Figure 19**, support vector machines [29] with a linear kernel show a low performance, while ensemble algorithms like random forest trees and gradient boosting [30] present higher accuracy rates, especially when 50 estimators or more are used.

When combining all sensor information by means of applying data fusion, the quality prediction accuracy increases near to 100%, as can be seen in **Figure 18**. This result and system allow for an in-cycle quality preventive alarm that can lead to an important reduction of scrap rate and exported non-quality, which automatically translate to a higher quality rate and a reduction of costs due to wasted raw material and energy consumption while improving OEE.

Other manufacturing processes can have different sampling rates or even create batch datasets where for each part a set of relevant values are recorded. Typically, large batch datasets present a high degree of data heterogeneity, compiling sensor values, reports, environmental data, etc. Moreover, part traceability may not be

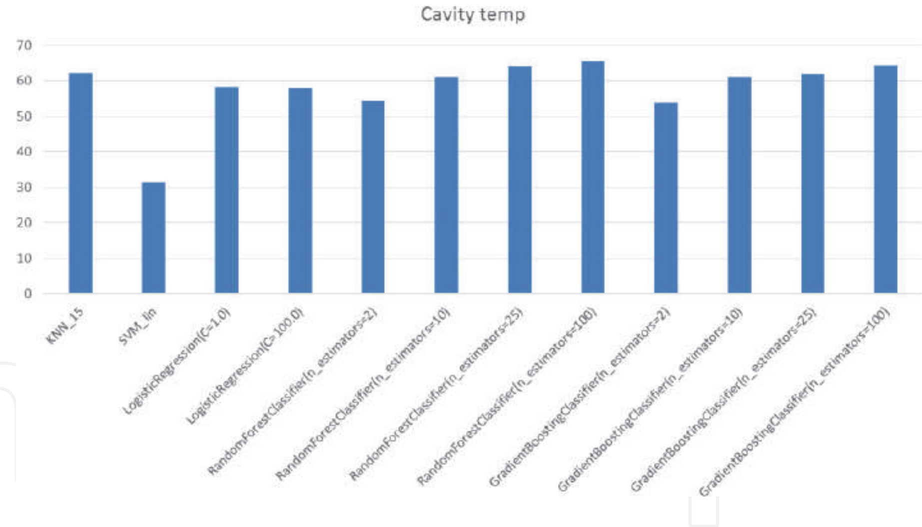

Figure 17.
Quality prediction mean accuracy for a 20-round cross-validation test using 70% for training and 30% of samples for test, using only the cavity temperature sensor data.

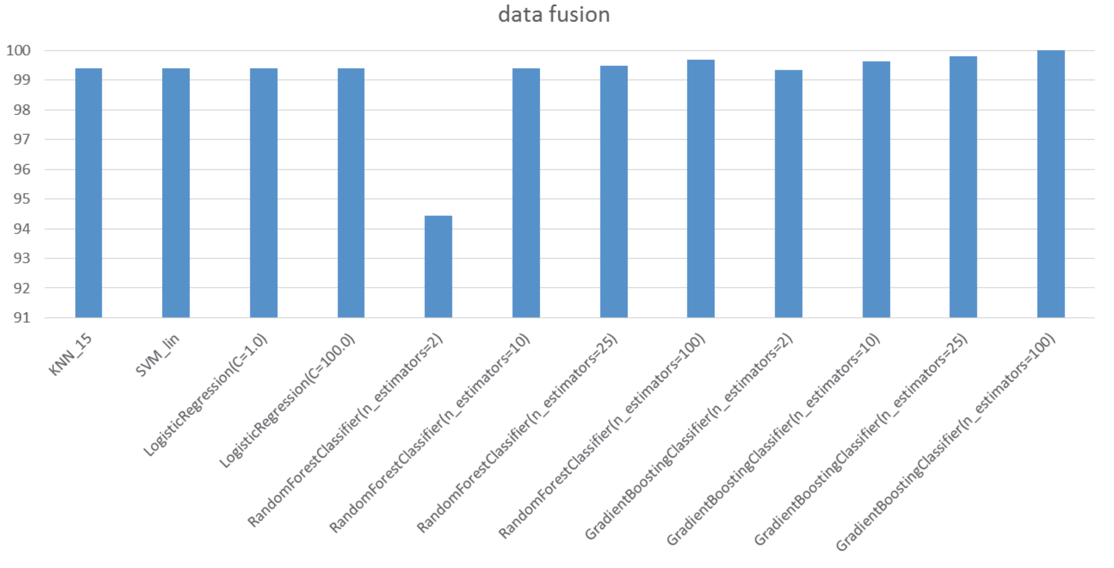

Figure 18.
Quality prediction mean accuracy for a 20-round cross-validation test using 70% for training and 30% of samples for test, combining the available cavity and machine sensor data.

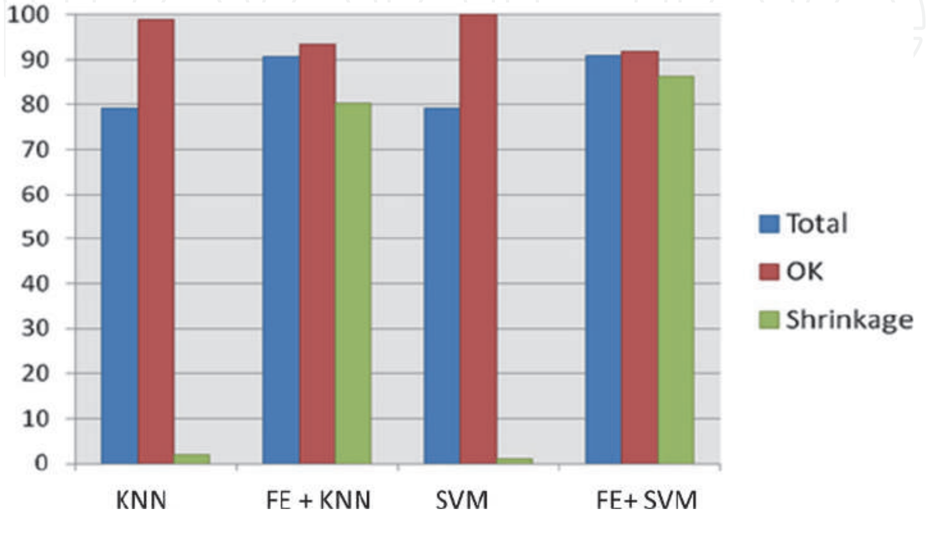

Figure 19.
Batch quality prediction with and without feature engineering (FE) for a heterogeneous and class-unbalanced dataset. Iron foundry case.

guaranteed, and thus quality rates may refer to the entire batch. In these scenarios, feature engineering [31] can provide a clear advantage for boosting the performance of the ML and AI algorithms. **Figure 19** shows the performance of KNN and SVM classifiers for a foundry dataset with more than 250 different parameters (chemical composition of the iron, climate, process data, sensor data, etc.) with and without feature engineering.

The dataset had two main difficulties: the extreme unbalance between classes (qualities) and the data heterogeneity. By applying feature engineering, the number of parameters can be reduced, focusing on the relevant ones. Bagging [32] and cost functions were used to face the class unbalance. Batch quality prediction based on process data can help in reducing the exported non-quality while providing knowledge discovery insights to find and correct the root causes of the undesired quality deviation.

5. Artificial intelligence for performance

Performance indicators consider any factor that causes the manufacturing process to run at lower speed than its maximum possible speed. For instance, slow cycle time affects performance indicators. For this reason, it is key to know the ideal cycle time, which is the fastest cycle time that can be achieved in optimal circumstances. Moreover, performance is also affected by idling time and minor stops.

Cycle time reduction is one of the main factors for improving productivity. A cycle time reduction contributes to reaching the optimal production throughputs, reduction of time to market, better scheduling, and a reduction of associated costs in terms of labor, energy, and raw material when combined with quality prediction and assessment. The reduction of cycle time has become a relevant topic both in research and in practical applications. Neural networks and machine learning algorithms can help to predict and optimize manufacturing cycle time in different sectors (e.g., see [33, 34]).

Preventive alarms generated by predictive quality systems based on AI and ML can prevent manufacturing at nonoptimal operation setups and thus prevent minor stops. Minor stops can also be reduced thanks to preventive maintenance systems. Case-based reasoning [35] systems can leverage past experiences to help manufacturing processes run faster. For instance, a CBR system can provide helpful recommendations for optimizing the cooling time based on the type of material and the thickness of the part that is being manufactured. The CBR system provides the most similar cases based on a defined similarity metric, and thus a previous cooling times of well-known and optimized processes can be taken as reference. Illustrating this case, the Des-MOLD project [36] developed an AI system based on CBR and argumentation [37] to help plastic injectors share their experiences and benefit from mold design and manufacturing process optimization [38].

6. Conclusions

Artificial intelligence and machine learning based solutions can provide a competitive advantage in today's manufacturing paradigm, redefined by the Industry 4.0 revolution and the massive data available thanks to CPS, virtual sensors, and IIoT devices. Leveraging this data has become a very relevant topic both in research and for practical applications due to its massive potential. Data-driven solutions are becoming more and more popular due to its potential both in terms of prediction and to its capacity to provide process insights for enhancing process owner's expertise.

This chapter has focused on how the leverage of the available process data by means of AI and ML solutions can impact into one of the most relevant manufacturing indicators: overall equipment efficiency. OEE has three main components: availability, quality, and performance. Each OEE component tackles a different challenge and thus may require a different approach. Through different experimental examples, each OEE component and how AI solutions can impact it have been described. It has been shown how predictive maintenance and virtual sensor solutions can help in reducing the undesired production breakdowns and thus increase equipment availability. Predictive quality solutions based on supervised algorithms, for either real-time cycle data or batch data, have been described, showing the importance of feature engineering for boosting prediction accuracy. And finally, equipment performance focusing on cycle time has been addressed by CBR for leveraging past experiences and providing process tuning types to run at the highest throughputs.

OEE will be further improved thanks to the new AI trends and technologies that are being researched right now, providing even more powerful and tailored solutions. Availability and performance indicators could be greatly improved when mature reinforced learning approaches are available at the production level, reducing setup times and optimizing cycle times thanks to the collaboration between human expertise and AI systems. Image processing through deep learning and convolutional neural networks can impact quality, especially for visual defects. Collaborative human-AI systems are envisaged as key for the next Industry 5.0.

Author details

Francesc Bonada[1]*, Lluís Echeverria[2], Xavier Domingo[2] and Gabriel Anzaldi[2]

1 Eurecat Technology Centre, Av. Universitat Autònoma, Cerdanyola del Vallès, Barcelona, Spain

2 Eurecat Technology Centre, Parc Científic i Tecnològic de Lleida, Lleida, Spain

*Address all correspondence to: francesc.bonada@eurecat.org

References

[1] Huang SH, Dismukes JP, Shi J, Su Q, Razzak MA, Bodhale R, et al. Manufacturing productivity improvement using effectiveness metrics and simulation analysis. International Journal of Production Research. 2003;**41**(3):513-527. Available from: https://doi.org/10.1080/0020754021000042391

[2] Jauregui Becker JM, Borst J, Van Der Veen A. Improving the overall equipment effectiveness in high-mix-low-volume manufacturing environments. CIRP Annals—Manufacturing Technology. 2015;**64**:419-422. DOI: 10.1016/j.cirp.2015.04.126

[3] Fleischer J, Weismann U, Niggeschmidt S. Calculation and optimisation model for costs and effects of availability relevant service elements. Proceedings of LCE. 2006:675-680

[4] Muchiri P, Pintelon L. Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion. International Journal of Production Research. 2008;**46**(13):3517-3535. DOI: 10.1080/00207540601142645

[5] Singh R, Shah DB, Gohil AM, Shah MH. Overall equipment effectiveness (OEE) calculation—Automation through hardware & software development. Procedia Engineering. 2013;**51**:579-584. DOI: 10.1016/j.proeng.2013.01.082

[6] Hedman R, Subramaniyan M, Almström P. Analysis of critical factors for automatic measurement of OEE. Procedia CIRP. 2016;**57**:128-133. DOI: 10.1016/j.procir.2016.11.023

[7] Mâinea M, Dută L, Patic PC, Căciulă I. A method to optimize the overall equipment effectiveness. IFAC Proceedings Volumes (IFAC-PapersOnline). 2010;**43**(17):237-241. DOI: 10.3182/20100908-3-PT-3007.00045

[8] Gabahne L, Gupta M, Zanwar D. Overall equipment effectiveness improvement: A case of injection molding machine. The International Journal of Engineering and Science (IJES). 2014;**3**(8):1-10

[9] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM; 2016. pp. 785-794. DOI: 10.1145/2939672.2939785

[10] Song Y, Liang J, Lu J, Zhao X. An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing. 2017;**251**:26-34. DOI: 10.1016/j.neucom.2017.04.018

[11] Prasad AM, Iverson LR, Liaw A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems. 2006;**9**:181-199. DOI: 10.1007/s10021-005-0054-1

[12] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research. 2012;**13**:281-305

[13] Breiman L. Stacked regressions. Machine Learning. 1996

[14] Björkström A, Sundberg R. A generalized view on continuum regression. Scandinavian Journal of Statistics. 1999;**26**(1):17-30. DOI: 10.1111/1467-9469.00134

[15] Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proceedings of the IEEE International Conference on Data Mining, ICDM. Pisa, Italy: IEEE; 2008. pp. 413-422. DOI: 10.1109/ICDM.2008.17

[16] Breuniq MM, Kriegel HP, Ng RT, Sander J. LOF: Identifying density-based local outliers. In: SIGMOD Record (ACM Special Interest Group on Management of Data). New York, USA: ACM; 2000. pp. 93-104. DOI: 10.1145/342009.335388

[17] Ting KM. Confusion matrix. In: Encyclopedia of Machine Learning and Data Mining. Boston, USA: Springer; 2017. DOI: 10.1007/978-1-4899-7687-1_50

[18] PREVIEW (H2020-FoF-2014). 2014. Available from: http://www.preview-project.eu/

[19] Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. Vol. 1. Montreal, Canada: IEEE; 1995. pp. 278-282. DOI: 10.1109/ICDAR.1995.598994

[20] Huggett DJ, Liao TW, Wahab MA, et al. Prediction of friction stir weld quality without and with signal features. The International Journal of Advanced Manufacturing Technology. 2018;**95**(5–8):1989-2003. DOI: 10.1007/s00170-017-1403-x

[21] Perzyk M, Kochanski A. Detection of causes of casting defects assisted by artificial neural networks. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture. 2003;**217**(9):1279-1284. DOI: 10.1243/095440503322420205

[22] Ribeiro B. Support vector machines for quality monitoring in a plastic injection molding process. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2005;**35**(3):401-410. DOI: 10.1109/TSMCC.2004.843228

[23] Denkena B, Dittrich MA, Uhlich F. Augmenting milling process data for shape error prediction. Procedia CIRP. 2016;**57**:487-491. DOI: 10.1016/j.procir.2016.11.084

[24] Purr S et al. Stamping plant 4.0—Basics for the application of data mining methods in manufacturing car body parts. Key Engineering Materials. 2015;**639**:21-30. DOI: 10.4028/www.scientific.net/kem.639.21

[25] Maimon O, Rokach L. Data Mining and Knowledge Discovery Handbook. Berlin, Heidelberg: Springer-Verlag; 2005. DOI: 10.1007/978-0-387-09823-4

[26] Available from: www.eurecat.org

[27] Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. 1901;**2**(11):559-557. DOI: 10.1080/14786440109462720

[28] Available from: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html

[29] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag; 2000. DOI: 10.1007/978-1-4757-3264-1

[30] Friedman JH. Greedy function approximation: A gradient boosting machine. The Annals of Statistics. 2001;**29**(5):1189-1232. DOI: 10.1214/aos/1013203451

[31] Zheng A, Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. Sebastopol, USA: O'Reilly Media Inc; 2018. ISBN: 1491953195, 9781491953198

[32] Breiman L. Bagging predictors. Machine Learning. 1996;**24**(2):123-140. DOI: 10.1023/A:101805431

[33] Gelmereanu C, Morar L, Bogdan S. Productivity and cycle time prediction using artificial neural network. Procedia

Economics and Finance. 2014;**15**: 1563-1569. DOI: 10.1016/S2212-5671(14)00626-1

[34] Kuo C, Chien C, Chen J. Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time. IEEE Transactions on Automation Science and Engineering. 2011;**8**(1):103-111. DOI: 10.1109/TASE.2010.2040999

[35] Riesbeck CK, Schank RC. Inside Case-Based Reasoning. Hillsdale, USA: Lawrence Erlbaum Associates Publishers; 1989. ISBN: 9780898597677

[36] Des-MOLD (FP7-NMP2-SL-314581). Available from: www.desmold.eu

[37] Toni F. A tutorial on assumption-based argumentation. Journal of Argument and Computation. 2014;**5**(1): 89-117. DOI: 10.1080/19462166.2013.869878

[38] Cabanillas D, Bonada F, Ventura R, Toni F, Evripidou V, Cartens L, et al. A combination of knowledge and argumentation based system for supporting injection mould design. In: Proceedings of 16th Catalan Congress of Artificial Intelligence (CCIA). Vic, Spain: IOS Press; 2013. pp. 293-296