mbr Multidisciplinary Business Review

Un journal de asfae chile

A systematic literature review and taxonomy proposition of machine learning techniques in smart manufacturing

Una revisión sistemática de la literatura y propuesta taxonómica de las técnicas de aprendizaje automático en la fabricación inteligente

Frederico de Oliveira Santos[a]
Ivanete Schneider Hahn[b]

Classification: Conceptual paper
Received: January 6, 2022
Revised: April 19, 2023; October 13, 2023
Accepted: October 29, 2023

Abstract

The purpose of this paper is to analyse the use of machine learning in smart manufacturing, describing techniques, technologies, industries, and purposes associated with industrial applications. We conducted a systematic literature review using Scopus, in which 26,032 documents were found. After applying quality criteria, 107 articles were analysed. The main findings show that machinery was the industry subsector with the major implementations regarding machine learning; process improvement is the main concern (interest) of all implementations; random forest was the most specific machine learning technique used; and diverse technologies associated with this context were identified such as: the industrial internet of things, digital twin, sensor technologies (soft, optical, barometric, ultrasonic), software technologies (Python, MATLAB, LabView, Google AutoML Platform) and equipment technologies (robotic, PLC, CNC). Most fault detection machine learning applications were focused on predictive maintenance, specifically in mechanical equipment (bearings, machines in general, and assembly lines). This study presents a novel taxonomy that identifies 85 specific machine-learning techniques used in smart manufacturing.

Keywords: Industry 4.0., Cyber-physical system, Machine learning, Operations, Technology.

Resumen

El objetivo de este artículo es analizar el uso del aprendizaje automático en la fabricación inteligente, describiendo técnicas, tecnologías, industrias y propósitos asociados con las aplicaciones industriales. Con base en una revisión sistemática de la literatura en Scopus, se encontraron 26 032 documentos y una vez cumplidas todas las preguntas de calidad, se analizaron 107 artículos. Los hallazgos principales muestran que la maquinaria fue el subsector industrial con las mayores implementaciones en lo que respecta al aprendizaje automático; la mejora de procesos es la principal preocupación (interés) de todas las implementaciones; el bosque aleatorio fue la técnica de aprendizaje automático más específica utilizada; y se identificaron diversas tecnologías asociadas a este contexto, como: el internet industrial de las cosas, el gemelo digital, las tecnologías de sensores (suaves, ópticas, barométricas, ultrasónicas), las tecnologías de software (Python, MATLAB, LabView, Google AutoML Platform) y las tecnologías de equipos (robóticas, PLC, CNC). La mayoría de las aplicaciones de aprendizaje automático de detección de fallas se centraron en el mantenimiento predictivo, específicamente en equipos mecánicos (rodamientos, máquinas en general y líneas de montaje). La originalidad de este artículo está en que diseñamos una taxonomía que incluye 85 técnicas específicas de aprendizaje automático utilizadas en la fabricación inteligente.

Palabras clave: Industria 4.0, sistema ciberfísico, aprendizaje automático, operaciones, tecnología.

[a] Instituto Federal Catarinense (IFC). Videira, Santa Catarina, Brazil. E-mail: fredautom@gmail.com. ORCID: https://orcid.org/0000-0002-7510-4854.
[b] Institution / Affiliation: Professor and Researcher at Universidade Alto Vale do Rio do Peixe. Caçador, Santa Catarina, Brazil. E-mail: ivischneider@hotmail.com. ORCID: https://orcid.org/0000-0003-0552-7496.

Introduction

Artificial Intelligence (AI) and Machine learning (ML) represent an important evolution in computer science and data processing systems which can be used in order to improve almost every technology-enabled service, product, and industrial application (Soori et al., 2023). More specifically, ML is a branch of artificial intelligence involving teaching a computer to read and interpret data (Srinivas & Young, 2023), and it is a subfield of artificial intelligence and computer science. It focuses on using data and algorithms to simulate the learning process of machines and enhance the accuracy of the systems (Magar & Farimani, 2023).

ML provides machines with the capability of learning from experience through data acquisition (Kotu & Deshpande, 2019). The main difference between AI and ML is that AI is developed for decision making and ML is developed for learning new things from data (Ozgulbas & Koyuncugil, 2019).

Over the past years, studies on ML and the combination of different ML methods have become a trend (Ren et al., 2022), and a growing body of research is applying ML in several areas, such as business - considering ML a tool that can be used in cryptocurrency research (Ren et al., 2022), for stock market prediction (Mintarya et al., 2023); in health and medical research, there are studies for prediction in medical and surgical research (Srinivas & Young, 2023), to identify the predictors of smoking (Bickel et al., 2023); ML based methodologies have also accelerated the prediction of the physical properties of materials (Magar & Farimani, 2023), and many others.

However, ML is still experimental because no universal learning algorithm exists, even though the number of ML algorithms is extensive and growing (Liu et al., 2021; Nguyen et al., 2019). ML is a crucial tool to provide suitable, efficient, and innovative solutions to Industry 4.0. as the smart manufacturing environment is shaped by great velocity, variety and volume of data, hidden patterns, and complex needs. Moreover, there is a lack of studies researching the manufacturing environment. So, by studying the manufacturing environment we contribute in at least two ways: the first, is to the literature, by contributing to the understanding of how AI and ML are already developed in this context; and secondly, we contribute by providing associated techniques, technologies, industries, and purposes with industrial applications.

In this context of opportunities, based on a systematic literature review, this study aimed to report the use of ML in a smart manufacturing context. In addition, a taxonomy of ML techniques was designed to contribute to the literature in order to report applied techniques and head the choice of the most specific suitable machine learning technique in a smart manufacturing environment considering its basic technique and several points such as the data analysis tasks, the way it is disposed of, amount of data, desired accuracy, time of response in training phase, data memory and timestamp.

To this end, the balance of the paper is structured as follows: Section 2 - reviews the relevant literature; Section 3 - the methodological procedures of the research are explained, followed by the analysis and discussion of the results; Section 4 - final considerations, research limitations, and recommendations for future research are presented.

Theoretical background

Digital transformation is creating a new kind of economy that acquires, treats, and disposes of real-time data generated by millions of connected people and machines. This data is captured not only by computers, but also by smartphones, sensors, social media platforms, satellites, and smart equipment in general. The spread of these devices enables the "datafication" of virtually any aspect of human social, political, and economic activity. It turns the knowledge-based economy (KBE) into the data-driven economy (DDE) which creates new and significant economic management challenges such as the emergence of machine knowledge capital as a rival to specialised human capital. Complementarily, DDE fosters the precise ability of computers to extract systematic data and instigates the design of techniques, algorithms, and methods to create useful information (Ciuriak, 2018). In this DDE perspective, the combination of smart sensors, computing infrastructure, and technologies such as the Industrial Internet of Things is becoming increasingly pervasive on the factory shop floor shaping Cyber-Physical Systems (CPS) (Upasani et al., 2017).

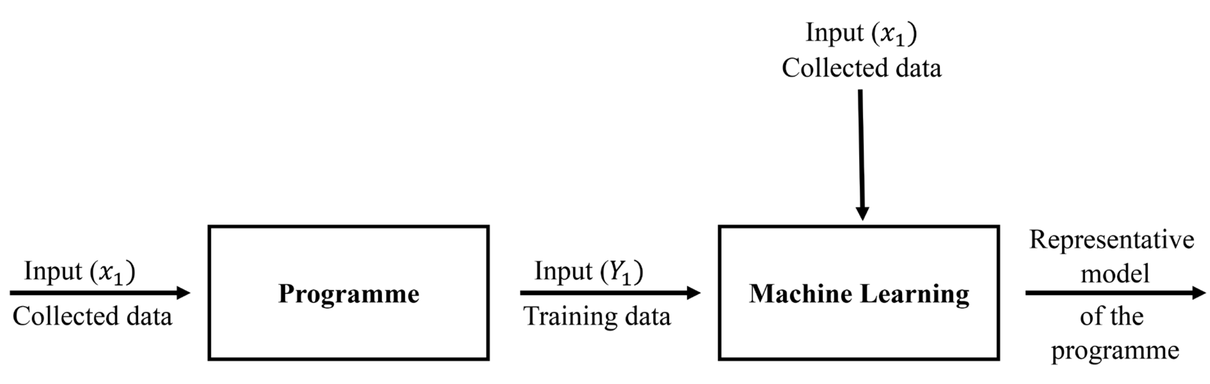
The great scale of data places new demands on organisations to quickly uncover hidden relationships and patterns. In this sense, data science – as a collection of techniques used to extract value from data that encompass artificial intelligence and machine learning – has proven to be extremely useful becoming an essential tool for any organisation that collects, stores, and processes data as part of its operations (Kotu & Deshpande, 2019).

Artificial intelligence (AI) emerges in this scenario, as a cognitive science with rich research activities in such areas as image processing, natural language processing, and robotics (Lee et al., 2018). The phenomenon of the implementation of AI in manufacturing, at the outset, generated "Intelligence Manufacturing" (IM) and following the evolution in manufacturing, IM evolved to "Smart Manufacturing" (SM) basically considering two pillars: the magnitude and impact of smart technologies such as the Internet of Things, Cloud Computing, Cyber-Physical Systems, and Big Data; the fusion of data and knowledge with society, physical space, and cyberspace (Yao et al., 2017). SM can be defined as the intensified application of advanced intelligence systems to enable rapid manufacturing of new products, dynamic response to product demand, and real-time optimisation of manufacturing production and supply-chain networks (SMLC, 2011; Terry et al., 2020).

The main difference between IM and SM is that IM is knowledge-based while SM is data-driven and knowledge-enabled (Yao et al., 2017). Therefore, features such as high efficiency, interoperability, virtualisation, decentralisation, real-time capability, and modularity – based on the main pillars of Industry 4.0, namely systems integration, big data analytics, augmented reality, cyber security, multi-agent systems, simulation, cloud computing, the internet of things and additive manufacturing – compounds smart manufacturing (Benotsmane et al., 2019).

Machine Learning (ML) can be considered a sub-field or one of the tools of artificial intelligence. It provides machines with the capability of learning from experience through data acquisition (Kotu & Deshpande, 2019). The main difference between artificial intelligence (AI) and machine learning (ML) is that AI is developed for decision making and ML is developed for learning new things based on data (Ozgulbas & Koyuncugil, 2019).

Figure 1. Representative model of a programme



Source: Author, adapted from Kotu and Deshpande (2019, p. 3).

Machine learning can be considered a subfield or one of the tools of artificial intelligence. It provides machines with the capability of learning from experience through data acquisition. This collected data used to teach machines is called training data. Consequently, machine learning becomes, a traditional programming model – a set of instructions to a computer - which transforms input signals (input collected data) into output signals (output training data) (Figure 1) using predetermined rules and relationships (Kotu & Deshpande, 2019).

Machine learning algorithms, also called "learners", take both the known input (input collected data) and output (output training data) to use as new entries for the learning process in order to ascertain a model for the programme which converts known input (input collected data) into final outputs (output trained data). Once the representative model is created, it can be used to predict the value of the interest rate, based on all the input variables (Kotu & Deshpande, 2019).

The science of ML is largely experimental because no universal learning algorithm exists. The computer cannot be enabled to learn every task it is given well. Any knowledge-acquisition algorithm needs to be tested with regards to learning tasks and data specific to the situation at hand (Bengio, 2016). However, it appears that digital technologies are entering a new era of complexity when massive data sets must be handled by machine learning and other advances meaning it is increasingly possible to break large problems down, by for instance cutting problems up into manageable challenges which can be modelled by computers. (Levy & Wong, 2014).

Considering unstructured and machine-generated data originating from the smart manufacturing processes and supervised learning the need for historical data is crucial and this learning approach is incapable of classifying new faults accurately, unsupervised learning approaches are initially more successful for dealing with the problems of the manufacturing industry sector (Wong et al., 2018).

Focusing on the learning process, Cielen et al., (2016) divide ML approaches into three types (supervised learning; unsupervised learning, and semi-supervised learning) –considering the human effort required to coordinate them and how they use labelled data: data with a category or a real-value number assigned to it that represents the outcome of previous observations. In a complementary way, Skilton and

Table 1. Machine learning approaches

Type	Description
Supervised Learning	The agent is trained by using examples from the problem space together with the desired output or action. Then the agent is provided with an input without the additional desired output and is required to make a prediction of the associated output, if the output differs from the desired output, then the agent is required to adapt (typically by being adjusted in some manner) so that it can produce the appropriate output for the given input.
Unsupervised Learning	These kinds of algorithms are used to train agents that are required to recognise similarities between inputs or identify features in the data provided. The agent is required to classify the provided data into clusters or segment the data into groups of similar entities.
Reinforcement Learning	These are algorithms that are used when the kind of training data is used in the case of supervised learning. Because this kind of learning fails to provide the same kind of error information that is commonly available with supervised learning, reinforcement learning often takes much longer and is less efficient when compared to supervised learning algorithms.
Model-based Learning	In these useful techniques, the agent builds a model that is a useful approximation to the training (input) data, or by constructing an explicit description of the target function. The main advantage this has over memory-based learning is its computational efficiency, and the efficient way memory is used, primarily because the agent can discard the training data once it has processed them.
Deep Learning	Its deals with problems involving image and voice. The specific kind of neural networks that were developed for these kinds of applications was called Deep Neural Networks, where each layer had the ability to recognise a set of features that would be used by the next layer in the network, moreover, the weights assigned to the connections between the nodes emphasised the importance of the feature.

Source: Skilton and Hovsepian (2018).

Hovsepian (2018), evidence that many paradigms have been utilised over the years in the field of ML and evince the following types of ML approaches (Table 1).

Machine knowledge refers to the knowledge contained in artificial intelligence. Human beings have now entered the four-dimensional society that comprises the natural world, the human world, the information world, and the intelligent-agent world (Li et al., 2019b).

Systematic review procedures

Database

The chosen database for this purpose was Scopus because: (1) it is the largest abstract and citation database of peer-reviewed literature; (2) it is compound of IEEE Xplore Library, ACM Digital Library, Willet Online Library, Engineering Village, and ScienceDirect – experienced engineering and computing databases; (3) it uniquely combines a comprehensive, curated abstract and citation database with enriched data and linked scholarly content.

Review design

Based on Biolchini et al. (2018) and Qiu et al. (2014) – the method of this systematic literature review was applied considering the following procedures: (a) Systematic literature review questions; (b) Search strategy; (c) Article selection; (d) Studies distribution; (e) Quality assessment; and (f) Data extraction. Only research articles (and original papers) in the final or in-press publication stage were analysed, once they were: disposed of in larger numbers; covering a greater diversity of themes; peer-reviewed; published more quickly emphasising their topicality.

Systematic review questions

The systematic review questions were classified into in two categories: general questions (GQ) and specific questions (SQ) (see Table 2). The GQ category is primarily guidelines to define the SQ and the variety of the proposed taxonomy. It tracks the main goals in the selected works and grants the opportunity to know general concerns and challenges involving the smart manufacturing industry. The SQ category is linked to the intersections between smart manufacturing and machine learning.

Search strategy

Focused on answering the proposed systematic review questions, keywords were selected to comprise a main search string split into research units and combined with Boolean operators considering acronyms, synonyms, and alternate spellings:

((“industry 4.0” OR “smart manufacturing” OR “smart factory” OR “industrial internet” OR “internet plus” OR “industry smartization”) AND

Table 2. Systematic Review Questions

Category	Question
General Systematic Review Questions (GQ)	GQ1: Which countries funds and which countries publishes research about machine learning in smart manufacturing?
	GQ2: Is it possible to design a taxonomy of machine learning techniques in a smart manufacturing scenario?
Specific Systematic Review Questions (SQ)	SQ1: In which manufacturing industries have machine learning techniques been used?
	SQ2: What are the main research concerns on the use of machine learning techniques in a smart manufacturing context?
	SQ3: What are the main technologies which associates machine learning to smart manufacturing?
	SQ4: What are the main machine learning techniques or algorithms applied in the context of smart manufacturing?
	SQ5: Which documents fuels research in the intersection between machine learning and smart manufacturing?

Table 3. Units of main search string

Identifier	Keywords	Description
Unit 1	("industry 4.0" or "smart manufacturing" or "smart factory" or "industrial internet" or "internet plus" or "industry smartization")	Establishes the fourth industrial revolution context.
Unit 2	("cyber-physical system" or "data science" or "artificial intelligence" or "machine learning")	Associates cyber physical systems (CPS) and artificial intelligence (AI) approaches to the referred context.
Unit 3	("technology" or "technique" or "algorithm" or "learner" or "learning system" or "industry" or "sector")	Depicts environments and solutions used in AI and CPS field.

Source: Authors, adapted from Skilton and Hovsepian (2018, p. 128).

("cyber-physical system" OR "data science" OR "artificial intelligence" OR "machine learning") AND ("technology" OR "technique" OR "algorithm" OR "learner" OR "learning system" OR "industry" OR "sector"))

In order to answer the questions in a modular way, the search string was split into three units, as detailed on Table 3.

Moreover, the option of the Scopus platform "search within" was initially configured to "all fields" and the period was not limited. Furthermore, preserving loyalty to the systematic review questions, we followed PIPOC (population, intervention, comparison, outcome, context) criteria based on Roberts and Petticrew (2006).

Article selection

After submitting the main search string (on March 30, 2021) 26,032 documents were found. This stage focused on discarding irrelevant studies. In this sense, we used the following exclusion criteria (EC): (1) search string on article title and keywords; (2) document type: articles and reviews; (3) Source type: journal; (4) Language: English; (5) terms must be in its title or keywords; (6) only available and non-duplicated documents. As result, the filtering process (see Figure 2) revealed 125 documents.

These 125 documents were, thereby selected as the filtered database to finally compose the results to be shown. Elsevier was the most representing publisher with 28,0% of analysed documents encompassing the following segments revealed: Elsevier Ltd; Elsevier B. V.; Elsevier Inc.; Elsevier USA. In the same way, the Institute of Electrical and Electronics Engineers (IEEE) represented 24,0% of the published documents encompassing the following segments: IEEE Computer Society; Institute of Electrical and Electronics Engineers Inc.

The analysed documents remaining after the filtering process were revealed in journals (according to exclusion criteria 3). The 125 filtered documents were published through 70 journals. The 9 journals which published the most regarding the theme are: Sensors (Switzerland) (13 articles); IEEE Access (11 articles); IEEE Transactions on Industrial Informatics (6 articles); Journal of Manufacturing Systems (4 articles); International Journal of Advanced Manufacturing Technology (4 articles); International Journal of Architectural Computing (3 articles); Applied

Figure 2. Filtering process

Documents revealed	EC1	EC2	EC3	EC4	EC5	EC6	EC7
• 26,032 (100%)	• Documents remained • 2,151 (8.26%)	• Documents remained • 842 (3.23%)	• Documents remained • 813 (3.12%)	• Documents remained • 786 (3.02%)	• Documents remained • 135 (0.52%)	• Documents remained • 127 (0.49%)	• Documents remained • 125 (0.48%)

Source: Research data (2021).

Sciences (Switzerland); IEEE Transactions on Automation Science and Engineering (3 articles); Computers Industry (3 articles). These publications correspond to 40% of all the filtered documents.

Distribution of studies

Although talking about artificial intelligence has been a topic of interest since the 1950s (Yao et al., 2017), machine learning has been applied notoriously in smart manufacturing only in the past five years. In 2016, the first year of publications (four studies in total) machine learning is spotted by concerns about processes. In this scenario, Cheng et al. (2016) revealed a process design proposing a platform to accomplish the goal of zero defects by applying the technology of Automatic Virtual Metrology. In addition, through a process improvement solution, Wang et al. (2016) proposed a large scale online multitask learning and decision-making system to achieve flexible manufacturing capabilities in a company which provided engine equipment for heavy trucks.

In the second year of publications, 2017, (also four studies) most of the publications were applied to the machinery manufacturing industry and concerns about maintenance appeared in two documents that year. The machinery industry and maintenance relationship, since then, has become very well-known which was verified in many other studies of this review, especially articles with tool wearing approaches, as Wu et al. (2017) demonstrated. The year of 2018 was marked by 10 publications and a new concern was revealed (cyber-security) when Settanni et al. (2018) analysed security challenges in a semiconductor manufacturing environment. In 2019, the number of publications duplicated (20), and other concerns emerged such as communication (Oyewobi et al., 2019) and business and finance (Yao et al., 2019).

In 2020, the number of publications greatly increased to 73 and other concerns arose such as Health, when Joloudari et al. (2020) proposed an integrated method using machine learning to increase the accuracy of coronary heart disease diagnostics; Sustainability, as Leong et al. (2020) proposed concerning a lean and green strategy in the processing sectors of a combined heat and power generation plant; Product design, as a case in point, Duc et al. (2020) presented a dynamic route-planning system for automated guided vehicles within a warehouse. Thus, as the years have passed, the diversity of machine learning applications has risen and the number of publications in this scenario has increased exponentially.

Quality Assessment

Aside from filtering documents and defining a clear dataset for the systematic review, we verified the quality of primary studies encompassed by this dataset, based on Qiu et al. (2014). So, the quality questions are listed on Table 4. Based on the 125 analysed articles, 107 of them obtained maximum scores, revealing positive answers to all the proposed quality questions.

Table 4. Quality Questions

Identifier	Description
QQ1	Is the document in the intersection of Machine Learning and Smart Manufacturing?
QQ2	Is there a clear statement about the aim of the document?
QQ3	Does the document contain a systematic review, a background, or a context?
QQ4	Does the document contribute in an effective way to the statistical analysis?
QQ5	Can the information obtained from the document be verified?
QQ6	Does the document answer all the generic systematic review questions?
QQ7	Does the document answer all the specific systematic review questions?

Source: Research data (2021).

Systematic Review Results

The following subsections expose the relevant contributions of the documents from the filtered database, that is, 107 articles. To do so, the useful information was underscored through the detailed answering of each systematic review question.

GQ1: Which countries funded, and which countries published research about machine learning in smart manufacturing?

The filtered database revealed that 4% of the analysed studies had private funding (partial or total). The identified private funding entities were: Nvidia Corporation (i.e., Maggipinto et al., 2018); Biesse Group and Accenture (in Calabrese et al., 2020); Nokia Bell-labs (in Pinho et al., 2020); and Electronic Components and Systems for European Leadership (in Settanni et al., 2018). The countries which funded the most (considering private and public resources) are detailed in Figure 3.

China funded 19 studies. Most of them in its own country. Only 2 articles were funded in other nations: one in the United Kingdom (Simeone et al., 2020); and the other in the United States (Ghahramani et al., 2020). The second greatest contributor which published the most studies is the European Union – with 11 studies – and the third, the United States with 8 articles.

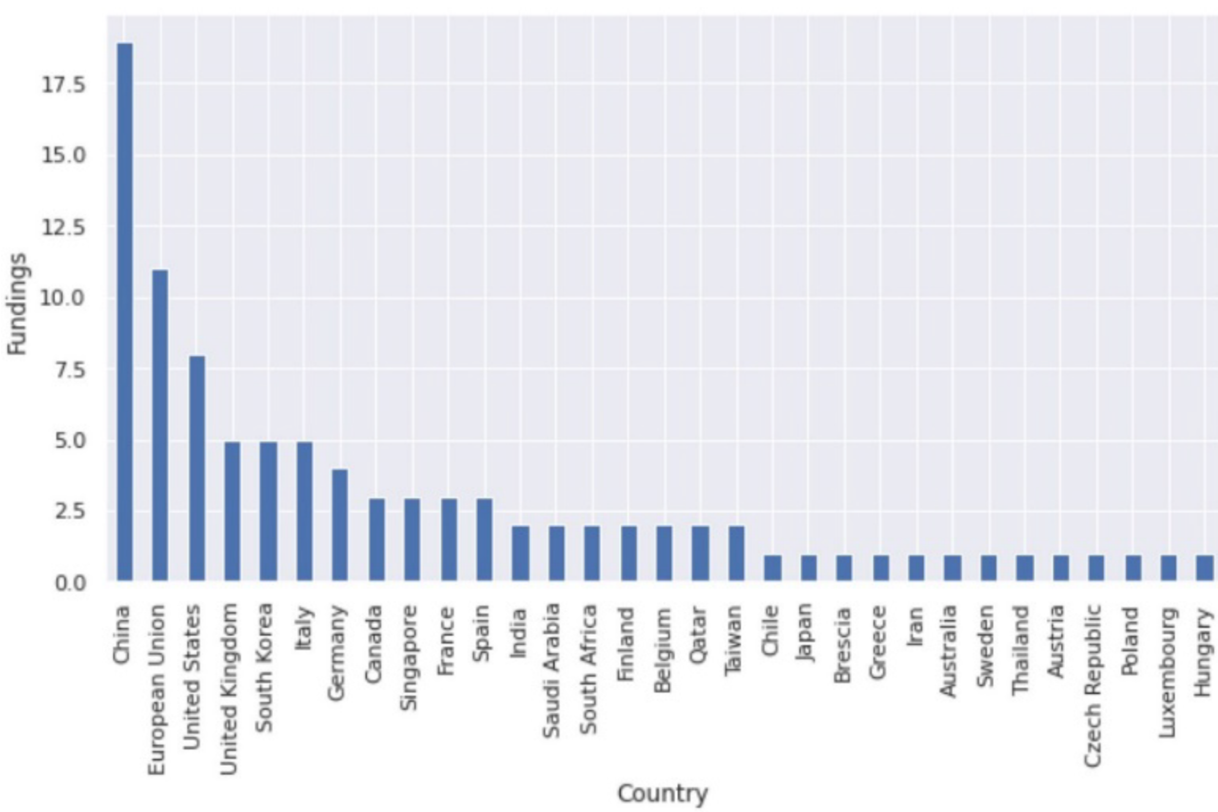
The countries which published the most – considering research developed with funding (private or public) and without funding – are detailed in Figure 4.

China also led this statistical result followed by the United States and Italy with 10 publications each. In this scenario, Latin America appeared with two publications, one in Chile and one in Brazil.

GQ2: Is it possible to design a taxonomy of machine learning techniques in a smart manufacturing scenario?

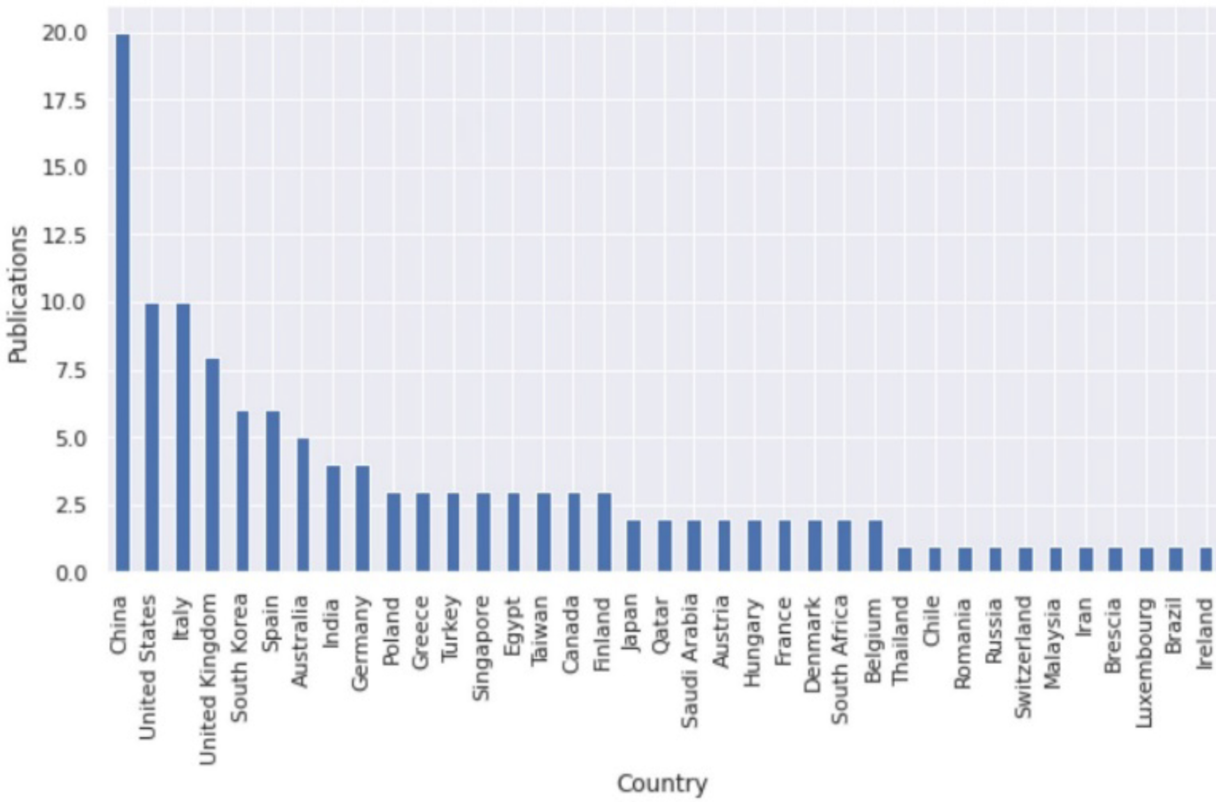
The taxonomy was designed based on the highlighted sources Çinar et al. (2020), Dalzo-

Figure 3. Countries which funded most research.



Source: Research data (2021).

Figure 4. Countries which published the most



Source: Research data (2021).

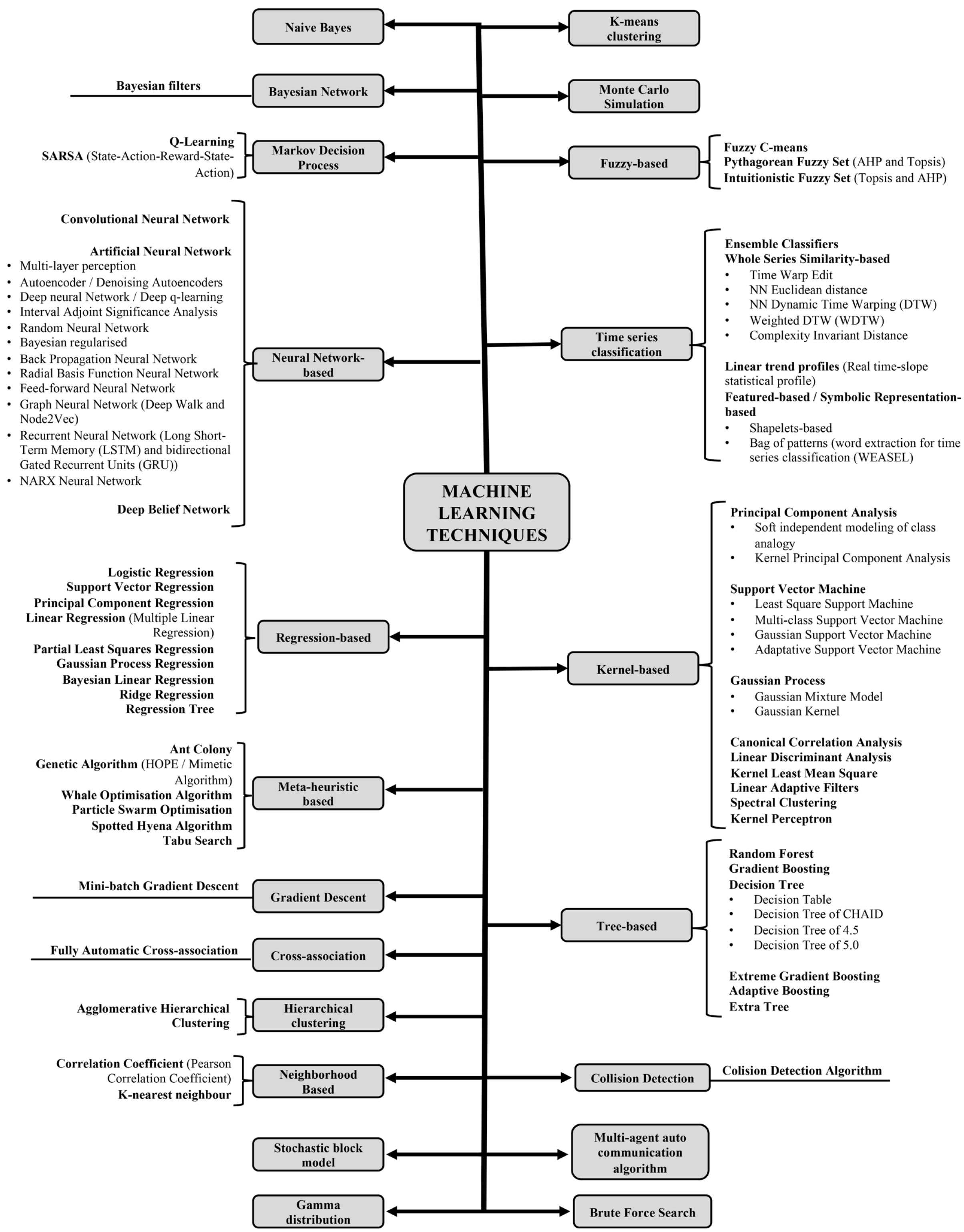
chio et al (2020), Géron (2017), Graves (2012), Wong et al. (2018), Wu et al. (2017) and Yu et al. (2019) due the fact that they contain the pillars of this taxonomy's structure. In addition, theories, and arguments postulated on the analysed documents of the filtered database, for instance Wu et al. (2017), also contributed to the structure, which is detailed in Figure 5.

In this sense, some comments are highlighted: 85 specific machine learning techniques were identified; artificial neural networks are at the very core of deep learning; Long-short term and bidirectional GRU cells are solutions used based on recurrent neural networks; Autoencoder is a type of artificial neural network (Aboelwafa et al., 2020), which is embraced by neural networks (Angelopoulos et al., 2020); Deep Neural Network is a type of artificial neural network comprising of many layers between the output and the input and supporting smart manufacturing by assisting in image and acoustic processing and thus product quality inspections, defect prognosis and fault assessments (Ratnayake et al., 2020); Wong et al. (2018) attests, recurrent neural network is an artificial neural network; Géron (2017) indicates that recurrent neural networks can be used as a deep learning technique with similar applications and capabilities as convolutional neural networks; the Kernel-based category was defined considering Géron (2017), and Luo et al. (2016). The neighbourhood-based category was defined considering Luo et al. (2016).

SQ1: In which manufacturing industries have machine learning techniques been used?

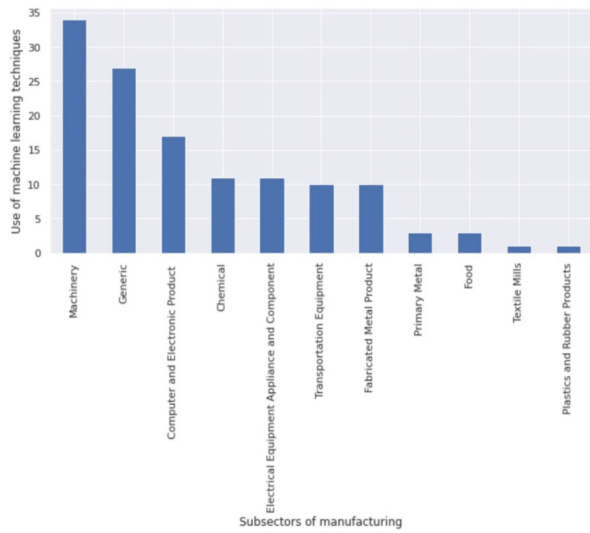
The classification of manufacturing industries was designed based on the North American Industry Classification System (NAICS), a joint system of classification of economic activities that makes the industrial statistics produced in the United States of America, Canada, and Mexico comparable (NAICS, 2017). The subsectors of manufacturing in which machine learning techniques were applied are presented in Figure 6.

Figure 5. Taxonomy of machine learning techniques in smart manufacturing



Source: Research data (2021).

Figure 6. Machine learning techniques applied in manufacturing subsectors



Source: Research data (2021).

The five most influential subsectors provide important information. In a chronological vision, considering the filtered database, Cheng et al. (2016) published the first document applying a back-propagation neural network in the **Machinery** subsector proposing a platform to accomplish the goal of zero defects in wheel machining through the technology of virtual metrology. This subsector is the one which most fueled the main concern when using machine learning techniques (process improvement) and was spotted by applications related to equipment – as presented by Roveda et al. (2020), who proposed a Model-Based Reinforcement Learning (MBRL) variable impedance controller to assist human operators in collaborative tasks; and process improvements – which Mamledesai et al. (2020) proposed as a model to teach the machines how a conforming component-producing tool appears and how a non-conforming component-producing tool appears. This provides more flexibility by including different quality requirements of the machine shops and defining whether the investigated tool produces a conforming or a non-conforming part – related to machines.

Machining processes were notoriously studied in the *Machinery* subsector, as demonstrated by Nasr et al. (2020) – by means of a system developed to obtain the optimal combination of machining parameters and the reinforcement ratio that led to minimising the feed force, depth force, and surface roughness in a machining process of graphene nanoplatelets.

Although the *Generic* subsector does not exist in the NAICS, it was created during the analysis of the filtered database to fit the studies which did not specify the subsector and had potential to be applied in diverse subsectors. Kim et al. (2020) represented this scenario by proposing a novel protocol fuzzing test technique that can be applied in a heterogeneous environment.

The third subsector, *Computer and Electronic Product Manufacturing*, brought together industries producing electronic products and their components. The manufacturers of computers, communications equipment, and semiconductors, for example, were grouped into the same subsector because of the inherent technological similarities of their production processes, and the likelihood that these technologies will continue to converge in the future (NAICS, 2017). Some studies were published in this scenario: Maggipinto et al. (2018) used deep learning techniques to automatically extract highly informative features based on the data of a real industrial dataset related to Etching – one of the most important semiconductor manufacturing processes – providing more accurate and scalable virtual metrology solutions; Schirru et al., (2018) focused on predictive maintenance, analysed the health factors of equipment wear also in semiconductor manufacturing with reference to etching equipment; The NAICS acknowledges the importance of these electronic industries, their rapid growth over the past several decades and the likelihood that these industries will, in the future, become even more important in the economies of the three North American countries (NAICS, 2017).

The *Chemical* subsector was shaped by studies which dealt with fibers. An important study, put forth by Khayyam et al. (2020), applied a hybrid machine learning algorithm for limited and big data modelling on carbon fiber production reducing the number of experiments and production costs.

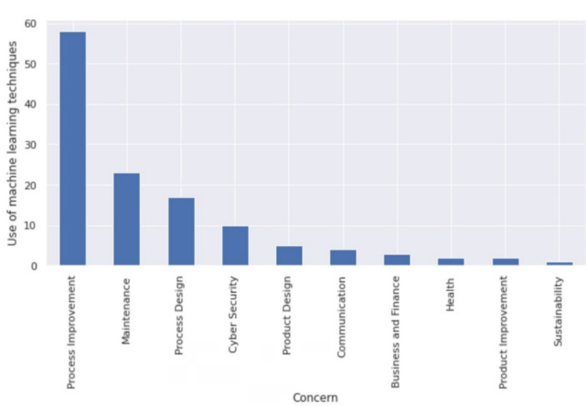
Another subsector that provided important information is *Electrical Equipment Appliance and Component* due to its relation to the Computer and Electronic Product subsector. It was represented by documents that compound three pillars: energy management, identified by a proposal put forth by Elsisi et al. (2021) using a deep learning and a IoT-based approach to control the operation of air conditioners in order to reduce energy consumption; power generation, represented by the proposed methodology of Ashraf et al. (2020) for implementing Industry 4.0 in a functional coal power plant focused on its efficiency; detection of equipment failure, outlined by the proposal of Cakir et al., (2021) of a condition monitoring system to detect faulty bearings in a universal motor shaft.

SQ2: What were the main concerns when using machine learning techniques in a smart manufacturing context?

The concerns that motivated the use of machine learning techniques can be summarised in categories as presented in Figure 7.

The *Process Improvement* category dealt with the efficiency of environments or industrial plants, and it encompasses: worker's skills and performance (Tao et al., 2019); production processes (Zhuang et al., 2021); and other kinds of processes, for example, the logistic process for

Figure 7. Concerns when using machine learning techniques



Source: Research data (2021).

using machine learning techniques to select the best green supplier (Çalik, 2021). This category also encompassed concerns about the ramp-up of smart manufacturing processes, as demonstrated by Doltsinis et al. (2020); the milling of innovative materials such as graphene (Nasr et al., 2020); and the machining of other important materials using, for example, tungsten (Wang et al., 2019).

The *Maintenance* category encompassed reactive, preventive, and proactive maintenance. In most of the cases it dealt with equipment and was linked to monitoring and visualisation (Moens et al., 2020), fault detections (Marino et al., 2021), reliability (Elsisi et al, 2020), tool wear predictions (Wu et al., 2017), and frameworks focused on better equipment performance– as set forth by Li et al. (2019a) who proposed a framework of cognitive maintenance providing technical solutions to real-time online maintenance tasks, avoiding outages due to equipment failures, and ensuring the continuous and healthy operation of equipment and manufacturing assets. This category also dealt with maintenance planning, as discussed in Upasani et al. (2017).

Process Design embraced, in general, proposals of new frameworks (Simeone et al., 2020) and architectures (Fischbach et al., 2020). Bakliwal et al. (2018), for instance, proposed a multi-agent system architecture to implement collaborative learning for the emerging concept of "social industrial assets". This category was also integrated by procedures. In this context, Jiang et al., (2021) proposed a procedure to predict the backend final test yield at the wafer fabrication stage itself in semiconductor manufacturing. Robotics was a well-known technology identified in this category, as denoted in Kaya et al. (2020).

Cyber Security encompassed cyber-attacks. In this scenario, Aboelwafa et al. (2020) introduced a novel method for false data injection attacks detection using the autoencoders machine learning technique. In this sense, Latif et al. (2020) proposed a novel attack detection scheme for IIoT using random neural network. In addition, Roveda (2020) proposed a solution that enforced privacy and trustworthiness in industrial IIoT data.

The *Product Design* category was related to the development of new products (machines and systems). In this sense, Reinhardt et al. (2020) introduced a study for carpentry task sequences (performed by collaborative robots) with the capture of computable actions. Also, Roveda et al. (2020) proposed a variable impedance controller to assist human operators in collaborative tasks. Further, Duc et al. (2020) developed a new system (a dynamic route-planning system) for automated guided vehicles within a warehouse.

Communication dealt with generic industries, diverse environment, and technologies. For example, machine-to-machine shopfloor and 5G technologies were illustrated in Messaoud et al., (2020) and – in an expanded way – wireless networks with cognitive radio were introduced in Ahmed et al. (2021).

The *Business and Finance* category executed documents developed also in the smart manufacturing context. For instance, Huo and Chaudhry (2021) evaluated the global expansion location decisions of Chinese manufacturing. In addition, Wang et al., (2020) proposed a smart customisation service to better address the semantic gap between customers (identifying their needs) and designers/manufacturers (contributing to the development, in the design stage, of customised solutions).

The *Health* category regarded the analysis of human risks and injuries as reported by: Pistolesi and Lazzerini (2020) who assessed the risk of low back pain and injury via inertial and barometric sensors; and Joloudari et al. (2020) who proposed an integrated method using machine learning to increase the accuracy of coronary heart disease diagnostics.

The *Product Improvement* category concerned making products better, and Elsisi et al. (2021) introduced a deep learning-based approach to control the operation of air conditioners in order to reduce its energy consumption in smart buildings. Finally, the *sustainability* category involved the organisations' strategy for improving its sustainability. This scenario was represented by the study put forth by Leong et al. (2020) which proposed a lean and green strategy in processing sectors of a combined heat and power generation plant.

SQ3: What were the main technologies which associated machine learning with smart manufacturing?

The identified technologies encompassed mostly software, systems, and equipment. The *Industrial Internet of Things* (IIot) is an intrinsic technology of Industry 4.0 and provides flexible, smart, and embracing solutions. It was the most used technology in the smart manufacturing context. It provided specific solutions, as the case of Vununu et al. (2018), who designed a system to detect faulty machine drills, to large and generic ones, and Luo et al. (2016), who proposed a large-scale web QoS prediction scheme for the industrial internet of things. In addition, this technology was identified in 8 of the 11 manufacturing subsectors.

Sensor technologies are crucial to achieving a closed-loop monitoring process and to the development of new solutions and to improve processes by providing accurate and reliable measurements. The analysis of the selected documents revealed diverse kinds of sensors such as: wearable barometric sensors (Pistolesi & Lazzerini, 2020); optical sensors (Simeone et al., 2020); soft sensors (Zheng et al., 2018); and ultrasonic sensors (Bowler et al., 2020). The sensors identified were applied in diverse manufacturing subsectors, such as generic, food, chemical, machinery, computer and electronic product, transportation equipment, and health.

Software technologies – besides encompassing edge, fog, and parallel computing – were supported by the following software: Python (Kazi et al., 2021), MATLAB (Osswald et al., 2020), LabView (Mishra et al., 2020), Spark platform (Calabrese et al., 2020, Wang et al., 2016), Google AutoML Platform (Google AutoML) (Sader et al., 2020), LinkSmart open source IoT platform (Soto et al., 2019), Amazon Web Services Machine Learning (Amazon ML) (Caesarendra et al., 2018).

Digital Twin, an intrinsic system of Industry 4.0 concept, was used in several articles such as Xia et al. (2021) and Ghosh et al. (2020). Image processing was another field of application of machine learning techniques. In the analysed documents it was strongly used for process improvements in diverse subsectors such as food

Multidiscip. Bus. Rev. | Vol. 16, N° 2, 2024, pp. 66-88, ISSN 0718-400X | DOI: https://doi.org/10.35692/07183992.16.2.6

(Simeone et al., 2020), machinery (Tannous et al., 2020) and computer and electronic product (Tsutsui & Matsuzawa, 2019).

Equipment technologies were frequently related to machines. One of the most associated equipment to technology was Robot. In this sense, Nicholas et al. (2020) presented a robust workflow for robotic 3D printing onto unknown and arbitrarily shaped 3D substrates. Programmable Logic Controller (PLC), a modular and powerful controller – as one of the most important equipment of automation systems – was used in diverse industries, such as the machinery subsector for maintenance (Yan et al., 2018) and the generic subsector for process improvement (Doltsinis et al, 2020). With the same characteristic of being useful, computer numerical control (CNC) was strongly used since the first publications in the machinery subsector (especially for machining). Over the years, it was also related to other subsectors and concerns as Kumar and Misra (2021) demonstrated.

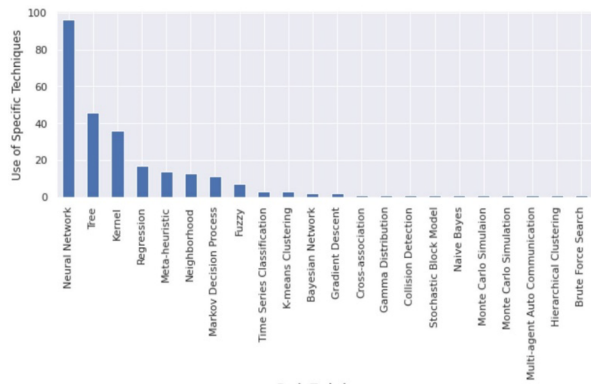
SQ4: What were the main machine learning techniques applied in the context of smart manufacturing?

Many different specific machine learning techniques applied in the smart manufacturing context were identified. The filtered database revealed the use of specific techniques in 292 situations. The three most expressive basic techniques (see Figure 8) are:

- **neural network-based techniques**: supported by the artificial neural networks category, were used in 96 situations representing 36,92% of all situations identified.
- **tree-based techniques**: supported by random forest specific techniques, were applied in 46 situations representing 17,69%.
- **kernel-based techniques**: supported by support vector machine specific technique, were revealed in 36 situations representing 13,85%.

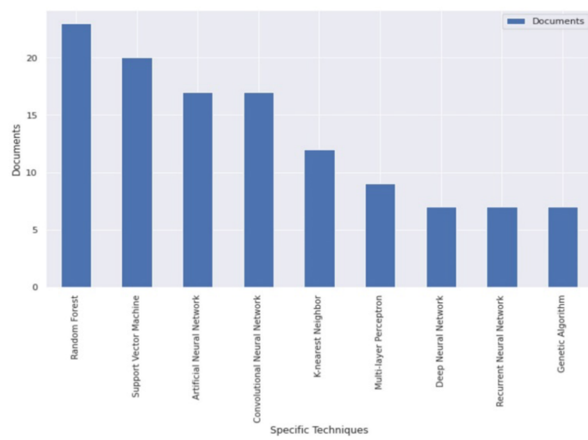
In addition, during the analysis of the filtered database, 85 specific techniques were identified. The nine most used of them are depicted in Figure 9.

Figure 8. Basic machine learning techniques used



Source: Research data (2021)

Figure 9. Most specific machine learning techniques



Source: Research data (2021)

The four most used techniques are described as follows. *Random forest* (used in 23 documents) is an ensemble learning algorithm developed by Breiman (2001). It is a specific tree-based technique and was made up of several decision tree classifiers. Its output category is determined collectively by these individual trees. The benefits of the use of random forest are: (1) it can manage high-dimensional data without choosing a feature; (2) trees are independent of each other during the training process; and (3) its implementation is simple. However, the training speed is generally fast, and, at the same time, the generalisation functionality is good enough (Çinar et al., 2020).

Support vector machine (used in 20 documents), the best-known kernel machine learning method and one of the most popular models in machine learning, it is a powerful and versatile technique capable of performing linear or nonlinear classification, regression, and even outlier detection. It is particularly well suited for the classification of complex but small or medium-sized datasets (Géron, 2017). In addition, it is widely applied in the predictive maintenance of industrial equipment, as demonstrated in some of the analysed studies, i.e., Çinar et al. (2020) and Vununu et al. (2018).

An *artificial neural network* model (used in 17 documents) – which takes its motivation from the human nervous system – is a parallel system that can resolve paradigms that linear computing cannot resolve. It is an adaptive system, i.e., parameters can be changed during operation (training) to suit the problem. In addition, it can be used in a wide variety of classification tasks, e.g., character recognition, speech recognition, fraud detection, and medical diagnosis (Dougherty, 2013).

As Li et al. (2019a) attest, artificial neural network algorithms refer to a wide class of intelligent computation and machine learning techniques with structures of many layers to process nonlinear information. In this context, a *convolutional neural network* (used in 17 documents) is a kind of hierarchical multi-layer model with a strong capacity to discover knowledge in big data, especially for image-based data, because vision is highly hierarchically organised (Cheng et al., 2018).

In some predictive maintenance scenarios, the information or signs of failure can also be perceived based on data in a two-dimensional format, for example, pictures or a frequency spectrum. In cases related to equipment (quite common in the analysed documents of the filtered database), filters in convolutional layers can extract local features from raw data and further build complex patterns for machine health monitoring by stacking these convolutional layers, which makes the convolutional neural network an ideal tool when the target is image-based data (Zhao et al., 2019).

SQ5: Which documents fomented research in the intersection between machine learning and smart manufacturing?

The two most cited studies were published in 2017 by the same author (Wu et al., 2017) a researcher of the Department of Industrial and Manufacturing Engineering, at Pennsylvania State University, United States. The first one, which had 157 citations, introduced a random forest-based prognostic method for tool wear prediction and compared its performance to two other machine learning techniques. This study was developed in the machinery subsector of manufacturing (the environment which was the most used to apply machine learning techniques) and focused on the remaining useful life of components (a common concern identified in many studies of the filtered database), especially on milling tasks. In addition, this document contributed to the most specific machine learning technique used (random forest).

The second one, with 127 citations, also was focused on prognostics and was applied to the machinery subsector. It introduced a computational framework to enable manufacturers to monitor machine health conditions and generate predictive analytics. This study also used the aforementioned machine learning technique (random forest) to implement the proposed solution.

Main Findings and Discussions

There is no unanimity among authors for categorising methods of machine learning. Khayyam et al. (2020) attest that there are four of them: supervised learning, unsupervised learning, reinforcement learning and semi-supervised learning. Angelopoulos et al. (2019) also affirm that there are four methods, but they are not the same: supervised learning, unsupervised learning, reinforcement learning and deep learning.

In the same way, there is no harmony in categorising the techniques of machine learning. For example, Dalzochio et al. (2020) divide “deep learning” and “neural networks” as two different classes of machine learning techniques and classify “recurrent neural networks” as part of “neural networks”. On the other hand, Abidi et al. (2020) and Khayyam et al. (2020) define recurrent neural networks as part of "deep learning".

Tree-based techniques were standard models and widely used as machine learning. As Calabrese et al. (2020) attest, one of the tree-based techniques key advantages is the fact that they are easily interpreted since the predictions are revealed as a set of rules. Additionally, they demand less computer effort for the training phase. They also consider the gradient concept trend and techniques such as gradient boosting and extreme gradient boosting can achieve higher performance – in terms of accuracy, precision, and recall –than, for instance, support vector machine (the second most used machine learning technique). In such an environment of easy interpretations, the k-nearest neighbor – one of the most used techniques in the machine learning context – has its usage reinforced by the following advantages: simple to understand; easy to implement and debug; it does not explicitly build any training model – it simply predicts the novel data based on the most similar historical data (Romeo et al., 2020). Thereby, neighbourhood-based techniques were widely used strategies to deal with web quality of service (QoS) predicting its missing entries as demonstrated, for example, in Luo et al. (2016).

Most fault detection machine learning applications were focused on predictive maintenance, specifically in mechanical equipment (bearings, machines in general and assembly lines). In addition, tool wear prediction was widely concerned. Moreover, the use of deep learning techniques in images processing tasks is increasing noticeably and the transfer learning method is an effective step towards qualitative and quantitative microstructure interpretation of images (especially steel) in metallurgical manufacturing, as Choudhury (2020) argued.

From a production perspective, based on Diez-Olivan et al. (2019), three levels of implementation of technologies in the Industry 4.0 context were identified:

1. **Vertical integration**: allied to automation challenges, this concept refers to the integration of diverse information and communication technology (ICT) systems into different hierarchical levels, from the bottom ones – e.g., smart sensors as shown in Mulrennan et al. (2018), actuators as shown in Nasr et al. (2020) – to the highest ones – e. g., execution and planning as shown in Alexopoulos et al., 2020).
2. **Horizontal integration**: it deals with the integration of ICT technologies into mechanisms and agents involved in different stages of manufacturing processes and business planning, that is, exchanging energy and information within a company – e.g., logistic, production and commercialisation, as depicted in Yao et al. (2019) – or among companies – value-added networks – as described in Çalik (2021).
3. **Circular integration**: it refers to a mix of vertical and horizontal integrations to link the end user to the product life cycle, as shown in Wang et al (2020).

Still, in the field of production, machine learning techniques and the technologies associated with it (such as numerical and graphical modelling, simulations, 3D design, and software related to it) are applied to design industrial robot trajectories – especially robot joints, as depicted in Azizi (2020) – contribute directly to the emergence of collaborating robots and, consequently, to economic and social development (Benotsmane et al., 2019).

Conclusion and Research Boundaries

This paper was designed to analyse the use of machine learning in smart manufacturing, describing techniques, technologies, industries, and purposes associated with industrial applications. To accomplish this, we conducted a systematic literature review in Scopus, and we analysed 107 articles. Machinery was the industry subsector with the most implementations of machine learning; process improvement was the main concern (interest) of all implementations; random forest was the most specific machine learning technique used; and diverse technologies associated with this context were identified such as the industrial internet of things, digital twin, sensor technologies (soft, optical, barometric, ultrasonic), software technologies (Python, MATLAB, LabView, Google AutoML Platform) and equipment technologies (robotic, PLC, CNC). We designed a taxonomy pointing out 85 specific machine-learning techniques used in smart manufacturing.

Data has become a critical input for decision making and production processes in several industries. Data science is providing a profound impact on the way industrial organisations design, manufacture and monitor their assets, reducing process downtime and increasing product quality. Specifically in the smart manufacturing field – a complex environment shaped by Cyber Physical Systems, intelligent agents, and advanced ICTs – there is no unanimity among authors in categorising types of machine learning and no consensus regarding a taxonomy of machine learning techniques used in the smart manufacturing context.

Technologies associated to Industry 4.0 and advancements in artificial intelligence are fueling the development of machine learning techniques making them wildly applicable in diverse engineering fields. It is molding a new era of complexity where massive data sets must be handled by ML techniques and computers in order to solve complex demands, manage challenges, and provide new solutions.

Systems developed with the machine learning concept – due to the fact of being based on specific techniques – are extremely flexible and feasible in terms of being applied and integrated into diverse tools as programming platforms – such as Python and Matrix Laboratory (MATLAB), maintenance planning systems – such as Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MES) – and equipment – such as robots and Computer Numerical Control (CNC). Consequently, machine learning can be applied in a vast field of environment, equipment, processes, and systems and it has been designing new solutions for the smart manufacturing context.

Algorithms based on machine learning can be used to predict a system's behaviour and improve its overall performance, enabling the development of tools capable of analysing data and perceiving the existing underlying trends and correlations. However, the most appropriate machine learning technique to be chosen must consider the available amount of data, the way it is disposed, its shape, in which technology it will be incorporated and, finally, the main purpose of data manipulation.

It is noteworthy that to achieve specific industry needs, the combination of machine learning and mathematical techniques immerged hybrid, adapted and modified solutions with higher potential for identifying complex system characteristics and solving its complex demands. In the smart manufacturing field, the machinery subsector was the most demanded environment for the use of machine learning and process improvement was the major concern. In that same context, China was identified as the country which promoted research the most and Elsevier was the most representative publisher of research. A detailed taxonomy revealing the use of 85 specific machine learning techniques in smart manufacturing was established, and associated to it, technologies involving smart sensors, robotic, robust software, equipment and systems of automation, semiconductors and virtualisation were identified.

Despite the existence of other approaches aligned with Industry 4.0 – such as Industrie Du Futur (in France), Industria Conectada (in Spain), ASIAN 4.0 (Southeast Asian Nations) and Society 5.0 (in Japan) – with considerable research in the field of this study, the fact is the string search contains only terms in the English language and excludes studies published in the official language of these countries and regions. It is a limiting factor.

Also, there are several further study recommendations for Machine Learning in the manufacturing context. Some of these are:

- Robustness and Reliability: One of the primary challenges in applying Machine Learning to manufacturing is ensuring the robustness and reliability of the models. Future research could explore methods to improve the robustness and reliability of Machine Learning models in manufacturing by accounting for uncertainties and variations in the manufacturing process.
- Explainability and Transparency: Another key challenge in applying Machine Learning

to manufacturing is the need for explainability and transparency. To gain the trust of stakeholders and ensure that the models are used effectively, future research could explore methods to improve the explainability and transparency of Machine Learning models in manufacturing.

- Data Quality and Data Collection: Data quality and data collection are essential for Machine Learning models in manufacturing. Future research could explore methods to improve data collection and ensure high quality data.
- Integration with Existing Systems: Machine Learning models need to be integrated with existing manufacturing systems to be effective. Future research could explore methods to integrate Machine Learning models with existing manufacturing systems seamlessly.
- Human-Machine Interaction: Machine Learning models can greatly benefit from human input and feedback. Future research could explore methods to improve human-machine interaction and make it easier for human operators to work with Machine Learning models in manufacturing.
- Scalability: Machine Learning models need to be scalable to be effective in a manufacturing context. Future research could explore methods to improve the scalability of Machine Learning models in manufacturing and make them more accessible to a wider range of companies and applications.
- Cybersecurity: As Machine Learning models in manufacturing become more prevalent, cybersecurity risks will become a more significant concern. Future research could explore methods to improve the cybersecurity of Machine Learning models in manufacturing and reduce the risk of cyber-attacks.

Overall, these are some of the main further study recommendations for Machine Learning in manufacturing, and continued research in these areas can help ensure that Machine Learning is effectively applied to manufacturing and provides real-world benefits to manufacturers. Also, we would like to recommend a methodological approach for further studies in this area, since there is a gap due to using meta-analysis as a statistical technique – especially because we found some conflicting results, and meta-analysis can help to reconcile conflicting results from different studies by providing a quantitative estimate of the effect size across studies.

While it is true that the document type entitled “review” provides a holistic vision and insights into a theme, Nevertheless, the documents found with the type “review” were eliminated when applying the exclusion criteria 5 – and it is another limiting factor of this systematic literature review – due to the fact that reviews could mention, and consequently duplicate, existing articles on the dataset. Nonetheless, documents classified as “review” were appreciated to increase the arguments of this research.

References

Abidi, M. H. et al. (2020). Automated Maintenance Data Classification Using Recurrent Neural Network: Enhancement by Spotted Hyena-Based Whale Optimization. *Mathematics*, 8, 11, 2020. https://doi.org/10.3390/math8112008

Aboelwafa, M. M. N. et al. (2020). A Machine-Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT. *IEEE Internet of Things Journal*, 7, 9, 8462-8471. https://doi.org/10.1109/JIOT.2020.2991693

Ahmed, R. et al. (2021). CR-IoTNet: Machine Learning Based Joint Spectrum Sensing And Allocation For Cognitive Radio Enabled Iot Cellular Networks. *Ad Hoc Networks*, 112 https://doi.org/10.1016/j.adhoc.2a\\\\\\\\020.102390.

Angelopoulos, A., Michailidis, E. T., Nomikos, N., Trakadas, P., Hatziefremidis, A., Voliotis, S., & Zahariadis, T. (2019). Tackling Faults in the Industry 4.0 Era—A Survey of Machine-Learning Solutions and Key Aspects. Sensors, 20(1), 109. MDPI AG. Retrieved from http://dx.doi.org/10.3390/s20010109

Alexopoulos, K., Nikolakis, N. & Chryssolouris, G. (2020). Digital Twin-Driven Supervised Machine Learning for the Development of Artificial Intelligence Applications in Manufacturing. *International Journal of Computer Integrated Manufacturing*, 33(5), 429-439 https://doi.org/10.1080/0951192X.2020.1747642.

Ashraf, W. M. et al. (2020). Optimization of a 660 MWe Supercritical Power Plant Performance - A Case of Industry 4.0 in the Data-Driven Operational Management Part 1 - Thermal Efficiency. *Energies*, 13(21), 1-33. https://doi.org/10.3390/en13215592

Azizi, A. (2020). Applications of Artificial Intelligence Techniques to Enhance Sustainability of Industry 4.0: Design of an Artificial Neural Network Model as Dynamic Behavior Optimizer of Robotic Arms. *Complexity*, id 8564140. https://doi.org/10.1155/2020/8564140

Bakliwal, K. et al. (2018). A Multi Agent System Architecture to Implement Collaborative Learning for Social Industrial Assets. *IFAC - Papers OnLine*, 51(11), 1237-1242, 2018. https://doi.org/10.1016/j.ifacol.2018.08.421.

Benotsmane, R., Kovács, G., & Dudás, L. (2019). Economic, Social Impacts and Operation of Smart Factories in Industry 4.0 Focusing on Simulation and Artificial Intelligence of Collaborating Robots. *Social Scenes*, 8, 5, 143-163. https://doi.org/10.3390/socsci8050143

Bengio, Y. (2016). Machines Who Learn. *Scientific American*, 314, 6, 46-51

Bickel, W. K., Tomlinson, D. C., Craft, W. H., Ma, M., Dwyer, C. L., Yeh, Y.-H., Tegge, A. N., Freitas-Lemos, R., & Athamneh, L. N. (2023). Predictors of smoking cessation outcomes identified by machine learning: A systematic review. *Addiction Neuroscience*, 6(January), 100068. https://doi.org/10.1016/j.addicn.2023.100068

Biolchini, J. C. A. et al. (2018). Scientific Research Ontology to Support Systematic Review in Software Engineering. *Advanced Engineering Informatics*, 21, 133-151 https://doi.org/10.1016/j.aei.2006.11.006

Bowler, A. L., Bakalis, S., & Watson, N.J.(2020). Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning. *Sensors*, v. 20, i. 7, p. 1813, 2020. https://doi.org/10.3390/s20071813

Breiman, L. (2001). Random Forests, *Machine Learning*, 45, 5-32 https://doi.org/10.1023/A:1010933404324

Caesarendra, W. et al. (2018). An AWS Machine Learning-Based Indirect Monitoring Method for Deburring in Aerospace Industries Towards Industry 4.0, *Applied Sciences*, 8, 11. https://doi.org/10.3390/app8112165

Cakir, M., Guvenc, M.A., & Mistikoglu, S. (2021). The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system, *Computers & Industrial Engineering*, 151. https://doi.org/10.1016/j.cie.2020.106948

Çalik, A. (2021). A novel Pythagorean fuzzy AHP and fuzzy TOPSIS methodology for green supplier selection in the industry 4.0 era. *Soft Computing*, 25, 2253-2265. https://doi.org/10.1007/s00500-020-05294-9

Calabrese, M. et al. (2020). SOPHIA: An Event-Based IoT and Machine Learning Architecture for Predictive Maintenance in Industry 4.0, *Information*, 11, 4, 202-219. https://doi.org/10.3390/info11040202. Çalik, A. (2021). A novel Pythagorean fuzzy AHP and fuzzy TOPSIS methodology for green supplier selection in the industry 4.0 era. *Soft Computing*, 25, 2253-2265. https://doi.org/10.1007/s00500-020-05294-9

Cielen, D., Meysman, A. D. B., & Ali, M. (2016). *Introducing Data Science: Big data, machine learning, and more, using python tools*. Manning: Shelter Island.

Çinar, Z.M. et al. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0, *Sustainability*, 12, 9, 1-42. https://doi.org/10.3390/su12198211

Ciuriak, D. (2018). The economics of data: Implications for the data-driven economy. In Data governance in the digital age, *Proceedings*... Waterloo, 2018.

Cheng, F. et al. (2016). Industry 4.1 for Wheel Machining Automation. *IEEE Robotics and Automation Letters*, 1, 1, 332-339. https://doi.org/10.1109/LRA.2016.2517208

Cheng, J. et al. (2018). Recent advances in efficient computation of deep convolutional neural networks, *Frontiers of Information Technology & Electronic Engineering*, 19, 64-77.

Choudhury, A. (2020). The Role of Machine Learning Algorithms in Materials Science: A State of Art Review on Industry 4.0. *Archives of Computational Methods in Engineering*, 28, 3361 - 3381. https://doi.org/10.1007/s11831-020-09503-4

Dalzochio, J. et al. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges, *Computers in Industry*, 123. https://doi.org/10.1016/j.compind.2020.103298

Diez-Olivan, A., Del Ser, J., Galar, D., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion*, 50(July 2018), 92–111. https://doi.org/10.1016/j.inffus.2018.10.005

Doltsinis, S. et al. (2020). A Decision Support System for rapid ramp-up of industry 4.0 enabled production systems, *Computers in Industry*, 116. https://doi.org/10.1016/j.compind.2020.103190

Dougherty, G. (2013). *Pattern Recognition and Classification: An introduction*. 1. ed. Camarillo: Springer

Duc, N. D., Tran Huu, T., & Nananukul, N. (2020). A Dynamic Route-Planning System Based on Industry 4.0 Technology. *Algorithms*, 13(12), 308. http://dx.doi.org/10.3390/a13120308

Elsisi, M. et al. (2021). Deep Learning-Based Industry 4.0 and Internet of Things towards Effective Energy Management for Smart Buildings, *Sensors*, 21, 4. https://doi.org/10.3390/s21041038

Fischbach, A. et al. (2020). CAAI – a cognitive architecture to introduce artificial intelligence in cyber-physical production systems. *The International Journal of Advanced Manufacturing Technology*, 111, 609-626. https://doi.org/10.1007/s00170-020-06094-z

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. 1. ed. Sebastopol: O'Reilly Media.

Ghahramani, M. et al. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. *IEEE/CAA Journal of Automatica Sinica*, 7, 4, 1026-1037. https://doi.org/10.1109/JAS.2020.1003114

Ghosh, A. K. et al. (2020). Machining Phenomenon Twin Construction for Industry 4.0: A Case of Surface Roughness. *Journal of Manufacturing and Materials Processing*, 4, 1. https://doi.org/10.3390/jmmp4010011

Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*. 1 ed. Berlin: Springer-Verlag

Hofmann, T., Scholkopf, B., & Smola, A.J. (2008). Kernel Methods in Machine Learning. *The Annals of Statistics*, 36, 3, 1171-1220. https://doi.org/10.1214/009053607000000677

Huo, D., & Chaudhry, H. R. (2021). Using machine learning for evaluating global expansion location decisions: An analysis of Chinese manufacturing sector. *Technological Forecasting and Social Change*, 163. https://doi.org/10.1016/j.techfore.2020.120436

Jiang, D., Lin, W., & Raghavan, N. (2021). A Gaussian Mixture Model Clustering Ensemble Regressor for Semiconductor Manufacturing Final Test Yield Prediction. *IEEE Access*, 9, 22253-22263. https://doi.org/10.1109/ACCESS.2021.3055433

Joloudari, J. H., Joloudari, E. H., Saadatfar, H., GhasemiGol, M., Razavi, S. M., Mosavi, A., Nabipour, N., Shamshirband, S., & Nadai, L. (2020). Coronary Artery Disease Diagnosis; Ranking the Significant Features Using a Random Trees Model. *International journal of environmental research and public health*, 17(3), 731. https://doi.org/10.3390/ijerph17030731

Kaya, I. et al. (2020). Creating a road map for industry 4.0 by using an integrated fuzzy multicriteria decision-making methodology, *Soft Computing*, 24, 17931-17956. https://doi.org/10.1007/s00500-020-05041-0

Kazi, M., Eljack, F., & Mahdi, E. (2021). Data-driven modeling to predict the load vs. displacement curves of targeted composite materials for industry 4.0 and smart manufacturing, *Composite Structures*, 258. https://doi.org/10.1016/j.compstruct.2020.113207

Khayyam, H. et al. (2020). A Novel Hybrid Machine Learning Algorithm for Limited and Big Data Modeling with Application in Industry 4.0. *IEEE Access*, 8, 111381-111393. https://doi.org/10.1109/ACCESS.2020.2999898

Kim, S. J. et al. (2020). Smart seed selection-based effective black box fuzzing for IIoT protocol, *The Journal of Supercomputing*, 76, 12, 10140-10154.

Kotu, V., & Deshpande, B. (2019). *Pattern Data Science: Concepts and practice*. 2. ed. Chennai: Morgan Kaufmann

Kumar, P., & Misra, J.P. (2021). Process modeling and optimization using ANN and RSM during dry turning of titanium alloy used in automotive industry, *Journal of Automobile Engineering*, 235, 7. https://doi.org/10.1177/0954407020969255

Latif, S., Zou, Z., Idrees, Z., & Ahmad, J. (2020). A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network. *IEEE Access*, 8, 89337-89350. https://doi.org/10.1109/access.2020.2994079

Lee, J. et al. (2018). Industrial artificial intelligence for industry 4.0-based manufacturing systems, *Manufacturing Letters*, 18, 20, 20-23 https://doi.org/10.1016/j.mfglet.2018.09.002

Leong, W. D., Teng, S. Y., How, B. S., Ngan, S. L., Rahman, A. A., Tan, C. P., Ponnambalam, S. G., & Lam, H. L. (2020). Enhancing the adaptability: Lean and green strategy towards the Industry Revolution 4.0. *Journal of Cleaner Production*, 273, 122870. https://doi.org/10.1016/j.jclepro.2020.122870

Levy, C., & Wong, D. (2014). *Towards a smart society*. Report, Big Innovation Centre

Li, B. et al. (2019a). Framework and case study of cognitive maintenance in Industry 4.0, *Frontiers of Information Technology and Electronic Engineering*, 20, 1493-1504. https://doi.org/10.1631/FITEE.1900193

Li, Z., Liu, R., & Wu, D. (2019b). Data-driven smart manufacturing: Tool wear monitoring with audio signals and machine learning, *Journal of Manufacturing Processes*, 48, 66-76. https://doi.org/10.1016/j.jmapro.2019.10.020

Liu, Y., Esan, C.O, Pan, Z., & An, L. (2021). Machine learning for advanced energy materials. *Energy and AI*, 3, 100049. https://doi.org/10.1016/j.egyai.2021.100049

Luo, X. et al. (2016). A large-scale web QoS prediction scheme for the Industrial Internet of Things based on a kernel machine learning algorithm, *Computer Networks*, 101, 81-89. https://doi.org/10.1016/j.comnet.2016.01.004

Magar, R., & Farimani, A. B. (2023). Learning from mistakes: Sampling strategies to efficiently train machine learning models for material property prediction. *Computational Materials Science*, 224 (February), 112167. https://doi.org/10.1016/j.commatsci.2023.112167

Maggipinto, M. et al. (2018). A Computer Vision-Inspired Deep Learning Architecture for Virtual Metrology Modeling With 2-Dimensional Data, *IEEE Transactions on Semiconductor Manufacturing*, 31, 3, 376–384. https://doi.org/10.1109/TSM.2018.2849206

Mamledesai, H., & Soriano, M. A. (2020). A Qualitative Tool Condition Monitoring Framework Using Convolution Neural Network and Transfer Learning, *Applied Sciences*, 10, 20, 7298 https://doi.org/10.3390/app10207298

Marino, R. et al. (2021). A Machine-Learning-Based Distributed System for Fault Diagnosis with Scalable Detection Quality in Industrial IoT, *IEEE Internet of Things Journal*, 8, 6, 4339-4352. https://doi.org/10.1109/JIOT.2020.3026211

Messaoud, S., Bradai, A., & Moulay, E. (2020). Online GMM Clustering and Mini-Batch Gradient Descent Based Optimization for Industrial IoT 4.0, IEEE Transactions on Industrial Informatics, 16, 2, 1427-1435. https://doi.org/10.1109/TII.2019.2945012

Mintarya, L. N., Halim, J. N. M., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Computer Science*, 216, 96–102. https://doi.org/10.1016/j.procs.2022.12.115

Mishra, D. et al. (2020). Real time monitoring and control of friction stir welding process using multiple sensors, *Cirp Journal of Manufacturing Science and Technology*, 30, 1-11. https://doi.org/10.1016/j.cirpj.2020.03.004

Moens, P. et al. (2020). Scalable Fleet Monitoring and Visualization for Smart Machine Maintenance and Industrial IoT Applications, *Sensors*, 20, 15, 1-15. https://doi.org/10.3390/s20154308

Mulrennan, K. et al. (2018). A soft sensor for prediction of mechanical properties of extruded PLA sheet using an instrumented slit die and machine learning algorithms, *Polymer Testing*, 69, 462-469. https://doi.org/10.1016/j.polymertesting.2018.06.002

NAICS. *North American Industry Classification System*, 2017. https://www.census.gov/naics/reference_files_tools/2017_NAICS_Manual.pdf. Access on: 27 jun. 2021.

Nguyen, G., Dlugolinsky, S., Bobák, M. et al. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: A survey. *Artif Intell Rev* 52, 77–124. https://doi.org/10.1007/s10462-018-09679-z

Nasr, M. et al. (2020). Milling of Graphene Reinforced Ti6Al4V Nanocomposites: An Artificial Intelligence Based Industry 4.0 Approach. *Materials*, 13, 24, 5707. https://doi.org/10.3390/ma13245707

Naumov, S. et al. (2018). Experience in Use of Remote Access and Predictive Analytics for Power Equipment's Condition, *Thermal Engineering*, 65, 4, 189-199.

Nicholas, P. et al. (2020). Integrating real-time multi-resolution scanning and machine learning for Conformal Robotic 3D Printing in Architecture, *International Journal of Architectural Computing*, 18, 4. https://doi.org/10.1177/1478077120948203

Osswald, P. V. et al. (2020). Optimization of the production processes of powder-based additive manufacturing technologies by means of a machine learning model for the temporal prognosis of the build and cooling phase, *Production Engineering*, 14, 677-691 https://doi.org/10.1007/s11740-020-00987-4

Oyewobi, S. S., Djouani, K., & Kurien, A. M. (2022). Visible Light Communications for Internet of Things: Prospects and Approaches, Challenges, Solutions and Future Directions. *Technologies*, 10(1), 28. http://dx.doi.org/10.3390/technologies10010028

Ozgulbas, N., & Koyuncugil, A.S. (2019). Statistical Roots of Machine Learning, Deep Learning, Artificial Intelligence, Big Data Analytics and Data Mining. In The World Congress on Engineering and Computer Science, San Francisco, 2019.

Pinho, V. M. D. et al. (2020). Vision-Aided Radio: User Identity Match in Radio and Video Domains Using Machine Learning, *IEEE Access*, 8, 209619-209629. https://doi.org/10.1109/ACCESS.2020.3038926

Pistolesi, F., & Lazzerini, B. (2020). Assessing the Risk of Low Back Pain and Injury via Inertial and Barometric Sensors, *IEEE Transactions on Industrial Informatics*, 16, 11, 7199-7208. https://doi.org/10.1109/TII.2020.2992984

Qiu, D., Li, B., JI, S., & Leung, H. (2014). Regression Testing of Web Service: A Systematic Mapping Study. ACM *Computing Surveys*, 47, 2, 1-46. https://doi.org/10.1145/2631685

Rahmatov, N. et al. (2019). Machine learning–based automated image processing for quality management in industrial Internet of Things. *International Journal of Distributed Sensor Networks*, 15, 10. https://doi.org/10.1177/1550147719883551

Ratnayake, D. et al. (2020). Analysis of Machine Learning Algorithms in Smart Manufacturing. In 8TH International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, 2020.

Reinhardt, D. et al. (2020). CoBuilt 4.0: Investigating the potential of collaborative robotics for subject matter experts, *International Journal of Architectural Computing*, 18, 4, 1-18. https://doi.org/10.1177/1478077120948742

Ren, Y. S., Ma, C. Q., Kong, X. L., Baltas, K., & Zureigat, Q. (2022). Past, present, and future of the application of machine learning in cryptocurrency research. *Research in International Business and Finance*, 63(January), 101799. https://doi.org/10.1016/j.ribaf.2022.101799

Roberts, H., & Petticrew, M. (2006). *Systematic Reviews in the Social Scene: A practical guide*. 1. ed. Oxford: Blackwell Publishing XV.

Romeo, L. et al. (2020). Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0, *Expert Systems with Applications*, 140, 112869. https://doi.org/10.1016/j.eswa.2019.112869

Roveda, L. et al. (2020). Model-Based Reinforcement Learning Variable Impedance Control for Human-Robot Collaboration, *Journal of Intelligent & Robotic Systems*, 100, 417-433. https://doi.org/10.1007/s10846-020-01183-3

Sader, S., Husti, I., & Daróczi, M. (2020). Enhancing Failure Mode and Effects Analysis Using Auto Machine Learning: A Case Study of the Agricultural Machinery Industry, *Processes*, 8, 2, 224. https://doi.org/10.3390/pr8020224

Schirru, A., Pampuri, S., & Nicolao, G. (2010). Particle filtering of hidden Gamma processes for robust Predictive Maintenance in semiconductor manufacturing. In IEEE International Conference on Automation Science and Engineering, p. 51-56, Toronto, 2010.

Settanni, G. et al. (2018). Countering targeted cyber-physical attacks using anomaly detection in self-adaptive industry 4.0 systems. *Elektrotechnik & Informationstechnik*, 135, 3, 278-285

Simeone, A. et al. (2020). Intelligent Industrial Cleaning: A Multi-Sensor Approach Utilizing Machine Learning-Based Regression. *Sensors*, 20, 13. https://doi.org/10.3390/s20133642

Skilton, M., & Hovsepian, F. (2018). *The 4th Industrial Revolution: Responding to the Impact of Artificial Intelligence on Business*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-62479-2 SMLC. *Implementing 21st Century Smart Manufacturing*. 2011. https://www.controlglobal.com/assets/11WPpdf/110621_SMLC-smart-manufacturing.pdf.

Soori, M., Arezoo, B., & Dastres, R. (2023). Machine learning and artificial intelligence in CNC machine tools, A review. *Sustainable Manufacturing and Service Economics*, December 2022, 100009. https://doi.org/10.1016/j.smse.2023.100009

Soto, J.A.C., Tavakolizadeh, F., & Gyulai, D. (2019). An online machine learning framework for early detection of product failures in an Industry 4.0 context, *International Journal of Computer Integrated Manufacturing*, 32, 4-5, 1-14. https://doi.org/10.1080/0951192X.2019.1571238

Srinivas, S., & Young, A. J. (2023). Machine L earning and Artificial Intelligence in Surgical Research. *Surgical Clinics of NA*, 103(2), 299-316. https://doi.org/10.1016/j.suc.2022.11.002

Tannous, M. et al. (2020). Haptic-based touch detection for collaborative robots in welding applications, *Robotics and Computer-integrated Manufacturing*, 64, 101952. https://doi.org/10.1016/j.rcim.2020.101952

Tao, W. et al. (2019). A self-aware and active-guiding training & assistant system for worker-centered intelligent manufacturing, *Manufacturing Letters*, 21, 45-49. https://doi.org/10.1016/j.mfglet.2019.08.003

Terry, S., Lu, H., Fidan, I., Zhang, Y., Tantawi, K., Guo, T., & Asiabanpour, B. (2020). The Influence of Smart Manufacturing towards Energy Conservation: A Review. *Technologies*, 8, 31. https://doi.org/10.3390/technologies8020031

Tsutsui, T., & Matsuzawa, T. (2019). Virtual Metrology Model Robustness Against Chamber Condition Variation Using Deep Learning, *IEEE Transactions on Semiconductor Manufacturing*, 32, 4, 428-433. https://doi.org/10.1109/TSM.2019.2931328

Upasani, K. et al. (2017). Distributed Maintenance Planning in Manufacturing Industries, *Computers & Industrial Engineering*,108. https://dx.doi.org/10.1016/j.cie.2017.03.027

Vununu, C. et al. (2018). A Deep Feature Learning Method for Drill Bits Monitoring Using the Spectral Analysis of the Acoustic Signals, *Sensors*,18, 8. https://doi.org/10.3390/s18082634

Wang, J. et al. (2016). Large-Scale Online Multitask Learning and Decision Making for Flexible Manufacturing, *IEEE Transactions on Industrial Informatics*, 12, 6. https://doi.org/10.1109/TII.2016.2549919

Wang, J. et al. (2019). Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction, *Computers in Industry*, 111. https://doi.org/10.1016/j.compind.2019.06.001

Wang, y., Li, x., & Tsung, F. (2020). Configuration-Based Smart Customization Service: A Multitask Learning Approach, *IEEE Transactions on Automation Science and Engineering*, 17, 4. https://doi.org/10.1109/TASE.2020.2986774

Wen, x., & Chen, H. (2020). Heterogeneous Connection and Process Anomaly Detection of Industrial Robot in Intelligent Factory, International, *Journal of Pattern Recognition and Artificial Intelligence*, 34, 12. https://doi.org/10.1142/S0218001420590417

Wong, W. C. et al. (2018). Recurrent Neural Network-Based Model Predictive Control for Continuous Pharmaceutical Manufacturing, *Mathematics*, 6, 242. https://doi.org/10.3390/math6110242

Wu, D. et al. (2017). A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests, *Journal of Manufacturing Science and Engineering*, 139, 7. https://doi.org/10.1115/1.4036350

Xia, k. et al. (2021). A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence, *Journal of Manufacturing Systems*, 58. https://doi.org/10.1016/j.jmsy.2020.06.012

Yan, H. et al. (2018). Industrial Big Data Analytics for Prediction of Remaining Useful Life Based on Deep Learning, *IEEE Access*, 6. https://doi.org/10.1109/ACCESS.2018.2809681

Yao, H. et al. (2019). Resource Trading in Blockchain-Based Industrial Internet of Things, *IEEE Transactions on Industrial Informatics*, 15, 6. https://doi.org/10.1109/TII.2019.2902563

Yao, X. et al. (2017). From Intelligent Manufacturing to Smart Manufacturing for Industry 4.0 Driven by Next Generation Artificial Intelligence and Further On. In 2017 5th International Conference on Enterprise Systems (ES), p. 311-318, Beijing, 2017.

Zhao, R. et al. (2019). Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing, 115, 3, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

Zheng, X., Wang, m., & Ordieres-Meré, J. (2018). Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0. *Sensors*, 18, 7. https://doi.org/10.3390/s18072146

Zhuang, C. et al. (2021). The connotation of digital twin, and the construction and application method of shop-floor digital twin, *Robotics and Computer-integrated Manufacturing*, 68. https://doi.org/10.1016/j.rcim.2020.102075