



Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review

Marta Fernandes¹ · Juan Manuel Corchado² · Goretí Marreiros¹

Accepted: 5 February 2022 / Published online: 4 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

When put into practice in the real world, predictive maintenance presents a set of challenges for fault detection and prognosis that are often overlooked in studies validated with data from controlled experiments, or numeric simulations. For this reason, this study aims to review the recent advancements in mechanical fault diagnosis and fault prognosis in the manufacturing industry using machine learning methods. For this systematic review, we searched Web of Science, ACM Digital Library, Science Direct, Wiley Online Library, and IEEE Xplore between January 2015 and October 2021. Full-length studies that employed machine learning algorithms to perform mechanical fault detection or fault prognosis in manufacturing equipment and presented empirical results obtained from industrial case-studies were included, except for studies not written in English or published in sources other than peer-reviewed journals with JCR Impact Factor, conference proceedings and book chapters/sections. Of 4549 records, 44 primary studies were selected. In 37 of those studies, fault diagnosis and prognosis were performed using artificial neural networks (n=12), decision tree methods (n=11), hybrid models (n=8), or latent variable models (n=6), with one of the studies employing two different types of techniques independently. The remaining studies employed a variety of machine learning techniques, ranging from rule-based models to partition-based algorithms, and only two studies approached the problem using online learning methods. The main advantages of these algorithms include high performance, the ability to uncover complex nonlinear relationships and computational efficiency, while the most important limitation is the reduction in model performance in the presence of concept drift. This review shows that, although the number of studies performed in the manufacturing industry has been increasing in recent years, additional research is necessary to address the challenges presented by real-world scenarios.

Keywords Machine learning · Fault detection · Fault prognosis · Predictive maintenance · Manufacturing industry · Industrial case-study

✉ Marta Fernandes
mmdaf@isep.ipp.pt

Juan Manuel Corchado
corchado@usal.es

Goretí Marreiros
mgt@isep.ipp.pt

¹ GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto (ISEP/IPP), Porto, Portugal

² BISITE Research Centre, University of Salamanca (USAL), Salamanca, Spain

1 Introduction

Machine maintenance, with its impact on machine downtime and production costs, is directly related to a manufacturing companies' ability to be competitive in terms of cost, quality, and performance [1, 2]. The purpose of maintenance goes beyond repairing an equipment after it malfunctions. Its main objective is to maintain the functionality of machinery and minimize breakdowns.

As the name suggests, predictive maintenance consists in the early detection of problems. Under a predictive maintenance program, maintenance is performed by monitoring the actual condition of machinery and repairing or replacing

components after a certain level of deterioration has been detected, instead of performing repairs after a fault has occurred [3]. This approach has several advantages over reactive and preventive maintenance strategies [4, 5], namely:

- Prevention of catastrophic failures.
- Extension of an equipment's useful life.
- Optimization of preventive maintenance tasks.
- Improved management of the maintenance inventory.
- Optimization of equipment availability.
- Improved productivity.

By preventing serious failures, reducing unexpected faults, and maximizing the mean time between failures (MTBF), predictive maintenance helps reduce workplace accidents and their severity, reduces the number of repairs and the mean time to repair (MTTR) and extends the useful life of equipment, all of which results in increased earnings, less maintenance and production costs and more sustainable manufacturing [4, 6]. According to Sullivan et al. [5], the successful implementation of a predictive maintenance program can lead to an average reduction of maintenance costs between 25% and 30% and a return on investment (ROI) of 1000%.

Predictive maintenance is a form of condition-based maintenance [4], which relies on the prediction and detection of incipient faults in the equipment based on parameter measurements that reflect a machine's real condition [7–9]. In condition-based maintenance, decision-making is supported by diagnostics and prognostics techniques [7].

Diagnostics, which involves performing fault detection and identification (FDI), is generally performed using hardware redundancy methods or analytical redundancy methods. Hardware redundancy consists in measuring the same parameters using more than one sensor and then comparing the duplicate signals by means of various techniques, such as signal processing methods [10]. Analytical redundancy methods are based on mathematical models of the system and can be divided in quantitative, or model-based, methods and qualitative, or data-driven, methods [10, 11]. Both methods compare predicted or estimated parameters to real, measured values, but while model-based methods estimate the parameters of interest based on a mathematical model of the system under normal operating conditions, data-driven methods employ historical data and artificial intelligence algorithms to predict such parameters or detect anomalous values.

While diagnostics deals with the detection, isolation and identification of faults, prognostics aims to predict faults in the monitored system before they occur [7]. Specifically, prognostics techniques are used to estimate how soon - i.e., estimation of the remaining useful life (RUL) - and

how likely a fault is to occur, but most of the literature on machine prognostics focuses on the former type of prediction [7]. RUL estimation methods, which can also be data-driven, aim to predict how long a machine will function before a fault occurs or if the machine is going to fail in a given time interval [7].

Since they don't require additional hardware, analytical redundancy methodologies are less expensive to implement than hardware redundancy methods [10, 11]. Given the emergence of Internet of Things (IoT) technologies in industrial settings it is now possible to obtain a real-time digital representation of the production processes and current status of the equipment [12], which has led to an exponential growth of the volume of industrial data [13]. Data-driven methods, in particular machine learning and data mining techniques, are well suited to extract knowledge from this wealth of data and have successfully been used in the context of predictive maintenance [9, 14]. Moreover, although model-based methods can produce good results if the model of the system is precise, building an accurate mathematical model of a system is an arduous task that makes model-based methods a less viable option for complex systems [7, 10]. Recent review papers [9, 15] focusing on the use of machine learning techniques for predictive maintenance have identified that commonly used data-driven methods include artificial neural networks [16–20], support vector machines [21–23], decision trees (including ensemble methods) [24, 25], k-means [26, 27] and logistic regression [28, 29], among others.

Predicting and detecting faults in industrial equipment are difficult tasks that require the choice of adequate techniques to obtain accurate results. The present study performs a systematic literature review of the machine learning methods used for the detection of mechanical faults and the prognosis of faults in manufacturing equipment in real-world scenarios. It is meant to serve as a foundation for the implementation of predictive maintenance systems and help identify future research opportunities. The literature on mechanical fault detection and fault prognosis is vast, but to the best of the authors' knowledge no systematic literature review on this specific topic of study exists.

The review focuses on the detection of mechanical faults because these types of faults are a leading cause of breakdowns in manufacturing equipment [30, 31]. As mentioned above, fault prognosis aims to predict the time left before a machine breaks down and/or the probability of failure, without seeking to identify the type of fault (diagnostics techniques can be used for this purpose) [7]. Therefore, primary studies focusing on both mechanical fault detection and fault prognosis were considered in this review.

Another important aspect of this review is that only real-world industrial cases are considered. When put into

practice in the real world, predictive maintenance presents a set of challenges for fault diagnosis and prognosis that are often overlooked in studies validated with data obtained from controlled experiments, testbeds, or numeric simulations. Manufacturing systems are characterized by complex, non-stationary processes where noise and other disturbances are a reality [8, 32, 33]. This conditions the choice and applicability of machine learning methods, as do other aspects of practical order such as the absence of historical fault data that occurs frequently in industrial settings and restricts the learning task to unsupervised and semi-supervised methods. For these reasons, this study aims to present an overview of the current landscape of fault diagnosis and prognosis in real-world scenarios using machine learning techniques.

The study here presented was guided by five research questions aimed at characterizing the relevant research in terms of publication sources and scientific fields, as well as examining the state-of-the-art machine learning methods for mechanical fault detection and fault prognosis in manufacturing equipment, their strengths and weaknesses, and their application in the context of data stream learning. A search for eligible publications was conducted in five academic databases, which, after applying a set of criteria, culminated in the selection of forty-four primary studies.

The rest of this document is organized as follows: Section 2 presents the review protocol developed for this study, including the definition of the research questions, search strategy, study selection criteria and the data extraction strategy. The results obtained from conducting the review and answering the research questions are described in Section 3 and discussed in Section 4. Finally, Section 5 presents the concluding remarks and provides directions for future work.

2 Methods

This study follows the PRISMA statement [34], which establishes a checklist and a flow diagram for reporting systematic reviews. However, the PRISMA statement is oriented towards the healthcare field, whereas the present review covers themes related to engineering and computer science. Healthcare research differs significantly from research performed in engineering and computer science and, as such, the PRISMA statement does not apply in its entirety. For this reason, this study is also guided by the procedure presented in [35], which adapts different medical guidelines for performing systematic reviews to the particularities of software engineering, but is applicable to other scientific fields as well. The three main phases of this procedure, namely planning, conducting, and reporting

the review, as well as related activities are presented in Table 1.

The need for this review was identified while researching the literature of interest for the first author's PhD thesis about machine learning methods for fault detection and prediction. As far as the authors are aware, no systematic literature review of the machine learning methods used for mechanical fault detection and fault prognosis in manufacturing equipment in real-world scenarios currently exists.

Before undertaking the necessary research work, a review protocol was developed to establish suitable research questions and define the search strategy, study selection criteria and the data extraction process. The protocol is described in more detail in the following subsections.

2.1 Research questions

The first step in developing the protocol consisted in formulating meaningful research questions to guide a state-of-the-art review of the topic of study (Table 2).

The first research question is intended to help understand where papers that describe the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment have recently been published. The purpose is to identify not only the publication venues where the studies have been published and whether they tend to cluster around specific venues or not, but also determine the types of venues in which they were published. The latter is of particular interest considering this review focuses on industrial case-studies. Because fault detection and prognosis are studied in a wide range of scientific fields, the second research question aims to identify the fields that most commonly use machine learning methods for that purpose and if multidisciplinary approaches are present.

The purpose of research question three is to survey the machine learning algorithms and methods employed in the recent literature about mechanical fault detection

Table 1 Systematic review process

Phase	Activities
Planning the review	Identification of the need for the review
	Development of the review protocol
Conducting the review	Identification of research
	Selection of primary studies
	Study quality assessment
	Data extraction and monitoring
Reporting the review	Data synthesis

Table 2 Research questions

ID	Research Question
RQ1	In which publication venues are studies about the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment published?
RQ2	In which scientific fields has the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment been researched?
RQ3	What machine learning algorithms and methods are currently employed for mechanical fault detection and fault prognosis in manufacturing equipment?
RQ4	What limitations and advantages do those algorithms and methods present?
RQ5	Which of those algorithms and methods are used for data stream learning?

and fault prognosis in manufacturing equipment. To answer this question, aspects such as which machine learning algorithms are most frequently used, what types of learning tasks are addressed, or whether hybrid and ensemble methods are used should be considered. It is also important to learn why these algorithms are being used and what their weaknesses are, a matter addressed by research question four.

Much of the data used to predict and detect faults in manufacturing equipment is acquired by sensors that monitor the machines and produce high-speed data streams. Classical machine learning methods are not adequate to learn from these data streams, a task that presents unique challenges [36]. For that reason, research question five focuses on machine learning methods meant for data stream learning. The main aim is to determine how widespread the use of these methods is for mechanical fault detection and fault prognosis in manufacturing equipment, but also to understand how such techniques are being used.

2.2 Search strategy

To identify recently published research about machine learning methods for mechanical fault detection and fault prognosis in manufacturing equipment in real world scenarios, the following search strategy was devised.

2.2.1 Information sources

The five academic databases listed in Table 3 were chosen after considering search systems that were appropriate for systematic reviews [37] and whose subject was compatible with the topic of study. Although IEEE Xplore is not ideally suited for systematic reviews, it is an important research database in the fields of engineering, electron-

ics, and computer science and can be used to supplement the results obtained from the other four databases [37].

2.2.2 Search string

The search string used to find publications with the potential of being included in this systematic review was built by combining several search terms using the Boolean operators OR and AND (Table 4).

To identify studies that use machine learning, the terms “mining”, “learning” and “knowledge discovery” were included. The decision to use other terms besides “machine learning” stems from the fact that there is considerable overlap between machine learning and data mining, and the terms are often used interchangeably. Moreover, although “mining” and “learning” are meant to represent “data mining” and “machine learning”, respectively, the choice of using broader terms was made with the intention of finding research that employs other, related terms, such as “pattern mining” or “data stream learning”. “knowledge discovery” was included as well because it often makes use of machine learning techniques and can be relevant in the context of fault detection and prognosis.

To find research pertaining to mechanical fault detection and fault prognosis, the inclusion of the terms “fault detection”, “fault prediction” and “fault prognosis” was an obvious choice. However, it also made sense to include the term “predictive maintenance” since studies about this topic often propose fault detection or prognosis methods.

The search string is purposefully broad, not containing any terms that allude to mechanical faults, manufacturing equipment or industrial case-studies. If the string included those terms, the search results would be too narrow and many studies that do not explicitly use those terms would be left out.

Since the chosen academic databases have slightly different rules for building search strings, after devising the general search string presented in Table 4, specific strings were created for each of them. The following example illustrates the search string specified for the Web of Science (the TS tag field indicates the search terms should be looked up in the title, abstract and keywords):

TS = ((“mining” OR “learning” OR “knowledge discovery”) AND (“fault detection” OR “fault prediction” OR “fault prognosis” OR “predictive maintenance”))

2.3 Study selection criteria

A set of inclusion and exclusion criteria was defined to select the relevant studies from the search results. As can be seen in Table 5, the studies that should be included in the systematic literature review are those whose subject

Table 3 Research databases

Search system	Databases	URL
Web of Science	Core Collection – SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH	https://www.webofknowledge.com/
ACM Digital Library	The ACM Guide to Computing Literature	https://dl.acm.org/
Science Direct		https://www.sciencedirect.com/
Wiley Online Library		https://onlinelibrary.wiley.com/
IEEE Xplore		https://ieeexplore.ieee.org/Xplore/home.jsp

matter is the use of machine learning techniques for the detection of mechanical faults or prediction of faults in manufacturing equipment. Only studies that meet one or more of the inclusion criteria are of interest for the purpose of this review.

The exclusion criteria presented in Table 6 are meant to filter out research that does not satisfy other important characteristics. Duplicate publications are to be eliminated, as are publications that are not written in English or studies that were published in venues other than conference proceedings, book chapters/sections or journals with impact factor as defined in Clarivate’s Journal Citation Reports (JCR). Additionally, only full-length articles published since 2015 that present empirical results obtained from industrial case-studies are to be considered.

2.4 Data extraction strategy

The data collected from the selected studies is meant to answer the systematic review’s research questions. For that purpose, a data form template was created to extract information from each of the selected studies in a consistent manner (Table 7). To determine the ‘scientific fields’ of publications, an examination of the scientific categories of the publication venues will be carried out. In the case of conferences, the necessary information will be obtained from the official websites, whereas for journals the categories defined by Clarivate’s Journal Citation Reports (JCR) will be taken into consideration. Whenever a given publication is indexed in more than one JCR category, the category with the highest ranking will be chosen. If two categories or more have the same ranking, the authors of this review will decide which category is more appropriate. The

Table 4 Search string

Field of study	Search terms
Machine Learning	(“mining” OR “learning” OR “knowledge discovery”) AND
Fault Prediction/Detection	(“fault detection” OR “fault prediction” OR “fault prognosis” OR “predictive maintenance”)

‘country of research’ will be defined based on the country of affiliation of the first author.

3 Results

As can be seen in Fig. 1, the execution of the previously presented protocol resulted in the selection of 44 primary studies. In the identification phase, the search queries performed in the Web of Science, Science Direct, ACM Digital Library, Wiley Online Library, and IEEE Xplore databases yielded a total of 4549 records. After removing duplicate entries, a total of 3377 studies remained. These records were screened based on publication details, such as publication venue (EC3) and language (EC5), as well as on the information provided by the title and the abstract. Of the 3377 studies evaluated, 2821 did not meet the selection criteria. Additionally, seven publications had to be discarded because the full text was not available. The remaining 549 publications underwent a more detailed full text assessment to determine if they met the inclusion criteria and provided empirical results obtained from industrial case-studies (EC6). 505 studies had to be excluded, while the 44 studies that met the described criteria were selected for inclusion in the systematic review.

3.1 Distribution of publications by year and country

As can be seen in Fig. 2, there is a clear trend of increase in publications from 2016 to 2019. The majority (88.6%) of selected studies have been published since 2018, with a noticeable surge in the number of publications that year. However, the number of annual publications has

Table 5 Inclusion criteria

ID	Criteria
IC1	The publication focuses on the use of machine learning algorithms and methods for mechanical fault detection in manufacturing equipment.
IC2	The publication focuses on the use of machine learning algorithms and methods for prognosis of faults in manufacturing equipment.

Table 6 Exclusion criteria

ID	Criteria
EC1	Duplicate publication.
EC2	Publication is not a full-length article.
EC3	Study not published in a peer-reviewed journal with JCR Impact Factor, in conference proceedings, or in a book chapter/section.
EC4	Study published before 2015.
EC5	Publication not written in English.
EC6	Study without empirical results obtained from an industrial case-study.

been decreasing since 2020. Considering the search for publications to include in this review was undertaken in October 2021, it's unclear if this trend will continue until the end of 2021. Moreover, since the COVID-19 pandemic affected the scientific community significantly, delaying research work and publications [38], it is reasonable to expect that many studies that were planned for 2020 and 2021 will only be published in later years. It is also worth noting that no study from 2015 was selected for inclusion in the systematic review.

These publications come from 21 different countries (Table 8), but the distribution of the number of publications per country is positively skewed, i.e., most nations only published one or two studies. Only three countries published

Table 7 Data extraction form template

Data extraction form
Title:
Authors:
Country of research:
Publication venue:
Publication details:
Type of publication venue:
Scientific field:
Research questions answered:

Content:

- Which machine learning algorithms are used?
- How are they applied to the diagnosis of mechanical faults or prognosis of faults in manufacturing equipment?
- What types of learning tasks are addressed (e.g., supervised, unsupervised, reinforcement learning)?
- Are hybrid learning methods used?
- What types of learning techniques are employed (e.g., ensemble learning, transfer learning, online learning, etc.)?
- What are the benefits/limitations of the machine learning techniques used in the study?
- Which data stream learning techniques are used?

more than two studies, but together they were responsible for publishing 43.2% of the studies included in this review (Fig. 3). China, Germany, and Greece were the countries that published the most studies, with Germany and Greece contributing with six studies each and China with seven studies.

3.2 RQ1: In which publication venues are studies about the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment published?

The 44 selected studies were published in 36 distinct venues, of which 17 are journals and 19 are conferences, with only five venues publishing more than one study about the topic of interest (Table 9). The top publication sources include IEEE Access with five studies and the 2019 31st International Conference on Advanced Information Systems Engineering (CAiSE), CIRP Annals, Sensors and The International Journal of Advanced Manufacturing Technology with two studies each. 16 of the 36 publication venues are affiliated with the Institute of Electrical and Electronics Engineers (IEEE), representing 17.6% of journals, 68.4% of conferences and 20% of the top publication venues.

More than 52% of these distinct venues are conferences, but only 45.5% of the selected studies were published in conference proceedings versus 54.5% that were published in journals (Fig. 4). This reveals that the average number of papers published in journals is greater than the average number of papers published in conference proceedings. Furthermore, four of the five publication venues where more than one study was published are scientific journals and together these four venues published 25% of all the studies included in this review, which seems to imply there is a preference for publishing in scientific journals.

3.3 RQ2: In which scientific fields has the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment been researched?

The results show the recent research on machine learning for fault detection and prognosis in the manufacturing industry has been explored mostly by the computer science community. As shown in Table 10, computer science approaches account for 47.7% of the selected studies. Lagging considerably behind, but still worth considering, are engineering and multidisciplinary studies, with 25% and 13.6% respectively.

Multidisciplinary approaches involve several disciplines, such as telecommunications and cybernetics, but contributions from the fields of computer science, engineering, and

Fig. 1 PRISMA flow diagram of study selection

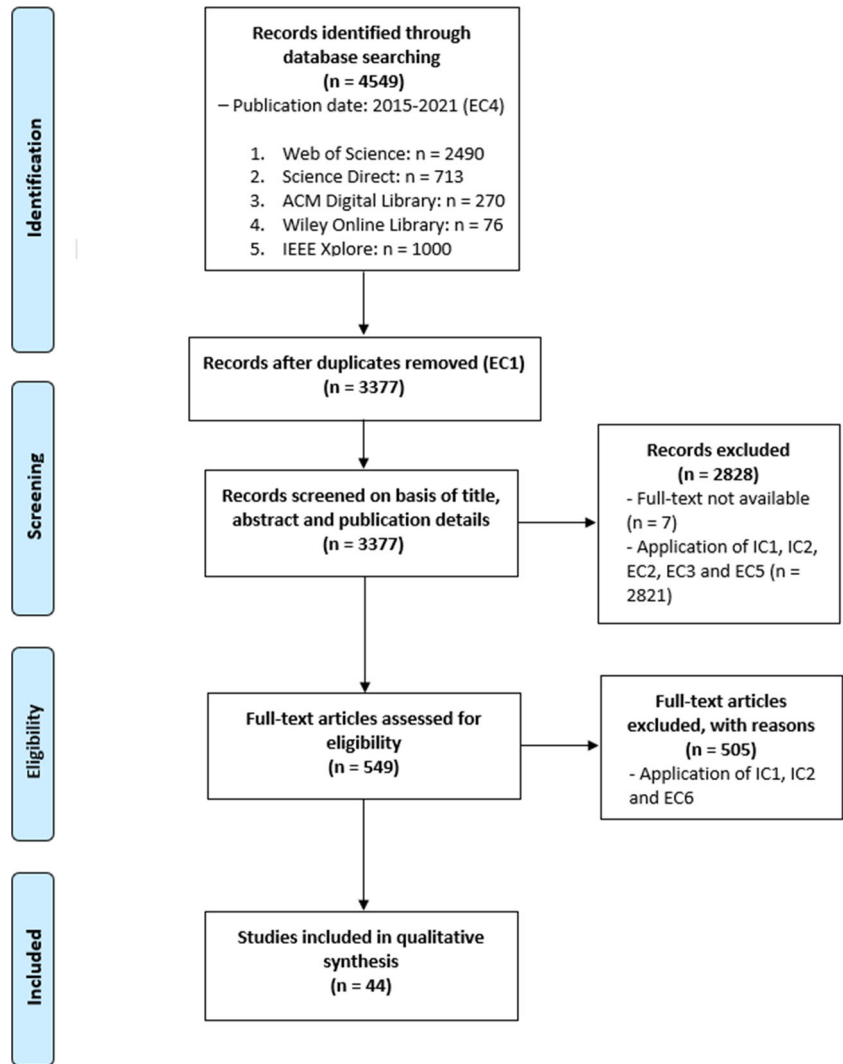


Fig. 2 Distribution of selected publications per year

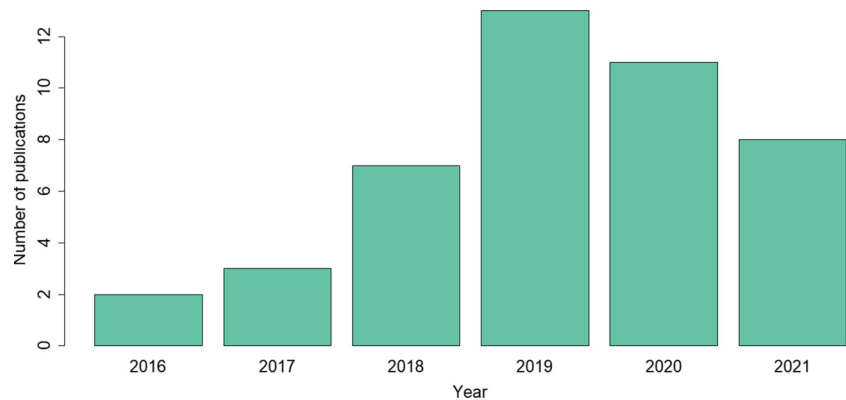


Table 8 Provenance of the publications included in the systematic review

Country or area	Number of publications	Country or area	Number of publications	Country or area	Number of publications
China	7	Turkey	2	Portugal	1
Germany	6	UK	2	Slovenia	1
Greece	6	USA	2	South Africa	1
Australia	2	Belgium	1	Spain	1
Italy	2	India	1	Sweden	1
South Korea	2	Netherlands	1	Switzerland	1
Singapore	2	Norway	1	Taiwan	1

automation and control systems are strongly prevalent even in this broader category.

3.4 RQ3: What machine learning algorithms and methods are currently employed for mechanical fault detection and fault prognosis in manufacturing equipment?

The selected primary studies employ a variety of machine learning algorithms and methods to perform mechanical fault detection and fault prognosis, including combinations of different algorithms. Most studies also perform comparative analyses between different machine learning algorithms to demonstrate the value of the proposed method or to select the most adequate algorithm. In the latter case, only the selected (or best performing) algorithms will be described in this review. These algorithms include: AdaBoost; agglomerative clustering (AC); autoencoder; autoregressive integrated moving average (ARIMA); back-propagation neural network (BPNN); classification and regression trees (CART); classification based on associations – classifier building algorithm (CBA-CB); convolutional neural network (CNN); deep neural network (DNN);

density-based spatial clustering of applications with noise (DBSCAN); discrete Bayes filter (DBF); eXtended classifier system (XCS); frequent pattern growth (FP-Growth); Gaussian mixture models (GMM); gradient boosting decision trees (GBDT); hidden Markov model (HMM); hierarchical clustering (HC); isolation forest (IF); k-means; K-multi-dimensional time-series clustering (K-MDTSC); k-nearest neighbors (K-NN); k-singular value decomposition (K-SVD); local outlier factor (LOF); logistic regression (LR); long short-term memory (LSTM); LSTM autoencoder; long short-term memory - generative adversarial network (LSTM-GAN); mean shift clustering (MSC); micro-cluster continuous outlier detection (MCOD); multilayer perceptron (MLP); naïve Bayes (NB); neighbourhood component analysis (NCA); partial least squares regression (PLSR); principal component analysis (PCA); quadratic discriminant analysis (QDA); quantitative association rule mining algorithm (QARMA); random forest (RF); random survival forest (RSF); recurrent neural network (RNN); simple linear regression; spectral clustering (SC); stacked sparse autoencoders (SSAE); support vector machines (SVM).

To facilitate the analysis of the data, the algorithms were grouped into different categories, as shown in Table 11.

Figure 5 illustrates that most studies included in this review (84.1%) use machine learning algorithms belonging to four categories, namely artificial neural networks with 12 publications, decision trees with 11 publications, hybrid models with eight publications and latent variable models with six publications. One of these studies uses both an artificial neural network and a hybrid model to address different problems. The remaining eight studies apply algorithms from a variety of categories. It is also worth noting that 13 studies make use of ensemble learning techniques.

The selected studies handle different types of learning tasks depending on the problems under consideration and the data that is available. As can be seen in Fig. 6, 53.3% of publications employ supervised learning techniques, 28.9% use unsupervised learning techniques, 15.6% make use of both supervised and unsupervised techniques and 2.2%

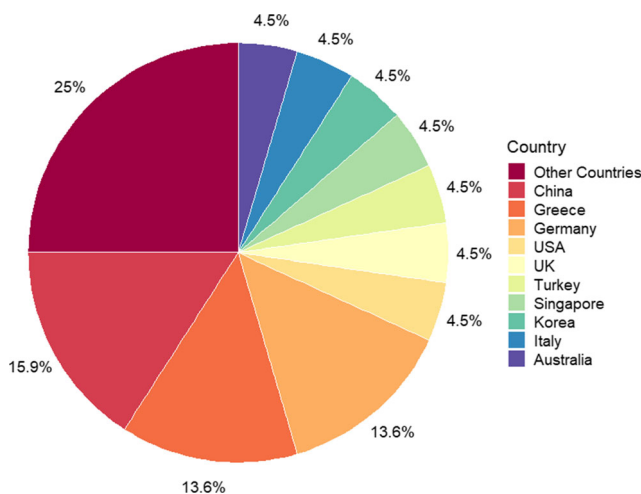
**Fig. 3** Share of publications by country

Table 9 Studies per publication venue

Publication venue	Studies	Venue type
(International conference) Machine Learning for Cyber Physical Systems (ML4CPS 2018)	[39]	conference
2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)	[40]	conference
2016 International Symposium on Leveraging Applications of Formal Methods, Verification and Validation (ISoLA)	[41]	conference
2017 13th IEEE Conference on Automation Science and Engineering (CASE)	[42]	conference
2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)	[43]	conference
2018 IEEE 20th Conference on Business Informatics (CBI)	[44]	conference
2018 IEEE International Conference on Big Data (Big Data)	[45]	conference
2018 IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA)	[46]	conference
2018 International Conference on Information and Computer Technologies (ICICT)	[47]	conference
2019 16th International Conference on Distributed Computing and Artificial Intelligence (DCAI)	[48]	conference
2019 16th International Symposium on Wireless Communication Systems (ISWCS)	[49]	conference
2019 Chinese Automation Congress (CAC)	[50]	conference
2019 Genetic and Evolutionary Computation Conference (GECCO)	[51]	conference
2019 IEEE 17th International Conference on Industrial Informatics (INDIN)	[52]	conference
2019 IEEE 5th World Forum on Internet of Things (WF-IoT)	[53]	conference
2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)	[54]	conference
2020 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining	[55]	conference
2020 International Conference on Development and Application Systems (DAS)	[56]	conference
Advances in Manufacturing	[57]	journal
Applied Sciences	[58]	journal
Business & Information Systems Engineering	[59]	journal
Complexity	[60]	journal
Computers & Industrial Engineering	[61]	journal
Electronics	[62]	journal
Engineering Applications of Artificial Intelligence	[63]	journal
Expert Systems with Applications	[64]	journal
IEEE Transactions on Industrial Electronics	[31]	journal
IEEE Transactions on Industrial Informatics	[65]	journal
Journal of Manufacturing Systems	[66]	journal
Simulation Modelling Practice and Theory	[67]	journal
Studies in Informatics and Control	[68]	journal
2019 31st International Conference on Advanced Information Systems Engineering (CAISE)	[69, 70]	conference
CIRP Annals	[71, 72]	journal
Sensors	[73, 74]	journal
The International Journal of Advanced Manufacturing Technology	[75, 76]	journal
IEEE Access	[77–81]	journal

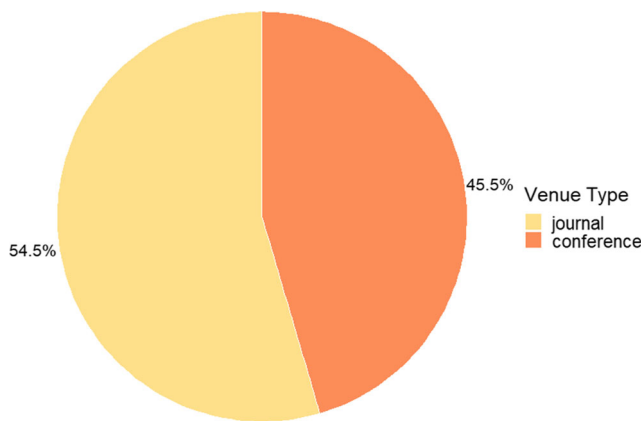


Fig. 4 Proportion of publications in conferences and journals

combine semi-supervised, unsupervised, and supervised techniques. The use of unsupervised techniques is motivated mostly by an absence of labeled data [47, 48, 50, 54, 59–62, 65, 70], although in some studies they are employed to detect outliers [74], reduce dimensionality [31, 75] or extract features [81]. In studies [39, 42, 45, 58, 66, 69, 71], labeled data was available, but was used to validate the unsupervised learning models.

3.5 RQ4: What limitations and advantages do those algorithms and methods present?

Of the 44 selected studies, 33 described the motivations for choosing a particular machine learning algorithm or combination of algorithms. Some of these motivations relied on the inherent strengths of the algorithms employed, while others considered the specific advantages an algorithm could have for fault detection and prediction, or for its implementation in industrial environments. In addition, the benefits provided by the proposed approach were also reported in several studies. On the contrary, only eleven studies presented the limitations of either the algorithms employed or the proposed approaches.

In the following subsections, Tables 12, 13, 14, 15 and 16 summarize the advantages and limitations of these machine learning algorithms and methods. After each table, they are described in more detail.

3.5.1 Decision trees

Models in the decision tree category have several characteristics that make them suitable for implementation in industrial contexts. In [41], the authors decided to use a classification tree due to its interpretability. CART models are white box classifiers whose outputs can be represented by a series of if statements. This allows factory engineers to analyze the model and understand the reasoning that led to a given decision. However, this advantage is lost when using ensembles of trees, since they combine several base models to obtain a more robust output.

Nonetheless, ensemble tree models are highly valued for their efficiency and effectiveness. Random forest models were used for these reasons in [44] and [53]. Specifically, in [44] a random forest was used to develop a proof of concept that would allow the authors to demonstrate that relevant results could be obtained from real world data in a short period of time. More complex algorithms would not have been appropriate in a situation where high predictive power and low implementation effort were necessary. Additionally, the ability of random forest models to reduce variance and increase generalizability was also taken into consideration, since the amount of training samples was relatively small, but the feature space was large. The predictive tool proposed in [76] also makes use of an ensemble method, specifically GBDT, due to the algorithm's low computational complexity and predictive power when handling large-scale datasets. The algorithm's ability to assess the importance of features was also essential to determine which time lag should be used as input to DPCA. However, the proposed approach has the disadvantage of lacking interpretability, not only because GBDT uses an ensemble of decision trees, but also because the data used to predict failures in the milling machine

Table 10 Proportion of studies per scientific field

Scientific field	Studies	Number of publications
Computer Science	[39, 41, 44–48, 51, 55, 59, 61, 63, 64, 67, 69, 70, 77–81]	47.7% (n = 21)
Engineering	[31, 54, 57, 58, 62, 65, 66, 71, 72, 75, 76]	25% (n = 11)
Multidisciplinary	[40, 43, 52, 53, 56, 60]	13.6% (n = 6)
Automation & Control Systems	[42, 50]	4.5% (n = 2)
Instruments & Instrumentation	[73, 74]	4.5% (n = 2)
Operations Research & Management Science	[68]	2.3% (n = 1)
Wireless Communications, Networking and Signal Processing	[49]	2.3% (n = 1)

Table 11 Machine learning algorithms and methods employed in the selected primary studies

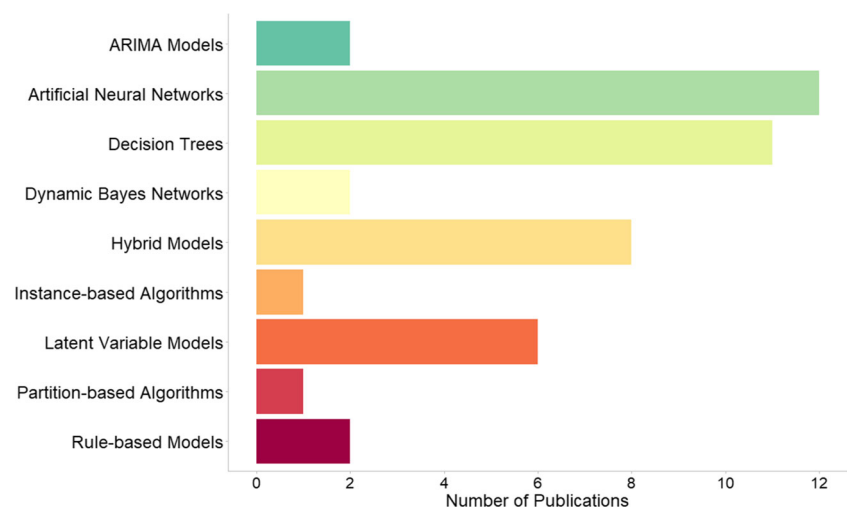
Category	Method	Publications	Learning task
ARIMA models	ARIMA	[48, 61]	unsupervised
Partition-based algorithms	MCOD	[69]	unsupervised
Decision Trees	Adaboost	[72]	supervised
	CART	[41, 46, 68]	supervised
	IF	[67]	supervised
	RF	[43, 44, 46, 52, 53]	supervised
	RSF	[71]	unsupervised
	GBDT	[76]	supervised
Dynamic Bayes Networks	DBF	[63]	supervised
	HMM	[42]	unsupervised
Hybrid models	ARIMA + LSTM	[50]	unsupervised
	DBSCAN + RF	[74]	unsupervised + supervised
	DBSCAN + SVM	[70]	unsupervised + supervised
	[HC / time series clustering] + RNN	[59]	unsupervised + supervised
	One-class SVM + K-Means + RF	[45]	semi-supervised + unsupervised + supervised
	Autoencoder + Simple Linear Regression	[58]	unsupervised + supervised
Instance-based algorithms	GMM + FP-Growth + CBA-CB	[75]	unsupervised + supervised
	CNN + NCA + Medium Gaussian SVM / CNN + NCA + ensemble subspace K-NN	[80]	supervised
	K-NN	[39]	supervised
Latent Variable Models	PCA	[65]	unsupervised
	GMM	[47]	unsupervised
	K-Means	[54]	unsupervised
	PLSR	[64]	supervised
	K-SVD	[60]	unsupervised
	K-MDTSC	[62]	unsupervised
Artificial Neural Networks	ANN	[57]	supervised
	BPNN	[40]	supervised
	CNN	[78]	supervised
	DNN	[77]	supervised
	LSTM	[70]	supervised
	MLP	[56]	supervised
	SSAE + BPNN	[31]	unsupervised + supervised
	SSAE + Softmax Classifier	[81]	unsupervised + supervised
	LSTM Autoencoder	[73]	supervised
	LSTM - GAN	[79]	supervised
	RNN	[55]	supervised
Rule-based models	Conditional Variational Autoencoder	[66]	unsupervised
	R4RE (“Rules 4 Rare Events” based on QARMA)	[49]	supervised
	XCSR	[51]	supervised

consists in the principal components obtained from the application of DPCA, which do not represent any physical properties or measurements of the system. The study presented in [67] used an ensemble method as well due to its

efficiency in terms of computation time and memory when handling large amounts of data.

In [71], the proposed model (manufacturing system-wide balanced random forest [MBRSF]) incorporated a

Fig. 5 Number of publications per category of machine learning algorithms



random survival forest because of its ability to handle bias and variance issues. The model captured complex fault patterns and diverse fault propagation pathways and made breakdown predictions for a time horizon not yet found in the manufacturing systems literature. Another advantage pointed out by the authors, was the theoretical guarantee provided for the prognostic performance due to the integration of the RSF model with data balancing techniques. Research undertaken by the authors demonstrated the MBRSF could attain a prognostic performance, with respect to an integrated Brier score, 90% better than other methods.

3.5.2 Artificial neural networks

Like ensemble methods, artificial neural networks suffer from a lack of interpretability, making them unsuitable for use in situations where it is necessary to know what factors contributed to a machine failure. They do, however, possess several advantages including good fault tolerance, the ability to learn complex nonlinear relationships and strong generalization abilities, which motivated their application in [57]. Likewise, in [77], a deep neural network was used

due to its ability to map the complex relationship between signals and the health status of industrial equipment. The use of a deep learning model was also considered because such models are capable of uncovering patterns in raw time series data, which eliminated the need to use signal processing techniques.

The method described in [78] explored the ability of convolutional neural networks to recognize and classify images by transforming time series data into images and using them as inputs to a CNN model. This approach has been shown to be suitable for maintaining temporal information and learning time-invariant features, thus resulting in improved classification performance. The proposed framework also included the option of using a parametric rectified linear unit (PReLU) function as an activation function to further improve performance when dealing with large datasets.

In [40], to overcome the BPNN's limitations the authors used a genetic algorithm to optimize the network's initial weights, thresholds and number of hidden layer neurons. With this technique, they were able to obtain faster convergence, more accurate fault predictions and less computational complexity.

Fig. 6 Types of learning tasks considered in the selected studies

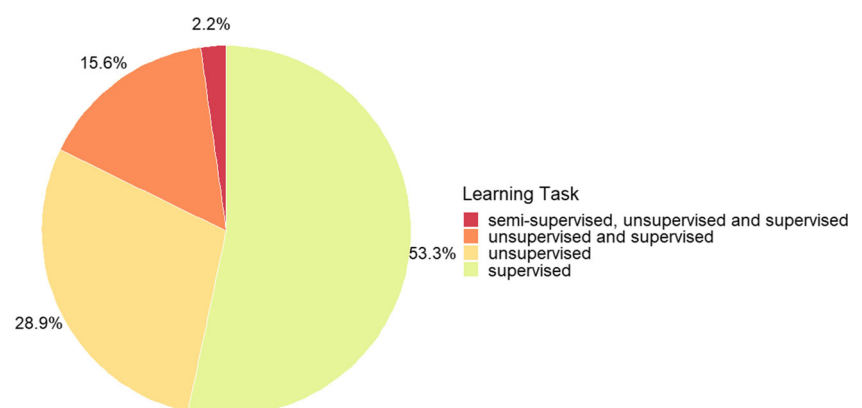


Table 12 Advantages and limitations of the decision tree algorithms employed for mechanical fault detection and fault prognosis

Publication	Advantages	Limitations
[41]	Algorithm: interpretability.	Not identified
[44]	Algorithm: high predictive performance; low implementation effort; reduces variance and increases generalizability.	Not identified
[76]	Algorithm: high predictive performance; low computational complexity; provides information about feature importance.	Proposed approach: lacks interpretability.
[53]	Algorithm: efficiency; effectiveness.	Not identified
[67]	Algorithm: computational efficiency	Not identified
[71]	Algorithm: handles bias and variance issues Proposed approach: theoretical guarantee for prognostic performance; captures complex fault patterns and fault propagation pathways.	Not identified

Similarly, the authors of the study proposed in [73] chose to detect faults and predict the RUL using LSTM-autoencoders because the combination of LSTMs and autoencoders has shown potential for accurate time series forecasting. According to the authors, LSTM-autoencoders have produced better forecasts than multilayer perceptrons, deep belief networks or LSTMs, due to their ability to identify the temporal patterns present in time series data and their superior feature extraction capability. However, the hyperparameters of the network impacts its performance significantly and choosing them can be a difficult task. The proposed approach, whereby one LSTM-autoencoder for each health state is trained, can be adjusted to handle different health states (labels) and be applied to different machines. However, the complexity of the architecture can increase rapidly, and the system might not be able to identify neighbouring health states. Additionally, this approach requires labelled data, which is not easily available in industrial settings.

Deep learning models are able to learn features from raw data as long as the training and test data share the same distribution and feature space. However, under time-varying conditions, such as those encountered in real industrial settings, this condition often does not hold. To handle this issue, the method proposed in [31] goes beyond simple pattern recognition and classification of existing faults by using deep learning to identify the dynamic properties of the machine tool. This enabled the early detection of fault features and the diagnosis of the machine's health status under time-varying operation.

The framework presented in [66] also addresses the issue of time-varying operations. The proposed method relies on two conditional variational autoencoder (CVAE) models

to estimate the health index of the machining centre and predict its future condition for a given operating regime. The authors of this study chose the CVAE due to its ability to remove noise from sensor data and extract meaningful features from the data automatically. Additionally, CVAEs are capable of learning complex conditional probability distributions regardless of the dimensionality of the feature space and can, therefore, be used to generate conditional data. This characteristic can be very helpful when handling industrial data since it facilitates the simulation of different production sequences regarding the current health state of a machine. Owing to these characteristics, the authors were able to develop a method that can estimate a machine's health under time-varying operations in a scenario where very little labelled data was available, as is often the case in real industrial settings.

In [81], the problem of different data distributions in the training and test data was also addressed, as was the issue of insufficient or low-quality training data, a common problem in the manufacturing industry. To solve these issues, the authors of the study combined deep transfer learning with digital-twin technology. The digital entity was used to simulate the entire product life cycle and generate vast amounts of data under different working conditions, while deep transfer learning was used to extract knowledge from the digital domain and apply it in the physical domain where the model was fine-tuned. With this approach, the authors of the study were able to explore shared knowledge and create a model that remained viable when put into production.

Changes in the data distribution were also the focus of the method proposed in [55]. SERMON is a model capable of modelling the temporal dependency present in streaming

Table 13 Advantages and limitations of the artificial neural networks employed for mechanical fault detection and fault prognosis

Publication	Advantages	Limitations
[31]	Proposed approach: early detection of fault features; diagnosis of machine's health status under time-varying operation.	Algorithm: performance decreases when training and test data don't share the same distribution.
[40]	Proposed approach: faster convergence; improved prediction performance; less computational complexity.	Algorithm: slow convergence speed; low precision; falls easily into local minimum; number of hidden layers difficult to determine.
[57]	Algorithm: fault tolerance; learns complex nonlinear relationships; strong generalization abilities.	Algorithm: lack of interpretability.
[81]	Proposed approach: knowledge sharing; generation of vast amounts of data through simulation of the entire product life cycle.	Algorithm: performance decreases when training and test data don't share the same distribution.; poor performance in case of insufficient or low-quality training data.
[78]	Proposed Approach: improved classification performance; maintains temporal information and learns time-invariant features.	Not identified
[77]	Algorithm: learns complex nonlinear relationships; uncovers patterns in raw time series data.	Not identified
[55]	Proposed approach: online learning capability; autonomous structural evolution; capable of adapting to drifts in the input data; capable of learning under finitely/infinately delayed label scenarios.	Not identified
[73]	Algorithm: capable of processing time series data; superior feature extraction capability; better forecasting ability. Proposed approach: can be adjusted to different types of machines and labels.	Algorithm: choice of suitable hyperparameters is complex and affects the performance of the network; Proposed approach: using multiple ANNs can be a difficult task; might not be able to identify neighbour states; requires labelled data.
[79]	Proposed approach: capable of generating large volumes of fault data; avoids mode collapse; improved accuracy and efficiency.	Not identified
[66]	Algorithm: denoising effect; automatic extraction of meaningful features; capable of learning complex probability distributions; well-understood and stable; can be used for conditional data generation. Proposed approach: estimation of machine's health status under time-varying operations; capable of handling sparse industrial data; product-specific health index can be used for scheduling maintenance and production.	Not identified

data and adapt to the changing characteristics of the data as it arrives in real-time. It does so thanks to the self-evolving architectures of its two RNNs. The SERN component can dynamically change the number of hidden units, while the MERN component can dynamically change the number of hidden layers as well. Moreover, to handle scenarios where labels might be delayed or inexistent, SERMON includes a mapping unit that employs an autoencoder to suggest possible data labels.

An LSTM-GAN was used in [79] as part of a PdM methodology that can not only monitor the health state of machines and predict faults before they occur, but also

provides the factory's maintenance staff with maintenance plans that are appropriate to deal with the issues detected by the state prediction and fault prediction models. The GAN was used to generate a large volume of synthetic fault data to improve the accuracy of the model. However, GANs may suffer from mode collapse. Due to the inclusion of memory units, gate structures and attention mechanisms, LSTM networks can alleviate the mode collapse issue. Moreover, LSTMs are capable of extracting patterns from long sequences of input data, making them ideal to detect abnormalities in condition monitoring data.

Table 14 Advantages and limitations of the hybrid models employed for mechanical fault detection and fault prognosis

Publication	Advantages	Limitations
[45]	Proposed approach: detects and classifies different mechanical faults in unlabeled data.	Not identified
[50]	LSTM: captures nonlinear relationships in sequential data. ARIMA: models linear associations present in time series data.	Not identified
[59]	Proposed approach: optimization of the performance of the proposed fault prognosis model. WPGMC clustering: interpretability. RNN: captures complex, nonlinear relationships in time series data.	Not identified
[74]	Proposed approach: uncovers patterns of wear and tear in unlabeled data. Random Forest: improved performance; robustness when handling numerical data and real-world problems.	Not identified
[58]	Proposed approach: improved accuracy. Autoencoder: can learn the relationship between the input data variables.	Proposed approach: anomaly threshold is defined arbitrarily; could use more a more sophisticated model to improve prediction accuracy.
[75]	Proposed approach: can learn from unlabeled data; applicable to different domains. GMM: capable of reducing the number of clusters. FP-Growth: handles large databases efficiently; can handle itemsets with low support threshold. Proposed approach: interpretability; can handle different types of sensor data; simple to set-up.	Not identified

3.5.3 Hybrid models

Hybrid machine learning models are created with the intention of solving tasks that a single algorithm, or type of algorithms, is not suited to handle. The analysis performed in [45] demonstrated that supervised learning algorithms are more appropriate for classification tasks but require labeled data and misclassify unknown faults. Semi-supervised algorithms can overcome these limitations, but are not capable of distinguishing between fault types, which is where unsupervised clustering algorithms can be useful. The combination of these different types of algorithms led to the development of a predictive maintenance system capable of detecting and classifying different mechanical faults from unlabeled data.

In [50], Cheng et al. decided to combine the strengths of ARIMA models and LSTM neural networks to optimize the performance of the proposed fault prognosis model. Since LSTM networks are artificial neural networks capable of handling long-term dependencies, they are ideal to capture nonlinear relationships in sequential data. Conversely, ARIMA models, which were developed for time series

analysis, are well suited to model the linear associations present in time series data.

The study presented in [59] employed clustering techniques and a recurrent neural network to overcome the problem of missing labels. The weighted pair-group method using centroid (WPGMC) was chosen for its ability to create homogeneous groups that could be more easily interpreted, while the RNN was selected because it possesses internal memory and is, therefore, able to capture complex, non-linear relationships in time series data. This ability is particularly important to uncover the patterns of wear and tear that occur in industrial equipment.

In [74] a random forest was used to perform fault detection due to its robustness when handling numerical data and real-world problems. Nonetheless, to improve the model's performance, DBSCAN was first used to detect outliers that might represent noise in the sensor data. This method improved the random forest's accuracy by 1.462% and further experiments demonstrated that using DBSCAN to detect and remove outliers improved the accuracy of other models as well.

Table 15 Advantages and limitations of the latent variable models employed for mechanical fault detection and fault prognosis

Publication	Advantages	Limitations
[47]	Algorithm: suitable to handle data generated from multimodal distributions.	Algorithm: assumes the data was generated from a mixture of finite Gaussian distributions.
[65]	Proposed approach: real-time analytics; scalable distributed implementation.	Not identified
[54]	Algorithm: cluster centers can adapt to new data; capable of classifying data despite noise and outliers. Proposed approach: capable of gaining a deep understanding about the equipment's performance.	Not identified
[64]	Algorithm: interpretability; suitable for small sample size and high-dimensional data. Proposed approach: Produces stable and consistent results.	Not identified
[60]	Algorithm: high computational efficiency; greater adaptability; robust to noise; uncovers patterns in raw time series data. Proposed approach: superior periodic impulse extraction.	Not identified
[62]	Algorithm: can handle raw multi-dimensional time series; solves the problem of empty cluster creation.	Algorithm: the time-series must be synchronous

The approach proposed in [58] constructs a health index by taking advantage of an autoencoder's ability to learn the relationship between the input data variables. Simple linear regression was subsequently used to predict future values of the health index and calculate the RUL. The proposed methodology is capable of learning from unlabeled data, and it was demonstrated that it can be applied in different domains. However, it is precisely because run-to-failure data wasn't available that the anomaly threshold had to be defined somewhat arbitrarily. Additionally, the prediction accuracy could be improved by using algorithms more sophisticated than simple linear regression.

The methodology proposed in [75] combines several learning models to perform fault prediction in a press module. An important concern when developing the fault prediction method was its interpretability, which is why the authors opted for an association rules approach. Additionally, the proposed approach requires few tuning of parameters and is generic enough to be applied to other types of sensor data.

3.5.4 Latent variable models

In the study presented in [47], the authors identified multimodal distributions when plotting the data. As a non-parametric method of density estimation, a GMM represents an appropriate choice for this kind of problem. However, this type of model has the disadvantage of assuming the data

is generated from a mixture of finite Gaussian distributions of unknown parameters.

In [65], Yu et al. developed a fault detection system using a distributed version of PCA. The selection of PCA took into consideration its real-time analytics ability when integrated with cloud computing, as well as the scalability of the distributed implementation. PCA was also a natural choice since labeled data was unavailable.

In [54], the authors opted for a cognitive analytics-based approach in order to gain a deeper understanding of how an industrial robot arm performed. Unlike what the authors identified as traditional data analytics frameworks, the proposed framework merges the information from the different data sources and analyses the correlation between the data features to understand how the robot arm operates under normal circumstances. K-means was used due to its ability to accurately cluster the data even in the presence of noise, but also because the cluster centres can be dynamically adapted when new data arrives.

PSLR was used in [64] because it is theoretically adequate to handle high-dimensional data and small sample sizes. PSLR was also chosen due to its explanatory power. Using correlation plots it's possible to determine the contribution of each variable to the prediction result. In addition, PSLR produces results that are stable, consistent, and can be easily maintained.

The authors of [60] took advantage of K-SVD's robustness to noise and its ability to capture the characteristic

Table 16 Advantages and limitations of other algorithms and methods employed for mechanical fault detection and fault prognosis

Category	Publication	Advantages	Limitations
Dynamic Bayes Networks	[42]	Algorithm: suitable to model time series data; can detect long-term degradation; can handle dynamic features in an unsupervised way. Proposed approach: handles data with asynchronous sampling rates; doesn't require comprehensive domain knowledge.	Algorithm: performance decreases when training and test data don't share the same distribution.
	[63]	Algorithm: capable of integrating data from heterogeneous sources; capable of incorporating information about uncertainty; able to rapidly adapt to changes; short execution time; low memory requirements; high performance.	Not identified
Partition-based algorithms	[69]	Algorithm: appropriate for online learning; low-memory and processing requirements; comprehensible results.	Algorithm: Sensitive to the input parameters.
Rule-based models	[49]	Algorithm: allows for quantification of the consequent item in closed intervals; incorporates online pruning of the generated rules within the search process. Proposed approach: improvement of the RUL estimates; reduced error rates in the test set.	Not identified
	[51]	Algorithm: appropriate for online learning; detects dependencies between variables. Proposed approach: provides valuable information to identify the origins of failures.	Not identified

components hidden in raw signals to denoise the original vibration signal. However, prior to doing so the K-SVD was improved (IKSVD) to make it significantly more efficient and adaptable. IKSVD in combination with fast spectral correlation (FSC) demonstrated to be superior to traditional approaches when it comes to extract periodic impulses from vibration data.

The algorithm proposed in [62] is based on k-means but employs a generalized notion of the Euclidean distance to handle multi-dimensional time-series and also addresses the issues k-means has with empty clusters. Moreover, the algorithm is capable of handling raw time series without needing any transformations such as the Fourier transform, or the wavelet transform. Because of the generalized distance defined for the algorithm, it is necessary for the time series data to be synchronous but this can be achieved with adequate data pre-processing.

3.5.5 Other approaches

The authors of [42] used a HMM since these models assume that a system's current hidden state is influenced by its previous hidden state. This means a HMM is an appropriate model for time series data and can be used to detect long-term degradation. The HMM is also capable

of handling dynamic features in an unsupervised way. In addition, the proposed approach, which combines a HMM with sliding windows and a genetic algorithm, can handle data with asynchronous sampling rates and doesn't require comprehensive domain knowledge. Nonetheless, since the parameters of the HMM depend on the feature values of the production cycles, when the contamination probability distribution of a production cycle differs substantially from the probability distributions of the production cycles used to train the HMM the model's performance decreases. This can be remedied by taking into consideration the advice of the maintenance experts when choosing production cycles to train the HMM.

The R4RE algorithm proposed in [49] is an improvement of the QARMA framework. As such, like QARMA, R4RE is fully distributed and guarantees that all the resulting rules are meaningful and meet the interestingness criteria defined by the user. The R4RE algorithm surpasses QARMA by overcoming two of its important limitations, namely it allows for quantification of the consequent item in closed intervals and incorporates online pruning of the generated rules within the search process. In the study presented in [49], these developments resulted in the improvement of the RUL estimates and reduced the error rates obtained in the test set.

The rule-based method (XCS) presented in [51] has a “covering” mechanism that enables the recalibration of the rule set for unseen data without needing to re-train and re-test the whole model, making it appropriate for online learning. Since the proposed model is also suitable to detect dependencies between variables and to recognize different failure patterns, the rules generated by the XCS can provide valuable information to identify the origins of failures.

In [63], the choice of a DBF to predict the degradation of machinery took into account the difficulties of implementing a predictive maintenance system. DBFs are well suited to predict faults in industrial settings due to their ability to integrate data from heterogeneous sources, to incorporate information about uncertainty and to rapidly adapt to changes. Moreover, these models have a short execution time, low memory requirements and high performance, which are key properties of industrial systems.

In [69], MCODE was chosen because of its low-memory and processing requirements, which make it ideal for processing streaming data, as well as for producing results that are easy to understand. This algorithm has the downside of being sensitive to the input parameters, which can influence the number of outlier reports.

3.6 RQ5: Which of those algorithms and methods are used for data stream learning?

Only the studies presented in [55] and [69] proposed methods for detecting faults directly from real-time data and applied stream learning techniques.

The model proposed in [55], named SERMON, consists of two RNNs that work in a cooperative manner to obtain better results in terms of modelling the temporal dependency present in streaming data, and the ability to self-evolve allows them to adapt to the changes (drifts) that characterize non-stationary data. This model is described in more detail in Section 4.3.

The algorithm used in [69], MCODE, is a state-of-the-art clustering algorithm developed for outlier detection in data streams that is applied using a sliding window over the most recent data [82]. As such, some parameters that affect the functionality of the algorithm and how it is used with streaming data must be defined. Parameters R and k define, respectively, the radius of the neighborhood and the minimum number of neighbors that must exist inside that radius for a point to be considered an inlier. The window size W constrains the amount of data that will be processed at each step, either as a time interval or as the number of datapoints, and the slide size S determines the speed/length of movement of the window. MCODE performs distance-based outlier detection, that is, a given object is considered an outlier if it has less than k neighbors inside radius R .

Asides from these studies, it is mentioned in studies [49, 51] and [39] that the proposed approaches are suitable or can potentially be used for online learning, but in none of them is that actually performed and described.

4 Discussion

The results presented in Section 3 will be discussed in more detail in this section with the purpose of identifying interesting trends and ideas. This section also aims to provide an overview of the challenges faced when using machine learning methods to detect mechanical faults and predict faults in real manufacturing scenarios and consider how future research efforts might address them.

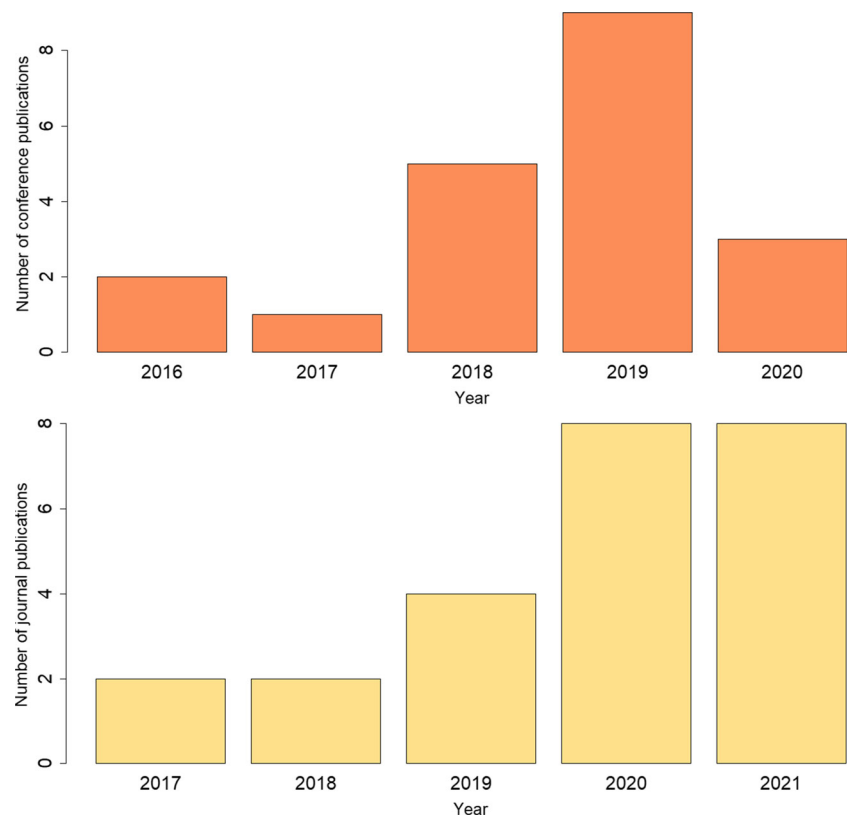
4.1 RQ1: In which publication venues are studies about the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment published?

As presented in Section 3.2, studies about the topic of interest have been published in a variety of conferences and journals, ranging from journals about operations research & management science to multidisciplinary conferences. While the journals considered in this review are all peer-reviewed journals with JCR Impact Factor (exclusion criterion 3), 58.8% of which are ranked Q1 [31, 59–61, 63–67, 73, 74], the same quality verification could not be performed for conferences since there is no ranking system that evaluates the quality of conferences across different scientific fields. However, it was observed that of the eleven computer science studies published in conferences, six were published in conferences ranked by the Computing Research and Education Association of Australasia (CORE). Two in conferences ranked C [41, 46], three in conferences ranked A [51, 69, 70] and one in a conference ranked A* [55].

As can be observed in Fig. 7, the number of publications in conferences decreased in 2017 in comparison with the previous year but increased considerably between 2017 and 2019. However, the number of publications in conferences decreased again in 2020 and there were no conference publications in 2021 (until October). On the other hand, while no study was published in journals before 2017, the number of journal publications has been increasing steadily since 2018 and the number of publications in 2021 has already equalled the number of studies published in 2020 (as of October 2021).

Figure 8 shows the distribution of publications in conferences and journals for the most prolific countries. While China published considerably more in journals than in conference proceedings, the opposite was true

Fig. 7 Publications in conferences and journals across the years



for Germany and Greece. However, since most countries contemplated in this review only published one or two studies it is difficult to discern which type of venue is favored.

4.2 RQ2: In which scientific fields has the use of machine learning for mechanical fault detection and fault prognosis in manufacturing equipment been researched?

It was shown in Section 3.3 that 21 of the 44 selected studies consist in contributions from the computer science community. This can be attributed to this review's focus

on machine learning, a discipline that arose from the intersection of computer science and statistics and is seen as a major branch of artificial intelligence [83, 84]. In addition, the presence of eleven engineering studies and six multidisciplinary studies is in line with the nature of fault detection and prognosis, which involves knowledge from different areas of engineering and computer science.

A more in-depth analysis reveals about as many computer science studies were published in conferences as in journals (Table 17). This can be attributed to the fact that, while most scientific fields prefer to publish in high quality journals, the computer science community typically favors publishing in prestigious conferences [85].

Fig. 8 Publications in conferences and journals for the top 3 countries

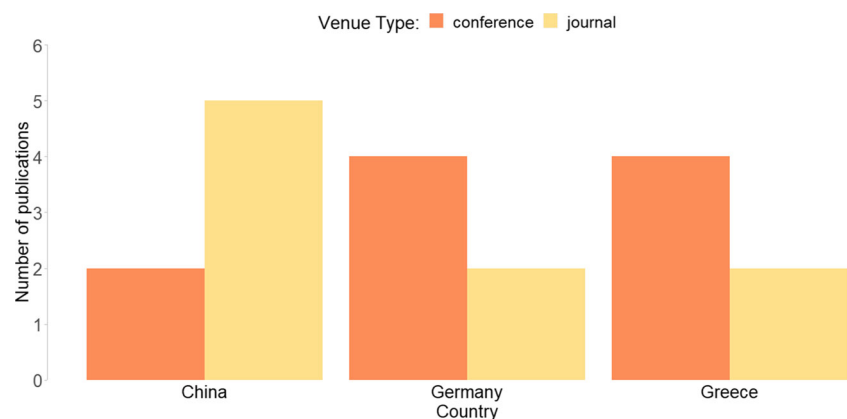


Table 17 Publication per scientific field and venue type

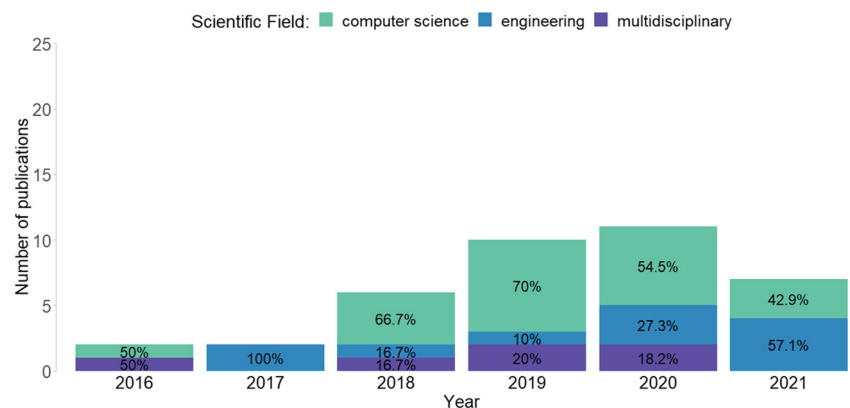
Scientific Field	Conference	Journal
Computer Science	11	10
Multidisciplinary	5	1
Engineering	1	10
Automation & Control Systems	2	0
Instruments & Instrumentation	0	2
Operations Research & Management Science	0	1
Wireless Communications, Networking and Signal Processing	1	0

This aspect is further supported by the fact that about 44% of the conferences where computer science studies were published have a CORE ranking of A or A*.

Most multidisciplinary studies and all the studies from the fields of automation & control systems and wireless communications, networking, and signal processing were published in conferences. On the contrary, the majority of engineering studies were published in journals, as were all of the studies from the remaining scientific fields.

Figure 9 shows the percentages of studies published per year by the top three scientific fields. Computer science studies were published almost every year between 2016 and 2021, except for 2017 when all the studies originated from the field of engineering. Although the largest proportion of publications between 2018 and 2020 came from the computer science community, the number of publications from this field has been decreasing since 2019. On the contrary, the number of engineering publications has been increasing since 2019 and, until October 2021, there were more publications from the field of engineering in 2021 than from computer science. The number of multidisciplinary studies published since 2018 has been very similar, with no notable increase or decrease in the number of yearly publications.

Fig. 9 Proportion of studies published per year by the top 3 scientific fields



As seen in Section 3.2, several studies were published in venues affiliated with the IEEE. This might be explained by the fact that 86.4% of the 44 selected publications consist in computer science, engineering or multidisciplinary studies, which are some of the areas of focus of that organization [86].

4.3 RQ3: What machine learning algorithms and methods are currently employed for mechanical fault detection and fault prognosis in manufacturing equipment?

In this subsection, the different algorithms and techniques presented in Section 3 are examined in more detail. The descriptions of the studies are organized according to the categories identified in Section 3.4: decision tree models, artificial neural networks, hybrid models, latent variable models and other approaches. Additionally, within each category, the studies are organized according to subcategories (where pertinent) and year of publication.

4.3.1 Decision Trees

Machine learning algorithms and methods belonging to the category of decision trees were some of the most commonly used for the tasks of detecting mechanical faults and predicting faults in manufacturing equipment in industrial environments. Ensemble methods, in particular, were widely used, as exemplified by the application of random forest models in 11.4% of the studies under consideration. In 2018, Amihai et al. [44] derived key asset health condition indices from raw vibration data and used a random forest model to forecast these metrics up to seven days ahead. A comparison of RMSE values for different look-ahead times demonstrated the random forest always performed better than a persistence model. In that same year, the authors of [46] tested different machine learning algorithms to predict equipment faults using process data from anode manufacturing machines. The best results were

obtained with a random forest model (accuracy = 99.2%; max_depth = 5-10) and with a decision tree model (accuracy = 99.2%; max_depth = 5), showing it was possible to predict faults 5 to 10 minutes before their occurrence. Paolanti et al. [43] implemented in 2018 a predictive maintenance system to predict the health status of the spindle's rotor of a CNC woodworking machine. To achieve this, the authors trained a random forest model on drive and vibration data collected from the machine to classify its condition into one of four classes, having obtained an average accuracy of 92%. In a study published in 2019, Binding2019, operational data and downtime data from a large central imprint printing press were used to predict failure events with a prediction horizon of 30 minutes. After analyzing the data, the authors focused on the prediction of mechanical failures in print units, such as leakages and deterioration of components. To achieve this, different classification models were trained and evaluated, namely logistic regression, random forest and extreme gradient boosted trees (XGBoost). Considering the F1-score for different decision thresholds, the random forest and XGBoost models yielded the best results, but the authors of the study chose to use the random forest model in the implementation of the predictive maintenance system. The predictive algorithms were also used to help identify print unit failures in the downtime data, in a manner similar to iterative semi-supervised labelling schemes. Also in 2019, the study described in, Aremu2019 used Kullback-Leibler divergence to construct a health indicator (HI) of multi-sensor systems to represent a system's deviation from its normal state. The usefulness of the HI for prognosis purposes was evaluated by comparing the RUL predictions for a semiconductor manufacturing equipment using the original data and the HI data. The results obtained using random forest regression and Gaussian process regression demonstrated the constructed HI always provided more accurate predictions, with the random forest model outdoing the Gaussian process in terms of RMSE (20.34 vs 24.7) and MAE (26.02 vs 28.63).

Other examples of decision tree ensembles include the work presented in [72] in 2017, where the authors described a procedure for fault prediction that leveraged cyclic manufacturing process data from similar work systems to improve the accuracy of the fault detection model. Since faulty cycles were rare, machine-to-machine (M2M) communication was used to acquire data from five injection moulding machines, thus increasing the amount of available fault data. To assess the effect of using data from several machines of the same type on model performance, three AdaBoost models, with decision stumps as estimators, were fitted to the data and evaluated using a machine-to-machine methodology (M2M): 1) a model trained and tested with data from all the work systems, 2) a model trained and tested with data from a single work system and 3) a

model trained with data from all the work systems except one whose data was used exclusively to test the model. Through a series of experiments, the authors demonstrated that the best performing model was the one trained and tested on data from all the machines (F1-score = 0.082), while the performance of the model tested with data that was not used for training was considerably worse (F1-score = 0.03). It should be noted that the performance of the proposed method depends on the degree of imbalance of the data. The F1-score results were low because the ratio of faulty to normal cycles (1:1484) was very low, but the model trained with data from all the machines performed considerably better than random guessing. Additionally, these results show that sharing data from similar work systems can improve the fault detection accuracy if data from the system of interest is also used to train the model. In 2020, the authors of [76] proposed a predictive tool for a cyber physical production system that uses GBDT to predict equipment failures in a CNC milling machine. Data was collected from the machine's central control system, as well as from external sensors used to monitor parameters such as vibration severity and amplitude. Rolling summary statistics of these variables, within 10, 30 and 60 second windows, were also added to the historical data. Additionally, information about the machine's operation mode, operator door mode (open or close) and program block number was used to infer when failure events occurred and label the data accordingly. GBDT was used initially to determine the relative importance of features and select the most appropriate time lag (60 sec) to use as input to dynamic principal component analysis (DPCA). DPCA was used to eliminate the autocorrelations present in the data and extract the principal components from the normalised 60-second lag feature space. This data was then fed to the GBDT to learn a binary classification model that predicted the probability of a production stop. Using the AUC score, the authors showed the predictive tool has an accuracy of 73% on unseen data.

Tree ensembles were also used for survival analysis [71] and anomaly detection [67]. In 2019, the study presented in [71] explored the rich data provided by the plant floor automation and information system (PFS) of a real-world automotive manufacturing line to learn complex and dynamic machine breakdown patterns. The authors proposed a manufacturing system-wide balanced random forest (MBRSF) model, whereby a random survival forest was used to estimate a hazard function from balanced system-wide data with the purpose of quantifying the likelihood of breakdown events over time. Experiments performed on 20 machines demonstrated the performance of the MBRSF was about 90% better, in terms of the integrated Briers score, than the performance of other survival models. In a study published in 2020, Kolokas

et al. [67] presented a methodology for fault prognosis that used an anomaly detection technique to predict faults from process data, but approached the problem as a case of binary classification. An isolation forest was used to detect anomalies in real industrial data, related to aluminum and plastic production, and correlate them with upcoming faults according to a predefined forecasting horizon. The model's performance was assessed using the Matthews correlation coefficient (MCC) to measure the correlation between the anomalies detected by the IF and the data's target labels, having obtained results up to $MCC = 0.73$.

CART models are also present, being the models of choice in 6.8% of the selected studies. In 2016, Linard and Bueno [41] described a new method for dynamic maintenance scheduling of large-scale printers. Labelled data obtained from printing test pages was used to train a decision tree that was deployed in real-time to predict whether failures would occur or not in the nozzles of the printers. The output of the decision tree was then used to update an automatic maintenance schedule defined by a timed automaton. The authors compared the performance of different classifiers but decided to use a decision tree not only because it provided the best results (precision = 0.788; recall = 0.631), but also due to its nature as a white box model since interpretability is particularly important in industrial contexts. In 2019, decision tree models were also used in [68] to estimate the failures of cold forging machines in an industrial company of the automotive industry. The decision tree model provided better results than the other evaluated algorithms, successfully predicting failures that occurred unexpectedly in the factory between 2014 and 2017 with an accuracy of 77%.

4.3.2 Artificial neural networks

Publications where artificial neural networks were employed for mechanical fault detection and fault prognosis account for more than a quarter of the studies under consideration (27.3%).

In 2016, Qing et al. [40] proposed a BPNN optimized by a multilevel genetic algorithm (MGA-BPNN) to predict the RUL of segment bearings in continuous casting equipment. The proposed model aimed to enhance the nonlinear learning and generalization abilities of the BPNN and thus obtain an improved forecasting model. Experimental results showed the MGA-BPNN model was better at predicting the RUL than either a BPNN or a BPNN optimized by a genetic algorithm and could be used as an effective means of fault prognosis.

In 2017, the authors of [57] describe a system framework for predictive maintenance based on industry 4.0 concepts. The system performed fault prognosis using an artificial neural network to uncover the hidden patterns of

degradation that led to a backlash error in a CNC machine center. After the model was trained using historical data, the artificial neural network was deployed in real-time to make predictions based on condition monitoring data. These predictions were used by a decision support system to formulate a maintenance strategy.

Luo et al. [31] proposed, in 2018, a method for early fault detection in CNC machine tools under time-varying conditions that relied on deep learning to identify impulse responses from vibration data. The deep learning model consisted of a layer of stacked sparse autoencoders (SSAE), meant to reduce the dimensionality of the input data, and a back-propagation neural network (BPNN) layer that classified the vibration signals into impulse and non-impulse responses (accuracy = 97.3%). The impulse responses selected by the deep learning model were used to identify the dynamic properties of the machine tool, which were then used to develop a health index that reflected the equipment's gradual deterioration process.

In the study presented in [70] in 2019, an LSTM network was built to predict faults in industrial ovens from sensor data and log events. The network was trained using consecutive time series and was used to predict the five subsequent future events, i.e., it predicted events 25 minutes into the future. Considering the data used to train the network was strongly imbalanced, its performance was assessed using the Matthews correlation coefficient (0.691), recall (0.790) and F1-score (0.803) as evaluation metrics. The results showed the values of the evaluation metrics decreased the further into the future a prediction was, but the network's performance was acceptable for all predictions.

Also in 2019, the fault diagnosis method proposed in [81] took advantage of digital twin technology to transfer fault information from the virtual entity to its physical counterpart. The digital twin consisted in a high-fidelity dynamic virtual model of a car body-side production line that simulated the entire product life cycle. This simulation data was used to build a diagnosis model that combined a SSAE layer to perform feature extraction from unsupervised data and a softmax classifier that used the extracted features as inputs and assigned probabilities to the class labels. Subsequently, deep transfer learning was used to relay the knowledge gained in the virtual space to a new fault diagnosis model built in the physical space. Monitoring data from the physical entity was used to improve the model, and an adaptation layer between the feature extraction and classification layers minimized the distance between the data distributions from the virtual space and the physical space. The virtual and physical entities of this digital twin-assisted fault diagnosis method cooperated with each other to provide accurate fault predictions (average accuracy = 97.96%) and adapt to new working conditions.

The study described in [78] introduced in 2020 a predictive maintenance framework to detect and classify the severity of mechanical faults in conveyor AC motors. Principal component analysis was used to reduce the dimensionality of time series data collected from the conveyor system to two channels, after which the data was encoded into images using the Gramian angular field method. The resulting images were used to train a convolutional neural network which outputted the “fault severity in the system”, i.e., based on the input images the CNN classified the motor’s state as “no fault”, “minor fault” or “critical fault”. To improve the model’s accuracy when using more extensive networks, the authors added the option to use a PReLU activation function instead of the more common rectified linear unit function (ReLU). The proposed approach was compared with an SVM and a CNN that used ReLU as its activation function. As shown in the experimental results, for small datasets the CNN’s performance was very similar using either PReLU or ReLU, with an accuracy of 100% in both cases. The SVM performed considerably worse, having obtained an accuracy of 55.2%.

In 2020, deep learning was also used to perform fault prognosis from times series data in [77]. The authors proposed a TensorFlow-enabled deep neural network to perform multiclass classification of the condition of a small trolley’s cylinder in an automobile production line. The performance of the proposed approach was compared to two other methods, namely PCA and HMM. The TensorFlow-enabled DNN performed better in all of the experiments, with an average accuracy of 80% versus 63% for the HMM and 50% for PCA. After training the DNN model offline using historical data, it was deployed in real-time to track the degradation of the equipment.

In another study presented in 2020 [56], a multilayer perceptron was used for fault prognosis of industrial packaging robots. Due to the facility’s lack of IoT technology, the data of interest, which consisted in failure notifications and associated information, was obtained from the enterprise resource planning (ERP) system. The MLP was composed of eight input nodes, 20 hidden layer nodes and four output nodes that indicated where and when a future failure would occur (accuracy = 91%). The authors of the study also performed a component-based reliability analysis whose results validated the MLP’s predictions (reliability = 75%).

Still in 2020, the method proposed by Das et al. combines two self-evolving recurrent neural networks to detect machine faults autonomously and in an online fashion [55]. The model, named SERMON, consists of two components: SERN, a Skip-connected Evolving Recurrent Network, and MERN, a Multilayer Evolving Recurrent Network. The two networks work in a cooperative manner to obtain better

results in terms of modelling the temporal dependency present in streaming data, and the ability to self-evolve allows them to adapt to the changes (drifts) that characterize non-stationary data. SERMON also includes a mapping unit (MU) that suggests possible data labels in the event of a delay in the arrival of the true label. SERMON was validated using data from a real-world industrial case study, i.e., to predict the condition of a 3D printing nozzle as either “healthy” or “clogged” based on nozzle shape features such as symmetry shape feature and slope feature, among others. The performance of SERMON was compared with seven other models in terms of classification rate, parameter count, hidden unit count and execution time and considered scenarios of no-delay, finite delay, and infinite delay in receiving labels. The classification rate and hidden unit count of SERMON was better than all the other models in all scenarios (average accuracy in a no-delay scenario: 72.08%; average accuracy in a finite/infinite delay scenario: 69.39%) and while one model (SkipE-RNN) obtained better results in terms of parameter count and execution time, its accuracy rate was always more than 10% lower than SERMON’s.

In 2021, Bampoula et al. [73] proposed an approach for fault detection and prediction based on LSTM-autoencoders. A prototype was tested in a steel production factory using three months of historical data obtained from a rolling mill machine and focused on the analysis of the surface temperatures and hydraulic forces of the machine’s two cylinders. The condition monitoring data obtained from the machine was segmented into time-series sequences according to three possible equipment health states, namely “good”, “bad” and “intermediate” operating conditions. Subsequently, each LSTM-autoencoder was trained using a dataset consisting only of time-series sequences corresponding to a given state. Afterwards, new data was fed to each LSTM-autoencoder and classified according to the highest accuracy obtained. That is, if the LSTM-autoencoder that obtained the highest accuracy was the one trained only with healthy data, the new data was classified as “good”. Finally, the authors considered the fatigue rate of the machine was constant and estimated the remaining useful life based on the classification accuracy. Performance results obtained with the prototype demonstrated that unnecessary preventive maintenance actions could be reduced, therefore decreasing the cost of maintenance operations.

In 2021, a PdM methodology was proposed in [79] that uses an LSTM-GAN to monitor the health state of machines, as well as predict when and in which machine a fault will occur. The proposed PdM methodology also includes a maintenance decision model that suggests maintenance operations according to the output of the prediction model. The methodology was tested in a

manufacturing factory located in China, where sensing devices monitored eight different machines (two automated guided vehicles, two robots, two milling machines and two turning machines) for over two years. Initially, the health state of the manufacturing system was predicted as being in one of four states: “good”, “watching”, “warning” and “fault”. If a machine was in “good” condition, no maintenance was required. If it was in either “watching” or “warning” states, a minor maintenance strategy would be implemented. In case of a “fault” state, the fault type and time of occurrence would be predicted by the fault prediction model and a major maintenance strategy would be implemented. Both the state prediction and the fault prediction abilities of the LSTM-GAN were compared with the results obtained using three other types of neural networks. The comparison analysis revealed the LSTM-GAN outperformed the other networks in both state prediction (average accuracy = 98.87%) and fault prediction (average accuracy = 98.92%).

In an additional study published in 2021 [66], the health model of a predictive maintenance system that takes into account time-varying operational conditions and allows for the subsequent scheduling of maintenance and production was introduced. The system uses condition monitoring sensor data, production data and future production orders to create a production schedule that incorporates the necessary maintenance actions. The proposed framework was validated in a real industrial use case with data from a multifunctional machining centre used to produce automotive components. The machine’s condition was assessed using two CVAE models: 1) HA-CVAE that takes as input condition monitoring data and the corresponding operating regime information and derives a set of health indexes that model the underlying trend of degradation under time-varying operational conditions, and 2) DS-CVAE, a data simulator used to generate realistic sensor data based on the conditional probability distribution learned from the training data for a specific operating regime and health state. The estimation of the health index was evaluated using metrics described in the prognostics and health management (PHM) literature to assess the trajectory of the health index over time (e.g., monotonicity, consistency). Validation experiments demonstrated that the proposed method was not only capable of estimating the machine’s health under different operating conditions and in a scenario where labelled data was scarce but was also able to predict the machine’s future health and degradation condition.

4.3.3 Hybrid models

Hybrid models integrate different machine learning models and techniques to solve problems that a single model is not

capable of handling, or to obtain better performance. 18.2% of the studies selected in this review made use of hybrid models, in some cases to address the absence of labeled data.

In 2018, Syafrudin et al. [74] proposed and described the implementation of a real-time monitoring system that combined IoT-based sensors, big data technology and a hybrid prediction model to predict faults in manufacturing equipment. The system was tested for 8 months in an automotive manufacturing company in Korea. In the proposed prediction model, the DBSCAN algorithm was first used to detect outliers in the sensor data that might represent noise introduced by problems in the sensing devices or by network connection issues. The detected outliers were removed from the dataset before it was used to train a random forest model which was then deployed in the monitoring system to perform fault prediction in real-time. When compared with other classification models, such as naive Bayes (accuracy = 93.57%), logistic regression (accuracy = 97.95%) and multilayer perceptron (accuracy = 96.78%), the DBSCAN + random forest hybrid model achieved better results (accuracy = 100%). Furthermore, the use of DBSCAN to remove noisy data improved the performance of the other classifiers as well, but the proposed model remained the best performing one.

In the same year, Strauß et al. [45] proposed a predictive maintenance approach that combined semi-supervised, unsupervised and supervised learning techniques to detect and classify mechanical faults in a heavy lift EMS at the BMW Group. Although fault data was available, the authors took into consideration the fact that this information is often scarce. As such, the problem of fault detection was initially approached using a semi-supervised method to perform anomaly detection. Three different models were built using normal data exclusively but were evaluated using a dataset containing both normal and fault data to assess their ability to detect data points that diverged from ‘normality’. Since there were several types of faults and the semi-supervised models were unable to distinguish between them, unsupervised models were used to cluster the fault data. By using semi-supervised and unsupervised models together, the authors were able to create a dataset that contained normal data, as well as instances of three different types of failures. This data was then used to train and evaluate eight supervised models, four of which had an F1-score of more than 90%. Model deployment in the predictive maintenance system took into consideration not only each model’s performance but also computational requirements - the final selected models were one-class SVM, K-means and random forest.

In 2019, the study presented in [50] proposed a fault prognosis model which combined an ARIMA model with a LSTM network to predict faults in a ball bearing automatic

production line. The ARIMA model was used to forecast the linear component of the time series (sensor data) collected from the production line, whereas the LSTM forecasted the nonlinear component obtained from the ARIMA model's prediction. The final predicted value resulted from the summation of the linear component with the prediction error of the nonlinear component of the ARIMA prediction. Experimental verification demonstrated that the proposed hybrid model performed better than either model by itself (MAE = 0.00425; RMSE = 0.03584).

As described in the artificial neural networks subsection, in 2019 Rousopoulou et al. [70] presented a solution for predictive maintenance of the industrial ovens used by a medical devices manufacturer. However, while the authors selected an LSTM network to predict faults from condition monitoring sensor data and log events, they decided to combine an outlier detection method with a classifier to detect faults in unlabeled acoustic data. Three outlier detection algorithms were tested and compared, namely DBSCAN, LOF and mean absolute deviation (MAD), with DBSCAN yielding the best results, i.e., it detected the most outliers. The detected outliers were marked as faulty, but since this class accounted for only 14% of the data, the synthetic minority oversampling technique (SMOTE) was used to increase the number of data points belonging to the minority class and balance the dataset. This data was then used to train an SVM capable of detecting new faults in live audio measurements with an accuracy of 85% and an F1-score of 0.86. A fault notification was issued by the system if the model detected five consecutive faults.

In another study published in [59], the development of a prognostic maintenance model in a context where no labeled data existed was described. The study was carried out at a German automotive manufacturer to address the situation where maintenance of a milling tool was performed subjectively by machine operators based on visual inspections. Although no labels were available, the proposed method was developed with the intention of uncovering latent information hidden in historic data that included maintenance and production records, control data and sensor data. After performing a thorough analysis of the data, taking domain knowledge into consideration, the authors approached the problem from two orthogonally related dimensions: 1) *time dimension*, that is, the time when a tool was replaced and 2) *condition dimension*, referring to information about damaged and undamaged tools that was inferred from the available data. Based on these dimensions, a 4-field matrix was defined to differentiate between correct and incorrect tool replacement decisions. Clustering techniques were applied along both dimensions to assign the data observations to the 4-field matrix. The *time dimension* was grouped using the agglomerative hierarchical clustering algorithm named

weighted pair-group method using centroid (WPGMC), with cluster one representing tool replacements that were performed at a late moment in the tool's lifetime and cluster two representing tool replacements that were performed early. When considering the *condition dimension*, time series' sequences were clustered into two groups using the MAD to measure the intensity of the sequences' oscillations. Sequences with a lower MAD value, i.e., weaker oscillations, were assigned to cluster one, while the remaining sequences were assigned to cluster two. The results obtained from clustering the data were orthogonally related and, based on that, the data observations were assigned to each of the quadrants of the 4-field matrix. Since time series' sequences of "type 1" in the 4-field matrix represented the replacement of an undamaged tool late in its lifetime, thus reflecting correct decisions made by the machine operators, this data was used to train and test a RNN to predict the tools' RUL. The RNN model was then used to predict the RUL of "type 3" observations (undamaged tools replaced too early), showing that using the prognostic model would have resulted in an extension of the tools' lifetimes for about one-third of these tool replacements.

In 2020, Tran et al. [80] described a method to detect drill faults from sound data recorded from a drill machine at Valmet AB in Sweden. To detect abnormalities in the drilling machine, sound data was collected when the drill was broken and when it was operating normally. The drill sounds were converted to images, specifically mel spectrogram images and scalogram images, and features were extracted from these images using a pre-trained CNN architecture – VGG19 architecture trained on the ImageNet dataset. NCA was then used to select the most representative features and reduce the dimensionality of the data. Afterwards, the features obtained from the mel spectrogram images were used to train several classifiers based on KNN and SVM, whose performance was compared to select the best model. The best overall accuracy was obtained by the Medium Gaussian SVM and the Quadratic SVM, but since the purpose of this study was the detection of broken drills, the selected model was the Medium Gaussian SVM, which attained an accuracy of 90.12% and a recall of 0.88 when classifying the broken sounds. A similar procedure was followed using the scalogram images, but in this instance the best performing classifier was the ensemble subspace KNN with an accuracy of 80.25%. These two approaches were compared with additional techniques and the results demonstrated that the proposed methods performed considerably better at classifying the drill sound signals.

In 2021, the authors of [58] proposed a framework based on autoencoders and simple linear regression to construct a health index that was subsequently used to

predict the RUL of industrial equipment. The proposed methodology is applicable to situations where no fault history data is available. To construct the health index, an autoencoder is used to learn the normal structure of the data. The health index consists in the difference between the input data and the reconstructed data which is calculated across all variables using the mean absolute error (MAE). Subsequently, the authors use simple linear regression to predict the trend of the health index as new data is fed into the system. If the trend of the health index increases and the slope parameter of the regression becomes large, it means the monitored equipment is displaying abnormal behaviour. The RUL is defined as the difference between the health index at the current time and the time at which a failure is predicted to occur. This methodology was applied in two industrial use-cases: a pump equipment and a robot arm. In the first case, data was collected from three different pumps and a small amount of failure data was obtained for purposes of model validation. Experimental results demonstrated the value of the health index rose before the occurrence of a failure, which was accurately predicted. For the robot arm use case, vibration sensors were attached to the edge of the arms of five different robots to collect data at an average rate of 1500 samples per day for periods of time between three and ten months. Once again, the proposed method was capable of correctly predicting the occurrence of faults. To verify the reliability of their proposal, the authors run additional experiments to predict the RUL with an isolation forest. A comparison of results based on the MAE and the root mean square error (RMSE) demonstrated the proposed method was better at predicting the RUL in all the experiments undertaken using data from both the pumps and the robot arms.

Also published in 2021, the study presented in [75] describes a methodology for fault detection and prediction in cold forming processes of a Phillips factory in the Netherlands using GMM, the FP-Growth algorithm and CBA-CB. In this study, information about the normal operating conditions of a press module was collected for over a year. The data included: material batch, maintenance logs and data from acoustic emission sensors. Using the matrix profile - a data structure for time series analysis -, two meta-time series were obtained from the acoustic emission data, which were used for anomaly detection, and for fault prediction using rule mining. For anomaly detection, the authors opted for a statistical approach since no labelled data was available that could guide the definition of the anomaly threshold. The matrix profile was also used for fault prediction by first mining salient subsequences to find common patterns. Subsequently, PCA was applied to the salient subsequences and the resulting principal components were clustered using GMM. The acoustic emission data was then segmented into non-overlapping time windows

and each pattern within a window was labelled according to previously discovered clusters. Finally, the acoustic emission data was integrated with the maintenance logs and FP-Growth was used to mine association rules from this data, which were then used to build a classifier using a modified version of the CBA-CB algorithm. The results obtained with the fault prediction module were compared with the performance of a majority classifier, which obtained a high micro F1-score but was unable to predict any events. The proposed method, on the other hand, was able to predict faults related to three of the four maintenance events of interest with a micro F1-score of 0.632. It should be noted these events are extremely infrequent, occurring less than 0.05% of the time.

4.3.4 Latent variable models

Publications that employ latent variable models account for 13.6% of the selected studies.

In 2018, Amruthnath et al. [47] described a methodology for mechanical fault detection of a furnace fan using unsupervised learning. The authors began by computing a 99.9% confidence interval using Hotelling's T-squared statistic, after which the data was clustered using a Gaussian mixture model fitted by expectation-maximization. The clusters obtained using the GMM model were then identified using the T-squared statistic and the maintenance crew's expert knowledge. Using only unlabeled vibration data, the proposed methodology was able to discover a healthy state, a faulty state, and a reset state (fan substitution).

The study presented in [65] in 2020 proposed a big data architecture for predictive maintenance. The fault detection system used sensor data, such as temperature and vibration, to predict failures in manufacturing equipment. Since the data was unlabeled, anomalies in the equipment were detected by a distributed version of PCA that was implemented using MapReduce. The output of the PCA model was combined with a deterministic mechanism that monitored the number of anomalies detected in a 5-minute time window to warn the factory engineers about an impending failure if the number of anomalies exceeded a certain threshold. The proposed architecture was tested from 2013 to 2018 in a real production environment.

In the same year, a framework was proposed in [54] that consists in a cognitive analytics-based approach for machine condition monitoring and anomaly detection. The authors used unsupervised learning and sensor fusion to continuously monitor the health of an industrial robot and detect anomalies in its operation, as well as predict the time when maintenance activities might be needed. The experimental study was performed using only data that reflected the robot's operation under normal conditions,

i.e., no information about faulty states was available. The data was obtained from three independent sources, namely the robot controller, an energy meter and accelerometer sensors. The proposed cognitive analytics framework involved performing several pre-processing tasks to the data obtained from each independent source so that it could be synchronized before it was fused and clustered using the k-means algorithm. Considering three positions for the robot arm (top, middle and bottom), the data was grouped into three clusters. Since this data represents the robot's operation under healthy conditions, the framework's AI engine model, which monitored data arriving in real-time, calculated the distance from the new data points to the centres of the clusters. If that distance exceeded a predefined threshold, the data point was flagged as an anomaly. The AI engine was also responsible for monitoring the deviation of the clusters' centres over time to determine when the robot might need maintenance actions.

Also in 2020, a method that combines FSC with an improved version of K-SVD was proposed in [60] to detect early weak faults in rotating machinery (or faults affected by strong background noise). IKSVD is an improved version of K-SVD that uses self-adaptive matching pursuit instead of optimal matching pursuit for sparse coding, which makes IKSVD considerably more adaptable and efficient. The proposed method was used to detect faults in a coal mill of a cement plant, which displayed a large vibration phenomenon related to the bearing pedestal. Traditional approaches using envelope demodulation spectrum and FSC could not extract the fault features of the rolling bearing, but the proposed method, which uses the IKSVD method to enhance the impulse feature components of the signal and FSC to extract the fault features from the denoised signal, was able to detect a fault in the small gear of the coal mill. Disassemblage of the mesh gears confirmed the presence of wear damage on the small gear tooth.

The study presented in [64] in 2021 describes a predictive maintenance approach that consists in predicting the current wear of a rotating metal bush at Tata Steel in the UK. Motivated by the lack of practical studies in industrial contexts, the authors of the study put together a predictive maintenance system and demonstrated how a data-driven model could be used for bearing wear prediction in situations where the data is scarce, high-dimensional and of poor-quality. To predict the condition of the bush, data was collected from the set-up sheet, the data warehouse and two sensor-based data sources that log process-related parameters. Three different models were applied to the data to assess their performance in predicting the condition of the bush, namely PLSR, an ANN and a random forest. RMSE and R^2 were used as performance metrics to compare the results obtained by each model. In experiments undertaken with different training sample sizes, the PLSR had the

largest R^2 and the smallest RMSE on average and, as such, was implemented in the predictive maintenance system deployed at Tata Steel to monitor the metal bush's condition. It should be noted that the purpose of this study was the prediction of the real-time condition of the bush and not its future condition. That is, the aim of the authors was developing a PdM system that predicted the bush's condition at a given moment and not the state of the bush after a period of time. This approach made sense because the condition of the component could only be determined at the end of each maintenance cycle, and, therefore, the factory personnel had no means of knowing if the bush really needed to be replaced at that time or if they could postpone (or anticipate) the replacement. By predicting the current condition of the bush, the proposed PdM approach could facilitate the transition from preventive maintenance to predictive maintenance and reduce maintenance costs at the factory.

In the same year, k-multi-dimensional time-series clustering (K-MDTSC), a modified version of k-means that is capable of handling multivariate time-series, was proposed to predict the wear of welding electrodes used in the body-in-white welding stage of a car manufacturing plant [62]. K-MDTSC is based on k-means but employs a generalized notion of the Euclidean distance to handle multi-dimensional time-series and addresses the issues k-means has with empty clusters. Moreover, K-MDTSC works directly with raw synchronous time-series, without requiring any transformation of the time-series data. To validate the proposed algorithm, voltage and current data was collected from the welding process performed by two robots at the body-in-white shop of the plant. After the data was pre-processed, K-MDTSC was used to cluster the multi-dimensional time-series and discover different welding profiles. The clusters were then characterized according to the wear and tear of the welding electrodes. After analyzing the results, the domain experts realized the wear of the electrodes wasn't having a negative impact on the welding process and that preventive maintenance operations were being performed before they were actually necessary. These preventive maintenance operations could, therefore, be postponed, resulting in a reduction in maintenance time and costs.

4.3.5 Other approaches

18.2% of the selected studies employed machine learning algorithms belonging to other categories, such as instance-based algorithms, rule-based models, or dynamic Bayes networks, to name a few.

The study presented in [42] in 2017 describes a predictive maintenance approach that combines sliding windows with a genetic algorithm and a hidden Markov model to estimate

and predict a hard masking deposition tool's long-term degradation. Since the available data was sampled at different rates, summary statistics were calculated over sliding windows to synchronize the data features. To handle the features' dynamic nature, a HMM was used to cluster the time series data and estimate the tool's degradation by considering past and present states of the tool's condition. The genetic algorithm was used in conjunction with the HMM to select the most suitable subset of features. Considering the tool's degradation was estimated and predicted in an unsupervised manner, the proposed method was evaluated using historical data and taking into consideration information provided by the maintenance experts of the semiconductor manufacturing company.

In 2019, Naskos et al. [69] proposed a method that was capable of detecting oil leakages in real time in the large tanks of a BENTELER Automotive factory. To detect outliers in real-time, the authors applied the micro-cluster continuous outlier detection (MCOD) algorithm to streams of sensor data. Additionally, domain knowledge of the production cycle was used to determine the operational status of the machinery and enhance the algorithm's performance. When compared with variants of the proposed method, including the application of MCO to raw data (no prior domain knowledge), the combination of MCO with domain knowledge obtained the best results.

In another study published in 2019, Graß et al. [39] described an anomaly detection method to detect faults in the fans of a reflow oven. Besides from the absence of labeled data, which conditioned the type of learning methods that could be applied to the problem, the authors also had to consider that different items were processed in the same production line. A reconfiguration of machine parameters occurs whenever the production of a new item begins, leading to different patterns of sensor measurements in the time series data that should not be interpreted as anomalies. To deal with this problem the authors began by clustering the data according to the different machine configurations. After this, for each cluster, the data was segmented, and suitable features were extracted for each segment. Finally, K-NN was used to define an anomaly threshold based on the mean distance between a given segment and its k nearest neighbors. The proposed approach was tested using seven years of historical data, successfully demonstrating its ability to detect fan malfunctions.

Continuing in 2019, the authors of [49] used a quantitative association rule mining method to predict the RUL of industrial equipment. The proposed algorithm, named "Rules 4 Rare Events" (R4RE), improves an algorithm previously proposed by the same authors, by allowing quantifications of the consequent item in closed intervals and integrating online pruning of the generated

rules. After being applied to sensor data collected from a real factory between October and December of 2018, the R4RE algorithm produced about 4500 rules that estimated the RUL (RUL-time) of the machines that were being monitored. Additionally, the authors used an expanded dataset to predict the RUL in terms of produced parts (RUL-parts), i.e., the number of units a machine can produce before a failure occurs. Measuring the RUL in terms of parts was considered by the authors as being a more robust measure, since it does not consider the periods when a machine was idle or turned off. When compared with other machine learning models, the R4RE model achieved the best results in the prediction of the RUL-time (RMSE = 34.2 ; MAE = 28.7; MAPE% = 20.1) and was among the top contenders in the prediction of the RUL-parts (RMSE = 668.7 ; MAE = 120.8; MAPE% = 3.76).

Still in 2019, Chen et al. [51] also used a rule-based method to predict the RUL of machinery. Specifically, the authors used a modified version of the eXtended Classifier System (XCS) to predict the RUL of a digital radio frequency matching box (RF-MB), a machine employed in the semiconductor manufacturing process. XCS is a rule-based machine learning method that can recalibrate its rule set through interaction with the environment. Whenever the rule set does not satisfy the current environmental condition, a special mechanism generates a new rule that matches it. However, XCS can only process binary input data and, as such, the authors applied a modified version of XCS (XCSR), which is capable of processing continuous-valued inputs. Moreover, since XCSR is a classifier, the estimation of the RF-MB's RUL was framed as a classification problem. Fisher discriminant analysis (FDA) was applied to the data to reduce the large number of variables, before using XCRS to predict the RUL with an accuracy of 97.3%.

In 2020, Ruiz-Sarmiento et al. [63] proposed a predictive model to estimate and predict the degradation of machinery used in the stainless-steel industry, specifically the drums of the heating coils of Steckel mills. The model consisted in a discrete Bayes filter that incorporated expert knowledge, configuration parameters, and real time sensor data. The expert knowledge was obtained from the factory specialists that helped identify suitable variables and interactions, as well as define the parameters that affected the machines' degradation. The DBF model was able to estimate the machinery's health status, but it was also used to simulate np manufacturing processes and predict the machinery's degradation after execution of those processes. The performance of the predictive model was evaluated and compared with other models using real data from a factory in Spain. The proposed model obtained the best results in all instances (average RMSE = 0.59).

The fault detection method of a predictive maintenance system developed for a mechanical metallurgy company is

described in another study published in 2020 [48]. Since labeled data was unavailable, the authors used a prediction-based anomaly detection technique to discover unusual occurrences in sensor data obtained from monitoring different CNC machines. An autoregressive integrated moving average (ARIMA) model was fit to each independent data feature and a 95% prediction interval was calculated for the model's forecast. Data points outside the bounds of the prediction interval were flagged as anomalies, but since isolated anomalies may not represent an impending fault a 30-minute time window was used to calculate the moving average of anomaly occurrences. The predictive system issued a fault alarm if the average of anomalies exceeded a user-defined value (default = 0.85). Additionally, since an imminent fault might affect more than one variable, the authors proposed a fault detection mechanism that correlated the anomalies detected for each variable and issued an alarm according to a threshold that considered the number of variables with correlated anomalies.

In 2021, Mohan et al. published a study describing a method that also uses an ARIMA model to help industries transition from industry 3.0 to industry 4.0 without having to undergo considerable structural changes [61]. The authors proposed using the ARIMA model to forecast the oil contamination level of a high-pressure sand moulding line in a foundry and subsequently calculate the RUL of the equipment. In this study, sensor data was collected from the hydraulic unit of the moulding line every three minutes and the data was used to forecast the oil's contamination level every three hours. The hydraulic unit's RUL was computed based on the model's 95% confidence level and on a threshold level for the oil contamination. Additionally, whenever the moving average of the oil contamination was greater than the threshold value, the window size used for calculating the moving average changed accordingly so that a warning message could be issued sooner. During the study period (September 2018 to December 2019), the breakdown time of the high-pressure moulding line was reduced by 84% and the number of breakdowns was reduced by 88%. Furthermore, the MTBF increased from 604 to 5349 minutes and the MTTR reduced from 83 to 46 minutes. Considering the downtime caused by the oil contamination specifically, the proposed approach managed to achieve zero downtime.

4.4 RQ4: What limitations and advantages do those algorithms and methods present?

As described in Section 3.5, a large variety of advantages was identified in the studies under consideration. In addition, some studies identified not only the advantages of the machine learning algorithms used but also of the methods developed to address a given problem.

Notwithstanding, some benefits were mentioned more frequently than others due to their importance in the context of fault detection and prognosis in industrial environments.

High performance [40, 44, 49, 50, 59, 63, 71, 73, 74, 76, 78, 79], as well as the ability to uncover complex non-linear relationships in the data [50, 57, 59, 66, 71, 77], were two of the reasons most frequently given for choosing an algorithm. The stochastic behavior of a manufacturing system and the intricate relationships between its components mean these systems are marked by unpredictability [87]. Additionally, in real-world scenarios different products are often manufactured in the same production line, requiring changes in machine configurations, components, and production materials [32]. These non-stationary conditions further complicate the task of detecting and predicting faults. It is, therefore, crucial that machine learning algorithms can discover the nonlinear and dynamic patterns that characterize these events. High performance is equally desirable and is directly related to an algorithm's ability to model the system. However, when detecting or predicting failures in the real-world the definition of performance must consider the tradeoff between false positives and false negatives. Depending on the business requirements, failure to predict a fault might have serious consequences, in which case false positives are preferable to false negatives. However, there are less critical situations where a false positive, which can imply unnecessary stoppages and use of resources, will be far more costly. This tradeoff must be carefully defined with the help of maintenance experts.

Computational efficiency is also seen as an important advantage for machine learning algorithms [40, 53, 60, 63, 67, 69, 76, 79]. While advanced algorithms can produce very good results, they can also be quite demanding in terms of computational resources. Artificial neural networks, for example, require a lot of CPU and GPU processing power. They also require a lot of memory, as do ensemble tree methods, and the amount of data being processed impacts the usage of computational resources as well. Furthermore, demanding computations can greatly increase energy consumption [88]. For these reasons, among others, the development of a predictive maintenance system requires significant investment [89]. As such, choosing machine learning algorithms that are computationally efficient can be a more cost-effective option, while also being more environmentally friendly, a concern whose importance and urgency cannot be understated.

Other advantages identified in more than one study include the algorithms' suitability for processing time series data, interpretability, ability to uncover fault patterns from unlabeled data, suitability for online learning, and usefulness for root cause analysis. Sensor data obtained from the monitorization of manufacturing equipment consists in time series data. Since time series are a

sequence of data points ordered by time that may have an internal structure, the methods used to analyze time series should ideally be capable of taking this structure into account. Many machine learning algorithms, however, are not suited for this task, which means a much greater feature engineering effort is necessary before they can be applied to time series data. For this reason, algorithms capable of handling raw time series data can be advantageous. A model's interpretability is also seen as beneficial [41, 54, 59, 64, 75], not only because it can be useful to debug or fine-tune a model, among other technical aspects, but also because understanding how a model came up with a result increases the user's trust in the model. It also enables better informed decision-making, and, in industrial contexts, it can assist in the identification of a fault's root cause. In fact, the latter was identified as an important characteristic in other studies as well [41, 51, 64].

Another quality that is often necessary is the ability to learn from unlabeled data, i.e., unsupervised learning [39, 42, 45, 47, 48, 50, 54, 58–62, 65, 66, 69–71]. For a variety of reasons, historical fault data can be hard to obtain. Faults might not occur very frequently, or they might not be logged correctly. It is also possible that records of these events exist but not for the same time periods as the available condition monitoring data. For whatever reason, the lack of historical fault data is a problem that arises frequently (e.g., labeled data was absent in 10 of the 44 selected studies - 22.7%). Even when labels are present, data representing the condition of machines during normal operation is usually much more abundant than fault events, which causes the representative classes to be imbalanced. When the class imbalance is extreme, applying techniques, such as resampling, to solve the problem might not produce viable results. In situations such as these, machine learning algorithms suitable for unsupervised learning can be used to extract knowledge from the data and assist in the detection of faults. However, validating these models is not a straightforward task if no historical data exists that can be used to evaluate their performance. In these circumstances, the development of unsupervised approaches should be guided by domain knowledge and tested in real production environments, as demonstrated in publications [47, 59, 60, 62, 65, 75].

As reported in Section 3.5, only eleven studies identified the limitations of the machine learning algorithms used [31, 40, 42, 47, 57, 58, 62, 69, 73, 76, 81]. While most of them are details specific to the chosen algorithms, such as the lack of interpretability of artificial neural networks, or the slow convergence speed of BPNNs, the decrease in model performance when the training and test data do not share the same distribution is a limitation with important implications for fault detection and prediction. This issue is related to

the concept of online learning and will be discussed in more detail in the next subsection.

4.5 RQ5: Which of those algorithms and methods are used for data stream learning?

In many real-world scenarios, like the manufacturing industry, the data generating processes are non-stationary, causing the distribution of the data to change over time in what is called concept drift [32, 90]. This means the historical data used to train a model and the data used to make predictions when the model is deployed come from different probability distributions, which affects the model's performance. This issue is particularly apparent in the context of predictive maintenance since monitoring data is acquired at a high frequency and arrives continuously in the form of data streams. To prevent learning models from becoming obsolete over time they have to be updated regularly with new input data [90, 91]. However, traditional machine learning models are trained using batches of data and are not adequate to process continuous flows of data, i.e., data streams. To handle this type of data, it is necessary to use machine learning algorithms capable of learning incrementally or through small batches of recent data. This type of learning, whereby algorithms process high-speed data while adapting to concept drifts, is known as online learning, or data stream learning [90, 92].

In the scope of this review, only two studies [55, 69] used a data stream learning algorithm, while three others [39, 49, 51] used algorithms suitable for online learning, but did not apply them in that context. Considering what has been said about the non-stationarity of manufacturing environments and how it brings about concept drift, it is reasonable to assume the performance of the approaches proposed in the other selected studies would degrade over time. One notable exception is the study presented in [81], where digital twin technology and transfer learning were used to address concept drift induced by changing working conditions. Other studies have used transfer learning to deal with different probability distributions in the training and testing data [32], but transfer learning alone isn't sufficient for continuous adaptation.

More studies focused on data stream learning techniques are necessary, particularly studies performed in real industrial environments, not only to validate the applicability of theoretical methods, but also to address several aspects of practical order, like pre-processing streaming data, detecting concept drift in semi-supervised and unsupervised settings, and handling legacy system, among others [92, 93]. As this systematic review has revealed, research of online learning techniques applied to fault detection and prognosis in the manufacturing industry is still in its infancy.

5 Conclusion

This study presents a systematic literature review of the machine learning methods used for mechanical fault detection and fault prognosis in manufacturing equipment in real-world scenarios. The review was conducted according to the PRISMA guidelines and the guidelines for software engineering systematic reviews described in [35]. Following the steps defined in the review protocol, an initial set of 4549 records published between January 2015 and October 2021 were identified (3377 without duplicates). After assessing them based on selection criteria, 44 primary studies were selected for inclusion in the systematic review. These studies were then examined in more detail based on five research questions aimed at characterizing the publication sources and scientific fields, the machine learning methods used, their advantages and limitations, and their application in the context of data stream learning.

About 84% of the selected studies employed machine learning techniques belonging to one of four categories: decision trees, artificial neural networks, hybrid models and latent variable models. However, although every study performed detection of mechanical faults or prognosis of faults in real manufacturing scenarios, each study is distinct in terms of the manufacturing context where the study was undertaken, the machinery for which faults were detected or predicted, and the characteristics of the data that was available. These differences are to be expected from industrial case-studies but made it difficult to compare the different techniques.

While the number of publications is considerably larger in the second half of the period considered for the review than in the first half, only 44 studies were selected for inclusion in this review from a preliminary group of 3377 (without duplicates). The literature on mechanical fault detection and fault prognosis in manufacturing equipment is extensive but, despite the economic, safety and environmental benefits predictive maintenance can provide, the number of studies performed in real-world manufacturing scenarios is still reduced. Studies developed under experimental conditions tend to disregard the numerous challenges presented by manufacturing environments, which raises questions about their applicability. Research interest in this topic of study seems to be increasing, but there are still several issues that need to be addressed.

An important problem that needs to be considered when performing fault detection and prognosis in the manufacturing industry is the inherent complexity of manufacturing systems and the time-varying properties of production processes. More research is needed to develop machine learning algorithms and methods that can handle noisy, non-stationary data and capture the nonlinear patterns of interaction between machinery components. A line of

research that can be pursued to deal with the issue of non-stationarity is online learning, also known as data stream learning. Online learning techniques that learn incrementally, or from small batches of recent data, are ideal to process high-speed streams of sensor data, while continuously adapting to the changes in the data's probability distribution caused by non-stationary environments (i.e., concept drift). Learning models that do not account for concept drift will eventually become outdated. As this review has shown, there is still a deficit of studies devoted to online learning methods, particularly where it relates to the detection of mechanical faults or prediction of faults in the manufacturing industry. As such, this line of research provides promising opportunities for future research.

Another concern common in real-world scenarios is the absence of labeled data, which restricts the learning task to unsupervised and semi-supervised methods. Due to this issue, almost half of the studies selected in this review employed unsupervised learning techniques, but more work is necessary not only to demonstrate the effectiveness of these models, but also to develop new methods capable of learning complex nonlinear relationships in the absence of labels, while adapting to concept drift. To successfully perform fault detection and prognosis in manufacturing environments, it is important to consider these factors collectively.

Predictive maintenance provides economic, safety and environmental benefits, but the development of a predictive maintenance system can be laborious and requires a significant upfront investment. To justify such an investment in terms of time and money, and derive benefits from it, it is essential that the models developed perform as accurately as possible, but it is also important to consider other aspects, such as computational efficiency or interpretability, in accordance with the business's needs.

Acknowledgements The present work has been developed under project PIANiSM (EUREKA - ITEA3: 17008; ANI|P2020 40125) and has received Portuguese National Funds through FCT (Portuguese Foundation for Science and Technology) under project UIDB/00760/2020 and Ph.D. Scholarship SFRH/BD/136253/2018.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Aboelmaged MG (2014) Predicting e-readiness at firm-level: An analysis of technological, organizational and environmental (TOE) effects on e-maintenance readiness in manufacturing firms. *Int J Inf Manag* 34(5):639–651. <https://doi.org/10.1016/j.ijinfomgt.2014.05.002>

2. Holmberg K, Adgar A, Arnaiz A, Jantunen E, Mascolo J (2010). In: Holmberg K, Adgar A, Arnaiz A, Jantunen E, Mascolo J, Mekid S (eds) *E-maintenance*. Springer, London
3. Muller A, Crespo Marquez A, Iung B (2008) On the concept of e-maintenance: Review and current research. *Reliab Eng Syst Saf* 93(8):1165–1187. <https://doi.org/10.1016/j.res.2007.08.006>
4. Mobley RK (2001) *Predictive Maintenance*. In: *Plant Engineer's Handbook*. Elsevier, pp 867–888
5. Sullivan G, Pugh R, Melendez AP, Hunt WD (2010) *Operations & maintenance best practices-a guide to achieving operational efficiency* (release 3). Technical Report, Pacific Northwest National Lab.(PNNL), Richland
6. Rødseth H, Schjølberg P (2016) Data-driven predictive maintenance for green manufacturing. In: *Proceedings of the 6th international workshop of advanced manufacturing and automation*. *Advances in Economics, Business and Management Research*. Atlantis Press, pp 36–41
7. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech Syst Signal Process* 20(7):1483–1510. <https://doi.org/10.1016/J.YMSSP.2005.09.012>, <https://www.sciencedirect.com/science/article/pii/S0888327005001512>
8. Kan MS, Tan ACC, Mathew J (2015) A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mech Syst Signal Process* 62:1–20. <https://doi.org/10.1016/j.ymssp.2015.02.016>, <https://www.sciencedirect.com/science/article/pii/S0888327015000898>
9. Zhang W, Yang D, Wang H (2019) Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Syst J* 13(3):2213–2227. <https://doi.org/10.1109/JSYST.2019.2905565>
10. Hwang I, Kim S, Kim Y, Seah CE (2010) A survey of fault detection, isolation, and reconfiguration methods. *IEEE Trans Control Syst Technol* 18(3):636–653. <https://doi.org/10.1109/TCST.2009.2026285>, <https://ieeexplore.ieee.org/abstract/document/5282515/>
11. Gertler JJ (2017) *Fault detection and diagnosis in engineering systems*. CRC Press. <https://www.taylorfrancis.com/books/9781351448796>
12. Boyes H, Hallaq B, Cunningham J, Watson T (2018) The industrial internet of things (IIoT): An analysis framework. *Comput Ind* 101:1–12. <https://doi.org/10.1016/j.compind.2018.04.015>
13. O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ (2015) Big data in manufacturing: a systematic mapping study. *J Big Data* 2(1). <https://doi.org/10.1186/s40537-015-0028-x>
14. Qin SJ (2009) Data-driven Fault Detection and Diagnosis for Complex Industrial Processes. In: *IFAC Proceedings Volumes*, vol 42. Elsevier, pp 1115–1125
15. Carvalho TP, Soares FAAMN, Vita R, Francisco RP, Basto JP, Alcalá SGS (2019) A systematic literature review of machine learning methods applied to predictive maintenance. *Comput Ind Eng* 137:106024. <https://doi.org/10.1016/j.cie.2019.106024>
16. Hu H, Tang B, Gong X, Wei W, Wang H (2017) Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks. *IEEE Trans Ind Inf* 13(4):2106–2116. <https://doi.org/10.1109/TII.2017.2683528>
17. Mathew V, Toby T, Singh V, Rao BM, Kumar MG (2017) Prediction of remaining useful lifetime (rul) of turbofan engine using machine learning. In: *2017 IEEE International Conference on Circuits and Systems (ICCS)*. IEEE, pp 306–311
18. Shao H, Jiang H, Wang F, Zhao H (2017) An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowl-Based Syst* 119:200–220. <https://doi.org/10.1016/j.knosys.2016.12.012>
19. Wang L, Zhang Z, Long H, Xu J, Liu R (2017) Wind turbine gearbox failure identification with deep neural networks. *IEEE Trans Ind Inf* 13(3):1360–1368. <https://doi.org/10.1109/TII.2016.2607179>
20. You D, Gao X, Katayama S (2015) Wpd-pca-based laser welding process monitoring and defects diagnosis by using fnn and svm. *IEEE Trans Ind Electron* 62(1):628–636. <https://doi.org/10.1109/TIE.2014.2319216>
21. Baptista M, Sankararaman S, de Medeiros IP, Nascimento Jr C, Prendinger H, Henriques EMP (2018) Forecasting fault events for predictive maintenance using data-driven techniques and arma modeling. *Comput Ind Eng* 115:41–53. <https://doi.org/10.1016/j.cie.2017.10.033>
22. Li C, Sanchez R-V, Zurita G, Cerrada M, Cabrera D, Vásquez RE (2015) Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing* 168:119–127. <https://doi.org/10.1016/j.neucom.2015.06.008>
23. Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A (2015) Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Trans Ind Inf* 11(3):812–820. <https://doi.org/10.1109/TII.2014.2349359>
24. Canizo M, Onieva E, Conde A, Charramendieta S, Trujillo S (2017) Real-time predictive maintenance for wind turbines using big data frameworks. In: *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, pp 70–77
25. Krishnakumari A, Elayaperumal A, Saravanan M, Arvindan C (2017) Fault diagnostics of spur gear using decision tree and fuzzy classifier. *Int J Adv Manuf Technol* 89(9-12):3487–3494. <https://doi.org/10.1007/s00170-016-9307-8>
26. Amruthnath N, Gupta T (2018) A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In: *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. IEEE, pp 355–361
27. Uhlmann E, Pontes RP, Geisert C, Hohwieler E (2018) Cluster identification of sensor data for predictive maintenance in a selective laser melting machine tool. *Procedia Manuf* 24:60–65. <https://doi.org/10.1016/j.promfg.2018.06.009>
28. Ahmad W, Khan SA, Islam MMM, Kim J-M (2019) A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models. *Reliab Eng Syst Safety* 184:67–76. <https://doi.org/10.1016/j.res.2018.02.003>
29. Yu J (2018) Tool condition prognostics using logistic regression with penalization and manifold regularization. *Appl Soft Comput* 64:454–467. <https://doi.org/10.1016/j.asoc.2017.12.042>
30. Haq R (2020) *Enterprise artificial intelligence transformation : a playbook for the next generation of business and technology leaders*. Wiley
31. Luo B, Wang H, Liu H, Li B, Peng F (2018) Early fault detection of machine tools based on deep learning and dynamic identification. *IEEE Trans Ind Electron* 66(1):509–518. <https://doi.org/10.1109/TIE.2018.2807414>
32. Bang SH, Ak R, Narayanan A, Lee YT, Cho H (2019) A survey on knowledge transfer for manufacturing data analytics, vol 104. Elsevier B.V.
33. Quintana G, Ciurana J (2011) Chatter in machining processes: A review. *Int J Mach Tools Manuf* 51(5):363–376. <https://doi.org/10.1016/j.ijmactools.2011.01.001>
34. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. In: *Journal of Clinical Epidemiology*, vol 62. Pergamon, pp e1–e34

35. Kitchenham B (2004) Procedures for performing systematic reviews. Keele, UK, Keele Univ 33(2004):1–26. https://www.researchgate.net/publication/228756057_Procedures_for_Performing_Systematic_Reviews
36. Gama J (2012) A survey on learning from data streams: Current and future trends, vol 1. <https://link.springer.com/content/pdf/10.1007/s13748-011-0002-6.pdf>
37. Gusenbauer M, Haddaway NR (2020) Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res Synthes Methods* 11(2):181–217. <https://doi.org/10.1002/jrsm.1378>
38. Subramanya SH, Lama B, Acharya KP (2020) Impact of COVID-19 pandemic on the scientific community. <https://doi.org/10.5339/QMJ.2020.21>
39. Grass A, Beecks C, Soto JAC (2019) Unsupervised Anomaly Detection in Production Lines. In: Beyerer JKO (ed) *Machine Learning For Cyber Physical Systems, ML4CPS 2018. Technologien für die intelligente Automation*, vol 9, pp 18–25
40. Liu Q, Zhang F, Liu M, Shen W (2016) A fault prediction method based on modified Genetic Algorithm using BP neural network algorithm. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 4614–4619
41. Linard A, Bueno MLP (2016) Towards Adaptive Scheduling of Maintenance for Cyber-Physical Systems. In: Margaria, T SB (ed) *Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques, Pt I, Lecture Notes in Computer Science*, vol 9952, pp 134–150
42. Kinghorst J, Geramifard O, Luo M, Chan HL, Yong K, Folmer J, Zou M, Vogel-Heuser B (2017) Hidden Markov model-based predictive maintenance in semiconductor manufacturing: A genetic algorithm approach. In: IEEE International Conference on Automation Science and Engineering. IEEE, pp 1260–1267
43. Paolanti M, Romeo L, Felicetti A, Mancini A, Frontoni E, Loncarski J (2018) Machine Learning approach for Predictive Maintenance in Industry 4.0. In: 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), pp 1–6
44. Amihai I, Gitzel R, Kotriwala AM, Pareschi D, Subbiah S, Sosale G, Strecker C (2018) An Industrial Case Study Using Vibration Data and Machine Learning to Predict Asset Health. In: Proper HA (ed) 2018 20TH IEEE INTERNATIONAL CONFERENCE ON BUSINESS INFORMATICS (IEEE CBI 2018), VOL 1. Conference on Business Informatics. IEEE; IEEE Comp Soc, pp 178–185
45. Strauß P, Schmitz M, Wöstmann R, Deuse J (2018) Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoTArchitecture and Machine Learning. In: Abe N, Liu H, Pu C, Hu X, Ahmed N, Qiao M, Song Y, Kossmann D, Liu B, Lee K, Tang J, He J, Saltz J (eds) *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*. IEEE; IEEE Comp Soc; Expedia Grp; Baidu; Squirrel AI Learning; Ankura; Springer, IEEE International Conference on Big Data, pp 1474–1483. <https://doi.org/10.1109/BigData.2018.8622076>
46. Kolokas N, Vafeiadis T, Ioannidis D, Tzovaras D (2018) Forecasting faults of industrial equipment using machine learning classifiers. In: Yildirim, T ML (ed) 2018 IEEE (SMC) International Conference on Innovations in Intelligent Systems and Applications, INISTA 2018. Aristotle Univ Thessaloniki; Democritus Univ Thrace; IEEE Systems & Cybernet Soc; IEEE; Yildiz Techn Univ
47. Amruthnath N, Gupta T (2018) Fault class prediction in unsupervised learning using model-based clustering approach. In: 2018 International Conference on Information and Computer Technologies, ICICT 2018. IEEE, pp 5–12
48. Fernandes M, Canito A, Corchado JM, Marreiros G (2020) Fault detection mechanism of a predictive maintenance system based on autoregressive integrated moving average models. In: Herrera F, Matsui K, Rodríguez-González S (eds) *Distributed Computing and Artificial Intelligence*, 16th International Conference. Springer International Publishing, Cham, pp 171–180
49. Christou IT (2019) Avoiding the Hay for the Needle in the Stack: Online Rule Pruning in Rare Events Detection. In: 2019 16th International Symposium on Wireless Communication Systems (ISWCS), pp 661–665
50. Cheng C, Zhang B, Gao D (2019) A Predictive Maintenance Solution for Bearing Production Line Based on Edge-Cloud Cooperation. In: 2019 Chinese Automation Congress (CAC), pp 5885–5889
51. Chen L-Y, Lee J-H, Yang Y-L, Yeh M-T, Hsiao T-C (2019) Predicting the Remaining Useful Life of Plasma Equipment through XCSR. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '19*. Association for Computing Machinery, New York, pp 1263–1270. <https://doi.org/10.1145/3319619.3326879>
52. Aremu OO, O'Reilly DO, Hyland-Wood D, McAree PR (2019) Kullback-leibler divergence constructed health indicator for data-driven predictive maintenance of multi-sensor systems. In: IEEE International Conference on Industrial Informatics (INDIN). Inst Elect & Elect Engineers; Tampere Univ; Finnish Soc Automat; IEEE Ind Elect Soc, pp 1315–1320
53. Binding A, Dykeman N, Pang S (2019) Machine learning predictive maintenance on data in the wild. In: IEEE 5th World Forum on Internet of Things, WF-IoT 2019 - Conference Proceedings. IEEE; Univ Limerick; IEEE Commun Soc; IEEE Consumer Elect Soc; IEEE Reliabil Soc; IEEE Sensors Council; IEEE Signal Proc Soc; IEEE Stand Assoc; IEEE Control Syst Soc; IEEE Council Elect Design Automat; IEEE Council RFID; IEEE Electromagnet Compatibil Soc.; pp 507–512
54. Farbiz F, Miaolong Y, Yu Z (2020) A cognitive analytics based approach for machine health monitoring, anomaly detection, and predictive maintenance. In: 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp 1104–1109
55. Das M, Pratama M, Tjahjowidodo T (2020) A self-evolving mutually-operative recurrent network-based model for online tool condition monitoring in delay scenario. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp 2775–2783
56. KOCA O, Kaymakci OT, Mercimek M (2020) Advanced Predictive Maintenance with Machine Learning Failure Estimation in Industrial Packaging Robots. In: 2020 International Conference on Development and Application Systems (DAS), pp 1–6
57. Li Z, Wang Y, Wang K-S (2017) Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. *Adv Manuf* 5(4, SI):377–387. <https://doi.org/10.1007/s40436-017-0203-8>
58. Kim D, Lee S, Kim D (2021) An applicable predictive maintenance framework for the absence of run-to-failure data. *Appl Sci* 11(11):5180. <https://doi.org/10.3390/app11115180>
59. Zschech P, Heinrich K, Bink R, Neufeld JS (2019) Prognostic Model Development with Missing Labels: A Condition-Based Maintenance Approach Using Machine Learning. *Bus Inf Syst Eng* 61(3):327–343. <https://doi.org/10.1007/s12599-019-00596-1>
60. Wang H, Du W (2020) Fast spectral correlation based on sparse representation self-learning dictionary and its application in fault diagnosis of rotating machinery. *Complexity*. <https://doi.org/10.1155/2020/9857839>
61. Mohan TR, Roselyn JP, Uthra RA, Devaraj D, Umachandran K (2021) Intelligent machine learning based total productive

- maintenance approach for achieving zero downtime in industrial machinery. *Comput Ind Eng* 157:107267. <https://doi.org/10.1016/j.cie.2021.107267>
62. Giordano D, Mellia M, Cerquittelli T (2021) K-mdtsc: K-multi-dimensional time-series clustering algorithm. *Electronics* 10(10):1166. <https://doi.org/10.3390/electronics10101166>
 63. Ruiz-Sarmiento J-R, Monroy J, Moreno F-A, Galindo C, Bonelo J-M, Gonzalez-Jimenez J (2020) A predictive model for the maintenance of industrial machinery in the context of industry 4.0. *Eng Appl Artif Intell* 87:103289. <https://doi.org/10.1016/j.engappai.2019.103289>, <http://www.sciencedirect.com/science/article/pii/S0952197619302489>
 64. Chen X, Van Hillegersberg J, Topan E, Smith S, Roberts M (2021) Application of data-driven models to predictive maintenance: Bearing wear prediction at tata steel. *Expert Syst Appl* 186:115699. <https://doi.org/10.1016/j.eswa.2021.115699>
 65. Yu W, Dillon T, Mostafa F, Rahayu W, Liu Y (2020) A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance. *IEEE Trans Ind Inf* 16(1):183–192. <https://doi.org/10.1109/TII.2019.2915846>
 66. Zhai S, Gehring B, Reinhart G (2021) Enabling predictive maintenance integrated production scheduling by operation-specific health prognostics with generative deep learning. *J Manuf Syst*. <https://doi.org/10.1016/j.jmsy.2021.02.006>
 67. Kolokas N, Vafeiadis T, Ioannidis D, Tzovaras D (2020) A generic fault prognostics algorithm for manufacturing industries using unsupervised machine learning classifiers. *Simul Model Pract Theory* 103:102109. <https://doi.org/10.1016/j.simpat.2020.102109>, <http://www.sciencedirect.com/science/article/pii/S1569190X20300472>
 68. Turkoglu B, Komesli M, Unluturk MS (2019) Application of Data Mining in Failure Estimation of Cold Forging Machines: An Industrial Research. *Stud Inf Control* 28(1):87–94. <https://doi.org/10.24846/v28i1y201909>
 69. Naskos A, Gounaris A, Metaxa I, Koechling D, Stirna J (2019) Detecting anomalous behavior towards predictive maintenance. In: Proper HA (ed) *Advanced Information Systems Engineering Workshops (CAISE 2019)*, Lecture Notes in Business Information Processing, vol 349, pp 73–82
 70. Rousopoulou V, Nizamis A, Giugliano L, Haigh P, Martins L, Ioannidis D, Tzovaras D, Stirna J (2019) Data Analytics Towards Predictive Maintenance for Industrial Ovens A Case Study Based on Data Analysis of Various Sensors Data. In: Proper HA (ed) *Advanced Information Systems Engineering Workshops (CAISE 2019)*, Lecture Notes in Business Information Processing, vol 349, pp 83–94
 71. Bukkapatnam STS, Afrin K, Dave D, Kumara SRT (2019) Machine learning and AI for long-term fault prognosis in complex manufacturing systems. *CIRP Ann-Manuf Technol* 68(1):459–462. <https://doi.org/10.1016/j.cirp.2019.04.104>
 72. Vrabčič R, Kozjek D, Butala P (2017) Knowledge elicitation for fault diagnostics in plastic injection moulding: A case for machine-to-machine communication. *CIRP Ann Manuf Technol* 66(1):433–436. <https://doi.org/10.1016/j.cirp.2017.04.001>, <http://www.sciencedirect.com/science/article/pii/S000785061730001X>
 73. Bampoula X, Siaterlis G, Nikolakis N, Alexopoulos K (2021) A deep learning model for predictive maintenance in cyber-physical production systems using lstm autoencoders. *Sensors* 21(3):972. <https://doi.org/10.3390/s21030972>
 74. Syafrudin M, Alfian G, Fitriyani NL, Rhee J (2018) Performance Analysis of IoT-Based Sensor, Big Data Processing, and Machine Learning Model for Real-Time Monitoring System in Automotive Manufacturing. *Sensors* 18(9). <https://doi.org/10.3390/s18092946>
 75. Avendano DN, Caljouw D, Deschrijver D, Van Hoecke S (2021) Anomaly detection and event mining in cold forming manufacturing processes. *Int J Adv Manuf Technol*:1–16. <https://doi.org/10.1007/s00170-020-06156-2>
 76. Zhang Y, Beudaert X, Argandoña J, Ratchev S, Munoa J (2020) A cpps based on gbdt for predicting failure events in milling. *Int J Adv Manuf Technol* 111(1):341–357. <https://doi.org/10.1007/s00170-020-06078-z>
 77. Chen B, Liu Y, Zhang C, Wang Z (2020) Time Series Data for Equipment Reliability Analysis With Deep Learning. *IEEE Access* 8:105484–105493. <https://doi.org/10.1109/ACCESS.2020.3000006>
 78. Kiangala KS, Wang Z (2020) An effective predictive maintenance framework for conveyor motors using dual time-series imaging and convolutional neural network in an industry 4.0 environment. *IEEE Access*:1. <https://doi.org/10.1109/ACCESS.2020.3006788>
 79. Liu C, Tang D, Zhu H, Nie Q (2021) A novel predictive maintenance method based on deep adversarial learning in the intelligent manufacturing system. *IEEE Access* 9:49557–49575. <https://doi.org/10.1109/ACCESS.2021.3069256>
 80. Tran T, Lundgren J (2020) Drill fault diagnosis based on the scalogram and mel spectrogram of sound signals using artificial intelligence. *IEEE Access* 8:203655–203666. <https://doi.org/10.1109/ACCESS.2020.3036769>
 81. Xu Y, Sun Y, Liu X, Zheng Y (2019) A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning. *IEEE Access* 7:19990–19999. <https://doi.org/10.1109/ACCESS.2018.2890566>
 82. Kontaki M, Gounaris A, Papadopoulou AN, Tsihlias K, Manolopoulos Y (2016) Efficient and flexible algorithms for monitoring distance-based outliers over data streams. *Inf Syst* 55:37–53. <https://doi.org/10.1016/j.is.2015.07.006>, <http://www.sciencedirect.com/science/article/pii/S0306437915001349>
 83. Luxton DD (2015) *Artificial Intelligence in Behavioral and Mental Health Care*. Technical Report, <https://www.sciencedirect.com/science/article/pii/B9780124202481000015>
 84. Mitchell TM (2006) *The Discipline of Machine Learning*. Technical Report, <http://www-cgi.cs.cmu.edu/~tom/pubs/MachineLearningTR.pdf>
 85. Meyer B, Choppy C, Staunstrup J, van Leeuwen J (2009) Viewpoint research evaluation for computer science. *Commun ACM* 52(4):31–34. <https://doi.org/10.1145/1498765.1498780>
 86. (2021). History of IEEE. <https://www.ieee.org/about/ieee-history.html>
 87. Eftymiou K, Pagoropoulos A, Papakostas N, Mourtzis D, Chryssolouris G (2012) Manufacturing systems complexity review: Challenges and outlook. In: *Procedia CIRP*, vol 3. Elsevier B.V., pp 644–649
 88. García-Martín E, Rodrigues CF, Riley G, Grahn H (2019) Estimation of energy consumption in machine learning. *J Parallel Distrib Comput* 134:75–88. <https://doi.org/10.1016/j.jpdc.2019.07.007>
 89. Mulders M, Haarman M (2017) *Predictive Maintenance 4.0, Predict the unpredictable* (PwC Publication). Technical Report, PwC and Mainnovation
 90. Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A (2014) A survey on concept drift adaptation. *ACM Comput Surv* 46(4):1–37. <https://doi.org/10.1145/2523813>, <http://dl.acm.org/citation.cfm?doid=2597757.2523813>
 91. Tsybmal A (2004) *The problem of concept drift: definitions and related work*. Computer Science Department, Trinity College Dublin.

92. Gomes HM, Read J, Bifet A, Barddal JP, Gama J (2019) Machine Learning for Streaming Data: State of the Art, Challenges, and Opportunities. SIGKDD Explor Newsl 21(2):6–22. <https://doi.org/10.1145/3373464.3373470>
93. Kreml G, Žliobaite I, Brzeziński D, Hüllermeier E, Last M, Lemaire V, Noack T, Shaker A, Sievi S, Spiliopoulou M, Stefanowski J (2014) Open challenges for data stream mining research. ACM SIGKDD Explor Newsl 16(1):1–10. <https://doi.org/10.1145/2674026.2674028>, <https://dl.acm.org/citation.cfm?id=2674028>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Marta Fernandes is a researcher with the Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD) in Porto, Portugal and has been an invited assistant professor with the School of Engineering, Polytechnic Institute of Porto since 2017. Her research interests include machine learning, data science, data stream learning, and fault detection and prediction. She is currently pursuing a Ph.D.

degree in computer science at the University of Salamanca, Spain and has received a Ph.D. Research Scholarship from the Portuguese Foundation for Science and Technology.



Juan Manuel Corchado was the vice president for research and technology transfer from 2013 to 2017 and the director of the Science Park with the University of Salamanca, Spain, where he was also the director of the Doctoral School until 2017. He has been elected twice as the Dean of the Faculty of Science, University of Salamanca. He has been a visiting professor with the Osaka Institute of Technology since 2015 and a visiting professor with University

Technology Malaysia since 2017. He is currently the director of the Bioinformatics, Intelligent Systems, and Educational Technology (BISITE) Research Group, which he created in 2000. He is also the president of the IEEE Systems, Man and Cybernetics Spanish Chapter and the academic director of the Institute of Digital Art and Animation, University of Salamanca, where he is also a full professor. He also oversees the masters programs in digital animation, security, mobile technology, community management, and management for TIC Enterprises with the University of Salamanca. He is also a member of the Advisory Group on Online Terrorist Propaganda of the European Counter Terrorism Centre (EUROPOL). He is also an editor and the editor-in-chief of specialized journals, such as the *Advances in*

Distributed Computing and Artificial Intelligence Journal, the International Journal of Digital Contents and Applications, and the Oriental Journal of Computer Science and Technology.



Goreti Marreiros is a professor in the Department of Informatics Engineering at the School of Engineering of the Polytechnic Institute of Porto (ISEP). She is also subdirector of the masters degree in engineering and artificial intelligence, director of the post-graduation in Big Data & Decision Making in ISEP and member of ISEPs Scientific Council. In recent years, she has taught and been responsible for several courses on decision support systems, multi-

agent systems and artificial intelligence at the graduate and post-graduate level. She has also participated in 15 projects as Principal Investigator (8 European projects - H2020; ITEA and ERASMUS+ - and 7 national projects - FCT; P2020) and in 20 projects as Researcher. Ms. Marreiros is the vice-president of the Portuguese Association of Artificial intelligence.