AI and Ethics (2022) 2:553–577
https://doi.org/10.1007/s43681-021-00132-6

REVIEW

Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis

Aurelien Teguede Keleko[1,2] · Bernard Kamsu-Foguem[1] · Raymond Houe Ngouna[1] · Amèvi Tongne[1]

Received: 5 October 2021 / Accepted: 30 December 2021 / Published online: 10 March 2022

Abstract

The purpose of this article is to study the issues of industrial maintenance, one of the critical drivers of Industry 4.0 (I4.0), which has contributed to the advent of new industrial challenges. In this context, predictive maintenance 4.0 (PdM4.0) has seen a significant progress, providing several potential advantages among which: increase of productivity, especially by improving both availability and quality and ensuring cost-saving through automated processes for production systems monitoring, early detection of failures, reduction of machine downtime, and prediction of equipment life. In the research work carried out, we focused on bibliometric analysis to provide beneficial guidelines that may help researchers and practitioners to understand the key challenges and the most insightful scientific issues that characterize a successful application of Artificial Intelligence (AI) to PdM4.0. Even though, most of the exploited articles focus on AI techniques applied to PdM, they do not include predictive maintenance practices and their organization. Using Biblioshiny, VOSviewer, and Power BI tools, our main contribution consisted of performing a Bibliometric study to analyze and quantify the most important concepts, application areas, methods, and main trends of AI applied to real-time predictive maintenance. Therefore, we studied the current state of research on these new technologies, their applications, associated methods, related roles or impacts in developing I4.0. The result shows the most common productive sources, institutes, papers, countries, authors, and their collaborative networks. In this light, American and Chinese institutes dominate the scientific debate, while the number of publications in I4.0 and PdM4.0 is exponentially growing, particularly in the field of data-driven, hybrid models, and digital twin frameworks applied for prognostic diagnostic or anomaly detection. Emerging topics such as Machine Learning and Deep Learning also significantly impacted PdM4.0 development. Subsequently, we analyzed factors that may hinder the successful use of AI-based systems in I4.0, including the data collection process, potential influence of ethics, socio-economic issues, and transparency for all stakeholders. Finally, we suggested our definition of trustful AI for I4.0.

Keywords Bibliometrics · Industry 4.0 · Predictive maintenance · Anomaly detection · Prognostics · Condition monitoring · Artificial intelligence · Machine learning · Deep learning · Ethic · Trustful AI

1 Introduction

Nowadays, manufacturers are facing increasing global competition on various strategies, and requirements such as reduction of production costs, ensuring quality and innovation of products [1]. Consequently, these manufacturers need resort to Industry 4.0 to remain competitive, and meeting its new challenges. According to [2], the 4th Industrial Revolution can be defined as a set of interconnected digital assets, and technologies that contribute to develop, automate, integrate, and exchange real-time data in the manufacturing process. The author of [3] defines it as the integration of several technologies such as sensors, cloud computing, cybersecurity, simulation, artificial intelligence (AI), Internet of

✉ Aurelien Teguede Keleko
aurelien.keleko_teguede@enit.fr

Bernard Kamsu-Foguem
bernard.kamsu-foguem@enit.fr

Raymond Houe Ngouna
raymond.houe-ngouna@enit.fr

Amèvi Tongne
amevi.tongne@enit.fr

1 National Engineering School of Tarbes-ENIT, Production EngineeringLaboratory, 47 Avenue d'Azereix, 65000 Tarbes, France

2 National Polytechnic Institute of Toulouse, 6 Allée Emile Monso, 31400 Toulouse, France

Things (IoT), Big data, or robotics. This new industry, therefore, meets the new requirements such as the digitalization of factories using cyber-physical systems, or communicating sensors [4]; the flexibility of the factory, and the production customization [5]; the use of logistics tools that favor, and optimize the exchange of information [6]; the use of simulation techniques for configuring the production system, and making the scheduling of activities more flexible [7]. The factory must be energy, and raw material-efficient [8], must respect some constraints be it socio-economic, ecological, and political [9]. Also, I4.0 promotes the training of the different actors [10], and the implementation of an economic strategy to be more competitive [11].

According to [12–14], the factory must be digitized to meet its new challenges. Thus, the increasing exploitation of industrial production systems, thanks to the presence of IoT, sensors, cloud computing, the widespread use of distributed control systems, and AI techniques have greatly contributed to the spread, and development of I4.0 [15]. Paper [14] shows that big data, and data mining have an essential role in this development. At the same time, according to [2, 14], there are nine main pillars of technological progress that form the foundation of I4.0. Within the broad research fields related to the works mentioned above, we focus mainly on Predictive Maintenance in the context of I4.0. Predictive Maintenance 4.0 (PdM4.0) is the study of trends, behavior pat terns, and correlations using some models, and real-time analysis. PdM4.0 is based on three fundamental steps (i) exploiting data collected; (ii) modeling, using different approaches among which data-driven, model-based, or a hybrid approach which combines the two previous ones; (iii) exploitation of knowledge for decision-making, and control of the physical phenomenon studied. Therefore, the resulting models allow extracting insights to anticipate breaking points and possible mechanical failures. They thus favor the decision-making process for maintenance activities to avoid downtime [16]. In this context, the industry can be transformed into a predictive industry [17]. Furthermore, its innovative technologies combined with machine condition monitoring systems offer new management opportunities [18], control, improvement of the efficiency, and reliability of industrial systems [19]. It should be noted that in most cases, productivity decreases are often due to anomalies or machine degradation, especially when they have not been detected. To that end, PdM4.0, machine condition monitoring, and AI has therefore become an important research area in I4.0 [20], which constitutes the focus of the present research study.

The rest of this paper is organized as follows: Sect. 2 deals with the contributions, objectives, and main issues of the study. Section 3 shows a brief description of the Industrial Revolution, different approaches to solve predictive maintenance challenges and potential ethical impacts related to the use of AI technologies, for PdM in industry 4.0. Most common predictive models used, especially AI-based modeling applied in Industry 4.0, is detailed in the Sect. 4. Section 5 describes the research methodology used, and the process of collecting scientific publications for the analyses conducted. A detailed and in-depth bibliometric analysis is carried out and presented in Sect. 6, followed by the discussion and main contributions of the research work in the Sect. 7. Finally, the conclusion of the study, the limitations, and the future research envisaged being described in Sect. 8.

2 Contributions and research objectives of the papers

Industry 4.0 and Predictive Maintenance have impacts on most aspects of the business value. In that respect, several bibliometric studies have been carried out to analyze these impacts. For example, several reviews concentrate on the impact of digitalization in specific sectors such as management, economics, or ecology in the literature. While [21] focuses on the different approaches and main topics related to I4.0, the author of [22] address decision-making based on system reliability in the context of I4.0. Furthermore, [23] shows the current trends of I4.0 via a comparative study with WoS, and Scopus databases. The authors [24, 25] explore the elements surrounding I4.0, and their developments in the socio-economic, service industry, and management context. Also [26] presents the challenges, and raises the relationships between sustainability, and I4.0. The authors [27, 28] focus on emerging techniques, and trends in equipment maintenance systems, while [29] presents the evolution of AI. Article [30] describes a literature review on Machine Learning for industrial applications. Authors of [31], carry out a bibliometric study mainly focused on the detection of bearing defects when using AI. The field of industrial maintenance is vast and includes several subfamilies maintenance methods or approaches. In our opinion, few bibliometrics studies deal with real-time predictive maintenance in the context of I4.0, which is the main focus of the present work. This targeted field of research allows us to identify potential anomalies in production to reduce machine downtime (among several other objectives). However, the development and performance of PdM4.0 systems can be hindered by several factors that we consider in our study.

Our paper provides a bibliometric analysis of the different AI techniques applied to PdM for that purpose. Furthermore, the article asks questions such as: What are the current trends of the AI models, methods, or architectures used in PdM4.0? What are the impacts, the characteristics, the performances, and the possible limitations of its approaches? What are the major challenges related to the application of their method at large scales? The main contribution is to investigate which current models, methods or techniques of AI are mostly used in the context of PdM for I4.0. We consider the following action scheme: A detailed bibliometric analysis applied to scientific papers collected on the WoS database that deals with fault detection and predictive maintenance for I4.0 was first performed. Associated analyses and visualizations were carried out using the Bibliometrix R tool [32], VOSviewer [33], and Power BI software. We are then highlighted the main trends, challenges in industrial maintenance and the relevant methods that support conditional monitoring, fault detection, prognostic, and diagnosis in real-time prediction. We also showed the trends of publication of indexed documents over time. Next, we studied the current state of research on these research works considered and their roles in developing I4.0. In that regard, we identified some insightful indicators such as the most productive authors, the leading universities (with the most cited articles) and extracted and analyzed the most frequent keywords, including the different emerging themes or technologies related to PdM4.0. We also identified the socio-economic impacts caused by the intensive use of AI-based systems applied to PdM in the industry, the issues identified, key challenges, and future research direction related to I4.0 for PdM4.0. Finally, answering the detailed above questions, we can provide a helpful guideline for researchers to better understand the research topic, the current state-of-the-art, challenges, and future directions of AI models applied to the PdM4.0.

RQ1: What are the main means of scientific publications and their frequency in the context of the study?
RQ2: What are the most productive, impact and source growth dynamics?
RQ3: What are the most important or popular authors, journals, universities, and countries?
RQ4: What are the most common technologies or tools used in industrial maintenance, their performances, and limits?
RQ5: What are the research trends in industry 4.0 and industrial maintenance 4.0?
RQ6: What are the potential ethical impact rules using AI techniques for predictive maintenance in I4.0?
RQ7: What are the key challenges, issues identified and future research directions in AI techniques applied to PdM4.0?

3 Overview of the industrial revolution and maintenance strategies

3.1 Revolution of industry

The industry has experienced four main revolutions [34–36]. The 1st Industrial Revolution (Industry 1.0) took place between 1780, and 1860 with the creation of mechanics, the exploitation of coal, and the development of the steam engine. The 2nd revolution (Industry 2.0) for the first time brought the mass production at lower cost with the introduction of electricity, and the development of transport. Industry 3.0 occurred between 1970, and 2010. It highlights new information technologies, electronics, and telecommunications. Finally, the fourth revolution (Industry 4.0) was presented for the first time at the Hanover Fair in 2013 [1]. According to [37] the definition of Industry 4.0 depends on the field of application, and research. Figure 1 shows the main industrial revolutions and their related inspection or control techniques.

3.2 Industrial maintenance strategies

According to the European standards [38], maintenance is a combination of actions and management techniques that can be applied to ensure the correct performance of the machine over time. Figure 2 represents the classification strategies of maintenance; each method is described in [39]. Corrective maintenance (CM) is the action performed when a machine has faults or breaks down. Thus, there is no work until the failure is repaired. However, preventive maintenance (PM) aims to reduce the probability of failure of components. It is performed at well-defined frequencies or periods. Recently, the predictive maintenance (PdM) and condition-based maintenance (CBM) strategies have attracted more attention from manufacturers [39]. Predictive maintenance is a technique to predict the future point of failure, or the lifetime of a machine component before it fails [40]. We can exploit the masses of data to train AI algorithms to optimize the production system. According to [19, 40, 41], its algorithms can detect patterns correlated with faults, failures, or detect degradation at an early stage to implement adequate countermeasures.

3.3 Types of control in industrial maintenance

In industrial maintenance, there are four main types of inspection of mechanical production systems [42]. The first type which is visual inspection consists of carrying out a physical, or periodic checkup of the system (Industry 1.0). The second type is instrument inspections, which is a combination of visual inspections, and the frequent use of instruments to monitor the system's condition (Industry 2.0). Real-time condition monitoring consisting of continuous monitoring by allowing experts to give their opinions on the system status or health (Industry 3.0) is the third type of inspection. Finally, the last type is predictive maintenance

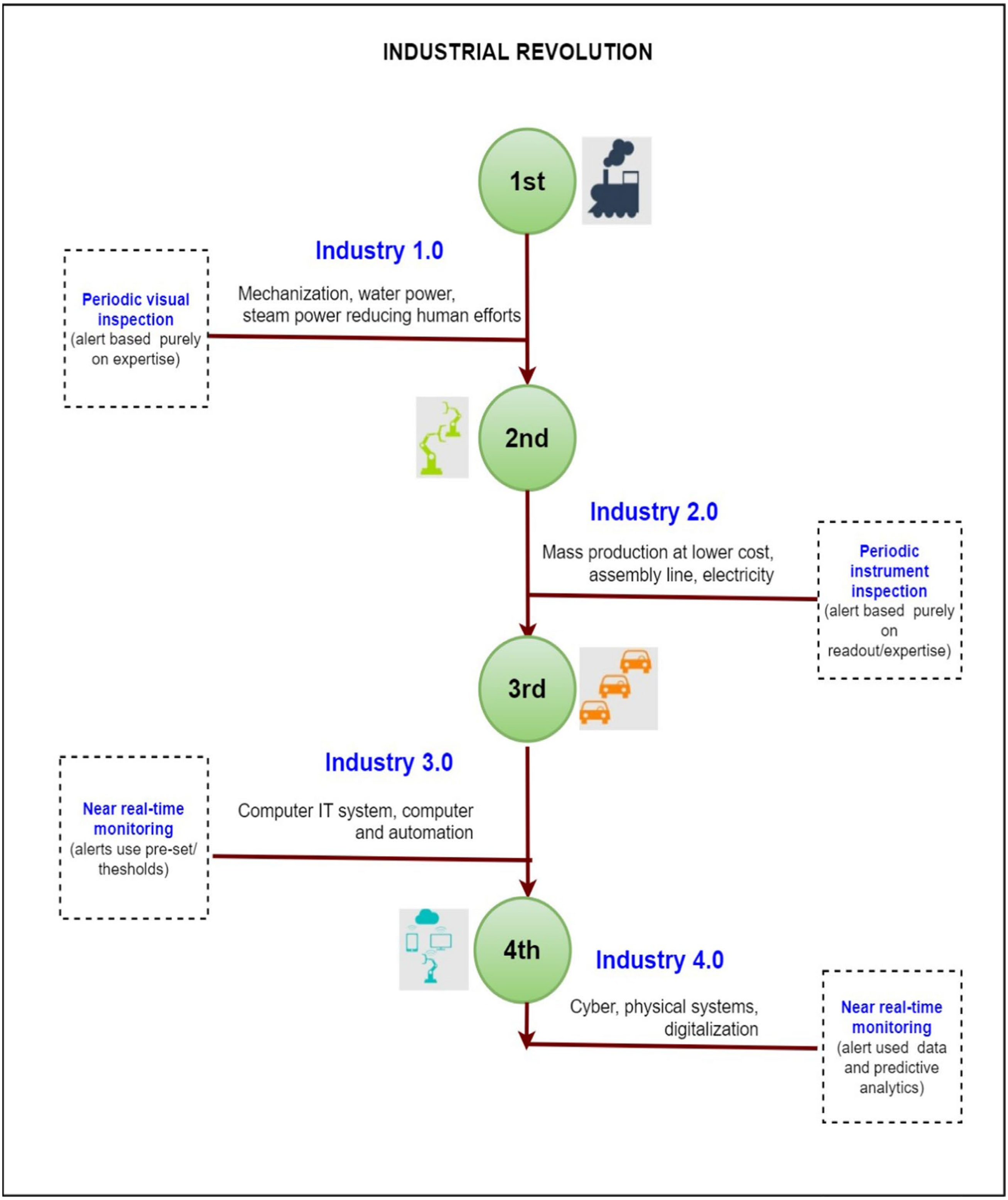

Fig. 1 Historical perspective of Industrial Revolutions, and relatives' inspection or monitoring techniques

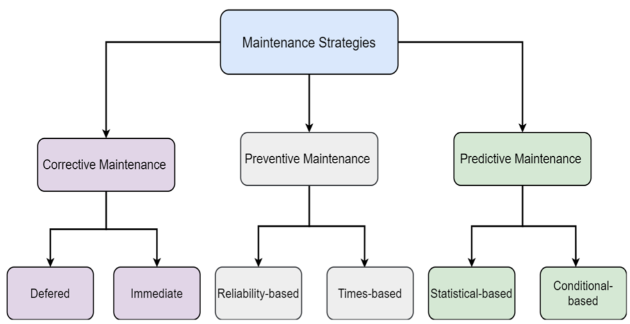

Fig. 2 Classification of monitoring or maintenance strategies in the context of Industry 4.0 (Adapted to [39])

which allows experts, and data scientists to exploit the data collected to predict the state of life of the machines.

3.4 Potential ethical impact of the use of AI for PdM in I4.0

It is widely acknowledged that AI is invading our lives. AI is creeping everywhere, from intelligent personal assistants to robotics (among the most common usages). Within the frame of PdM4.0, it can assist in cognitive tasks by providing a wide range of solutions to prevent downtime and equipment failure and even enable a system to reconfigure itself. In fact, a vital difference of the 4th Industrial Revolution from its predecessor is that we are now dealing with autonomous systems, not only automation [43]. Although fascinating, autonomous systems are worrying: to what extent is the AI algorithm's development, outcome, and impact correct and fair? To what extent can it identify the contexts in which it is right and fair and those in which it is not? What are the consequences of a loss of AI control? What are the influences on human welfare and integrity? These questions have led to the ethical issues of AI that have increased in the literature in recent decades.

As a study of what is morally wrong or right, by its very essence, following its etymology ("study of behaviors"), ethical questions define the practical principles of action. Different approaches have been developed to address the related issues, depending on the direction we give to our actions: either we act according to some moral values (virtue ethics), or according to the beneficial consequences they generate (consequentialist ethics), or according to their conformity to a principle regarding some obligations, duties or rights (deontological ethics). In some cases, our actions may be subject to conflicting ethical choices, leading to ethical dilemmas [44]. The Moral Machine from the Massachusetts Institute of Technology (MIT) illustrates such a context in which an autonomous vehicle may have to choose among ethical dilemmas: saving more lives, protecting passengers, upholding the law, avoiding intervention, gender or age or species preference, social value preference as mentioned in [45]. The authors claim that exploring ethical dilemmas should be the first step to building ethical systems. Besides, while making an automated decision, [46] noted that a virtual agent could make a judgment on its ethics (individual ethical decision) or take into consideration those of other agents (within the same decision process) which may have their ethics (collective ethical decision). More recently, [47], who studied trust in AI within the field of production management, identified possible antecedent variables related to trust and which were evaluated in human-AI inter-action scenarios. Their study proposed design guidelines for socially sustainable human–machine cooperation in future production management. The proposed framework is based on the SOR (Stimulus-Organism-Response) model, using decision situation characteristics as stimulus variables (predictability, error costs), AI characteristics (perceived ability, perceived comprehensibility) and human characteristics (digital affinity, expert status) as organism variables. They constructed a structural equation model in which implementation showed that AI characteristics and decision situation ones have a significant positive effect on the response, i.e., trust; for the human characteristics, they found that only one variable was statistically significant (i.e., digital affinity). Above all, following these studies and others on ethics in AI, we believe that addressing ethical AI is a moral obligation and a duty of AI developers for PdM4.0. Therefore, we consider that ethics applied to PdM4.0 would make it possible to be proactive, support innovation positively, and not stifle its potential. Otherwise, the design of AI algorithms used in PdM4.0 may remain opaque. It can generate biases, discrimination and worldviews without us always opening the 'black box' that makes them ethical and trustworthy.

4 Most common AI techniques used in PdM4.0

4.1 Main modeling techniques

The main approach for anomaly detection, prognostic and diagnostic in PdM is represented on Table 1. Knowledge-based modeling is an approach that is focused on knowledge and reasoning to solve problems [48]. Furthermore, this technique is based both on the conditional 'If–Then' rule, and in the knowledge known as 'Past' or 'previous' carried out in the process, also it is particularly useful to reduce the complexity of a physical model. In practice, it is often combined with other approaches as a hybrid method [49, 50]. Knowledge-based modeling can be classified into three sub-groups: rule- based [51], case-based [52], and fuzzy knowledge-based approach [53]. However, this approach is ineffective in the sense that it is impossible to apply the

Table 1 Modeling approaches for fault detection, and diagnosis in predictive maintenance

Modeling Approaches	Some sub-model
Physics-based	Kalman Filters [68]
	Markov models [69]
	Monitor-based [70]
	Fault trees [71]
Knowledge-based modeling	Bayesian Decision [68]
	Expert Systems [72]
	Binary Trees [73]
Data-driven modeling	Genetic algorithms [74]
	RF [75]
	Data mining [76]
	CNN [77]
Hybrid modeling	SAE and SVM [78]
	SVM and Naive Bayes [79]
	RF and LSTM [80]

rules without having experience, or precise knowledge of the process being studied.

Physics-based modeling requires the construction of a dynamic model by integrating various constraints, defects, or degradation linked to the non-stationary process [54, 55]. This approach has some advantages especially, the model parameters are directly related to the physical quantities, as a degradation, or deformation of the phenomenon can be explained by the variations of its parameters. The results can be easily interpreted. Although physics-based approach helps to better understand the physical universe compared to data-driven models; it is limited in its ability to extract knowledge directly from data that are mostly based on available physics. Sometime, the models generated are often too complex leading to incorrect results [56, 57].

Data-driven, or Data science modeling approach exploits both sensor data, to extract knowledge, or patterns useful for characterizing the condition of the system studied. It is based on statistical techniques, stochastic models [58], neural networks models [59, 60], data mining, and machine learning [61]. In addition, this approach is the most widely used in PdM and is a compromise between the application, and the accuracy of the model [62]. However, this method becomes unusable, and loses all interest, or use when the model is no longer capable of capture new changes associated with the process. Moreover, it does not characterize the law, or physics of the industrial process.

Hybrid, and digital twin modeling are a combination of a physical model, or a data-driven model [63, 64], or a knowledge-based model. Also, this approach continuously adapts to operational changes based on collected data, and online information [65–67]. Furthermore, Hybrid models provide better results, especially in terms of interpretability, and understanding of physics knowledge. However, they can be costly in terms of computing time, and in some contexts, the modeling of physics can be challenging or impossible.

4.2 AI models applied for PdM in I4.0

In the industrial context, AI is aimed at supporting decision-making. There are three main levels of support: descriptive, predictive and prescriptive AI. At the first level, AI consists of providing a reliable synthesis of the massive information that is available in the form of dashboards or Key Performance Indicators (KPIs). The second level is based on a set of rules and probabilistic or statistical approaches to provide forward-looking projections to better predict possible risks regarding the state of the system's degradation. In addition, to providing predictive insights, prescriptive AI proposes recommendations or feedback for facilitating and optimizing maintenance operations. The models used to perform these operations can be divided into two families: Machine Learning (ML) and Deep Learning (DL). Note that depending on the nature of the explanatory data and the target variables, approaches can be classified as supervised, semi-supervised, unsupervised, and reinforcement-based learning.

4.2.1 Machine learning techniques

Decision Tree (DT) is an approach to represent information in the form of trees structure with recursive partitions on the data space. DT is based on the principle of "divide and conquer", which means, the tree is built from a data set, and then it has decomposed on different subsets or branches until it reaches the last node or decision leaf (which can represent the limit of the division). Its subsets are obtained through divisions according to the Gini index. Moreover, DT is mainly composed of a main node named "root" (best predictor) among all subsets (less important). The algorithm can be exploited to solve classification or regression tasks and can be used in several industry applications [81, 82]. Decision rules or results produced by the algorithm are simple and easy to understand. However, the algorithm can generate very complex trees resulting in the overfitting problem. Furthermore, DTs suffers from instability and poor performance, compared to other ML algorithms that will be presented later.

Random Forest (RF) model has been developed by the author [83]. RF are based on combinations and aggregation (voting) of a set of random trees so that each node is evaluated independently. Furthermore, RF model is an improvement of decision trees, particularly in the correction of the instability and the variance reduction. In addition, RF offer the possibility of extracting the significant variables involved in model construction. The parameters of the model are easy to be calibrated, robust to noise and can be parallelized. RF are used in several applications especially for classification or regression problems [75, 84, 85]. They are often used as a benchmark in ML competition. However, learning can be difficult (latency of the algorithm) as far as a large amount of data and a significant occurrence of missing data are concerned.

Support Vector Machine (SVM) model is developed by the author [86]. SVM deals with a generalized linear model using the hypothesis space of a linear function in a high-dimensional feature space by creating an optimal partition hyper plane (maximum distance between the bridge margins and the nearest data). Optimization problems in this constrained setting provides convex solutions. Moreover, SVM has become more popular for their applications in image classification, face and handwriting analysis. Particularly, the authors [75, 87] apply SVM for conditional monitoring of mechanical or electronic machine. In addition, SVM uses kernel function to guarantee better discrimination, and the regularization of the hyper-parameters of the model helps to avoid overfitting problems. Some versions of hybrid SVM algorithms have been presented in [88], generally, they give higher performance than classical SVM model. However, kernel models can be sensitive to noise data or noisy classes, to overfitting problems when selecting the optimal model. Also, the estimation of the optimal parameters can be greatly challenging since an explicit model of nonlinear kernels does not exist. Finally, the computing time or the GPU memory is important when the data to process are increasing.

K-Nearest Neighbor (K-NN) model is a non-parametric classification algorithm. Its objective is based on the classification of new samples classes with higher similarity, in this case, the K-instances nearest to the reference set are computed on a Euclidean distance metric [89]. K-NNs are very often used in industrial applications, for pattern recognition problems or recommendation systems. This approach does not require any hypothesis on the data; Furthermore, they are simple, efficient, and easy to perform. The authors [81] exploit an improved version named WKNN for fault detection and isolation tasks of complex systems. Besides, the distance-weighted k-nearest neighbors (WKNNs) are more efficient than K-NN when the classes are separated. Nevertheless, K-NN can be inefficient because of the choice of the method of computing the distance and the number of K-nearest neighbors. Moreover, K-NNs can be inefficient due to their choice of the distance computation method and the number of K-nearest neighbors. When we use a large amount of data, the algorithm becomes much slower, this is a real obstacle to apply K-NN in real-time predictive systems. In addition to the models discussed previously, there are several classes of ML models, in particular Naïve Bayes, Discriminant Regression (LDA, QDA), penalized models (Ridge, Lasso, Elastic net) or ensemble models (Bagging, eXtreme Gradient Boosting “XGBoost”). Despite their various benefits and applications, these approaches can become unstable and inefficient (high-dimensional learning and overfitting problems) in the following cases: high data volume, complex equipment data, unbalanced classes, missing and noisy data. Today, scientific, and technological advances have allowed deep neural networks learning to emerge as a real improvement over traditional machine learning algorithms mentioned above.

4.2.2 Deep learning techniques

Convolutional neural networks (CNN) [90] are acyclic deep learning networks, composed principally by two types of artificial neural cells: processing (convolutional) and pooling. Concerning information or feature extraction on all input samples, CNNs are based on more convolution kernels named feature extractors. To reduce the number of parameters, these kernels and weights are distributed over the entire bidirectional input matrix. CNN has shown its efficiency in various applications such as pattern recognition or signal processing. Moreover, they required very few pre-processing, since they perform their own filters during training, which explains their robustness to noisy data. However, the design of this architecture remains a major challenge for researchers. Several variants of optimized algorithms, and architectures have been proposed in the literature. The AlexNet and its variant [91] are composed of five convolutional layers and three fully connected layers combined with regularization methods (data augmentation, dropout and Norm L_1, or L_2). The Network AlexNet has won many competitions, however, it has limitations related to the image’s fixed resolution, thus, the SPP network has been developed to overcome this problem. The Visual Geometry Group (VGG) network increasing the depth of the network by convolutional layers with very small convolutional filters. There are other architectures such as GoogLeNet, RCNN (Regions with CNN features) and FCN (Fully Convolutional Networks). Despite their many advantages, their network (black box models) is complex, and the decision-making rules are not explainable. Besides, the increasing number of hidden layers can have an impact on the performances of the networks.

Auto-encoders (AE) are non-recurrent neural networks with hidden layers smaller than the input layers. AE are formed by an encoder and a decoder. Its objective consists of representing in an optimal way the input data. Thus, the algorithm tries to learn a new representation (encoding) from the given input data set and to reduce its dimension. To predict an output target value, the algorithm performs an optimization operation by minimizing the reconstruction error of its own inputs. Also, there are different architectures of AE. The sparse AE that seeks to extract sparse feature on the raw data by penalizing both hidden unit bias and hidden layer activation output. A variant named “low density autoencoder” helps to detect objects without a priori knowledge of the class labels, the resulting model is robust to translation and rotation operations. The denoising and contractive AE have a similar network and the ability to capture details about the data. Their network structures are based on the same principle as the one shown in the previous model. Besides, denoising AE tends to introduce noise in the training set and then selects the correct information on the input of a biased model. While the “contractive” autoencoder adds explicit regularization (matrix norms such as the Frobenius norm) to its reconstruction error function, the denoising network forces the model to learn a function that is robust to slight variations in input values. AE are efficient and have many applications, such as anomaly detection, data denoising, transfer learning, or random fake data generations. In particular [92] use AE for the real-time remote sensing of the degradation states of the machines. Moreover, a hybrid deep SAE-SVM model is used by [78] for intelligent fault diagnosis in industry. However, the computing time of AE can be important because the problem does not prevent their exploitation for on-line learning.

Generative Adversarial Networks (GAN) are unsupervised learning algorithms that can generate “fake data” very similar to the original ones. GAN algorithm is based on the game theory, where two networks generator (G) and discriminator (D) are in competition. The first network is the generator of a fictitious image sample, and the second one takes the role of an adversary, checking if the data are real or from the generator. If the last one is not satisfied with the results, it returns it to the generator so that it can generate a new sample image. In addition, GANs have been the purpose of several extensions as Wasserstein GAN (WGAN) which uses the optimal transport plan to generate the data from noise, the discriminator calculates the Wasserstein distance between the distribution of the generated and real data [93]. WGAN is allowed to improve the stability of the optimization process like the search of the model hyper-parameters. Other metrics have been applied to generate or discriminate the data while improving the corresponding optimization problems, we can mention Lipschitz-GAN (LGAN), WGAN with gradient penalty (WGAN-GP), Spectral Normalization for GAN (SNGAN), First Order GAN (FOGAN), Vanilla and Least-Squares GAN. These approaches contribute to the reduce of the computing time, and they are used in many applications such as pattern recognition, generating or simulating data (texts, pictures, sounds, or videos). However, GANs are limited by the instability of unsupervised learning algorithms, and the generation of speech data is very complicated. Thus, it is not easy to turn the model training process without losing accuracy.

Finally, there are also other architectures that we have not introduced in this paper, the Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) or the Restricted Boltzmann Machine (RBM), which are applied to sequence processing problems such as time series. However, all the approaches mentioned above are also known as black box models, and their decision-making rules are not systematically explained.

4.3 Characteristics and techniques classifications

The application of AI techniques in industries can be influenced by various characteristics, including. We have the hardware and software infrastructure which provides security, interconnectivity of systems, and information processing abilities (Edge computing and Cloud). Digital twins and decision-making help in testing the different scenarios virtually and to make decisions. In this case, decision-making is based on the level of trustworthiness, and effectiveness of the model developed. Here, the evaluation and interpretation of uncertainties or errors rates do not have the same significance and thus depend on the targeted objectives. In addition, AI-based model is highly conditioned by the characteristics of the data (reliability, volume, variety, velocity, veracity, and availability). Furthermore, we could classify these models according to some aspects: (a) Nature of the task such as supervised (regression, classification), unsupervised (clustering, association), reinforcement or semi-supervised learning. (b) Type of variables to be analyzed (nominal, ordinal, discrete or continuous); (c) data structure (texts, pictures, signals, videos, images, sounds); (d) data quantity and quality (presence of missing, incomplete, mislabeled, noisy or biased data).

5 Methodology for the study

5.1 Bibliometric analysis

The Bibliometric is considered as the oldest bibliographic research method in information science. According to [94], it can be defined as a method for evaluating, and visualizing scientific research papers. According to [95] bibliometric analysis is a field of research that involves analyzing trends in scientific research papers on a specific topic, subject, or area. Also, bibliometric is seen as a statistical analysis applied to a set of documents, or books. Note that some organizations use this type of analysis as a distribution criterion to allocate financial aid to researchers [94]. The objective being to provide motivation, guidance for research, or to highlight the trend, the impact of the units. Finally, it provides motivation, and guidance for research. In bibliometrics, the units of analysis frequently used are journals, documents, references, keywords, authors, and their affiliations, universities, or countries and their collaborations. Keywords can be selected in relation to titles, abstracts, document, or bodies. These keywords can be provided either by the original authors (author keywords) or indexed against referenced bibliographic data sources also known as Keywords Plus. Words represent the terms, or phrases most frequently used in the titles of the references of a scientific document [96]. Besides, they are generated by algorithms that can deeply capture the content of a document. Moreover, the authors [97] have made a comparative study between the keywords author, and the keyword Plus. Unlike the keyword author's, the keyword Plus is more complex, and does not necessarily appear in the title of the article.

Therefore, for a bibliometric study, we can analyze several types of relations between the units, we have similarity relations, co-occurrence relations, and direct links between the units. These relationships can be represented as graphs, or networks. Authors [98] present a taxonomy of the most used bibliometric techniques. Bibliometrics analysis can be applied in many fields such as logistics [99], economics, biology [100], and in industry 4.0 [23, 101]. In this article, we look for the articles using the WoS search engine according to certain criteria to avoid possible sources of error [102]. Furthermore, we evaluate the collected publications with some statistical metrics such as productivity, number of citations, frequency of citations, publications, impacts measure, and hybrid measures.

5.2 Recommended workflow for science mapping

In this subsection, we propose the four-step workflow guideline for scientific mapping research using bibliometric analysis [103]. The first step consists of defining the research questions and selecting the appropriate bibliometric methods to answer them. The second step is focused on data collection, the researchers must identify the databases in relation to the thematic study. In addition, they must perform filtering, exclusion, and selection operations to extract relevant publications. They must also consider the period to capture the evolution of the case study over time. The third step is a bibliometric analysis, they can be carried out using several statistical software [98]. The last step is data visualization, and interpretation according to the results; there are several tools available to achieve this goal [32, 98].

5.3 Web of science and data collection

To carry out a bibliographic study we can use several bibliographic databases [104, 105] such as the web of science (WoS), Scopus, Springer, Google Scholar or Science Direct. For our case study, we focus on the WoS search engine, our motivations are the following: (a) WoS is a bibliometric analysis tool that allows evaluating statistical indicators of publications; (b) unlike Scopus, WoS contains more multidisciplinary publications with a high impact in each field [106]; (c) in contrast to Scopus, WoS contains more multidisciplinary publications with high impact in each field, also, we exclude Scopus to avoid duplicate documents, and Google Scholar for the reduced performance compared to the quality of the search obtained. In fine, we also exclude IEEE, Science Direct, and Springer because they only index their own publications [104].

5.4 Scanning and keywords search

To identify important publication keywords in bibliometrics, there are several approaches [107, 108]. We applied a variant of the TF-inverse document frequency (TF-IDF) method described by [109], that helps in the identification of an important term by combining their popularity and their discrimination. This approach has several advantages, for example, TF-IDF weights are more relevant for keyword frequency than TF-KAI weights [110]. According to this index. We found that keywords such as AI, real-time and PdM are the most important and correlated (significant increase) to productivity on I4.0. To define the relevant publication sample, we used these keywords to perform several queries on the WoS engine. The search also considers the years of publication, the title, the abstract, and the author/indexed keywords of the articles. We performed the search on 10th March 2021 in the WoS database.

> **(**Fourth Industrial Revolution **OR** Industry 4.0 **OR** Mechanic* **OR** Real-Time**) AND (**Artificial Intelligence **OR** Machine Learning **OR** Deep Learning **OR** Artificial Neural Network**) AND (**Predictive maintenance **OR** Decision making **OR** Diagnostic **OR** Prognostic **OR** Monitoring**) AND (**Time span: 2000-2021**)**

The research produces the bibliographical data for indexed documents (4065) including some information about papers such as titles, type of article, author publications, affiliations, countries, keywords, abstracts, number of citations, source conference, publisher name, address, years of publication, volume, issue number, and a list of cited references.

6 Analysis, and results

In this section, we focus on the main bibliometric analysis metrics [32]. Its metrics can be obtained on several levels such as sources, articles, authors, references, keywords, universities, or countries. We can, therefore, classify these elements by their impacts, productivity, their frequency of citations and network collaboration. We can also visualize co-occurrence networks, the thematic, and the trend of keywords. These analyses provide new information, and thus help to improve knowledge about scientific research.

6.1 Main information about the collection

Table 2 shows the main information about 4065 collected publications obtained on the WoS search engine according to the criteria. We have a total of 11268 keywords, and more than 450.00 authors (Author 14108, Author Appearance 18681, Author of single-authored documents 140, Author of multi-authored documents 13968, and single-authored documents 145). Also, we have 2308 source conferences, and more than 124771 references (Fig. 3). The pie chart (Fig. 4) shows the distribution of retrieved documents over the last 20 years. Firstly, we can see that articles are more representative with 2464 (60.62%) documents, secondly, and thirdly we have respectively the proceeding papers 137 (33.83%), and the reviews 204 (5%).

Table 2 Main information and statistics regarding the collection published between 2000, and 2021 on WoS

Description	Results and statistics
Article (2463)	Articles (2291)
	Book chapter (10)
	Proceedings paper (70)
	Article data paper (4)
Review (400)	Classic review (198)
	Early access (6)
Proceeding paper	1375
Editorial material	16
Meeting abstract	5
Editorial material	16
Period	Years (2000–2021)
	Keyword Plus (5170)
	Author's keyword (11,268)
	Author (14,108)
	Author appearances (18,681)
Author publication	Single-authored doc. (140)
	Multi-authored doc. (13,968)
Author collaboration	Single-authored doc. (145)
Source conference	2308
References	124,771
Average year of publication	3.76
Average citations per document	8.363
Average citations per year per document	1.752
Collaboration index	3.56
Co-Authors per documents	4.6
Documents per author	0.288
Author per documents	3.47

6.2 Annual scientific publication trend

In this subsection, we answer the question **RQ1**. We defined scientific productivity as a metric that measures the frequency of publication, or author impact on a specific discipline. Figure 5 shows that over the last 20 years, there is an exponential increase in the number of publications. Furthermore, Table 3 shows the evolution reaches its peak in 2020 with more than 1095 papers published (25 or 96%) compared with the previous year. At the end of the first quarter of the year 2021, we record about 175 indexed documents.

6.3 Most productive, impact and source growth dynamics

To answer the question **RQ2**, we analyze the review, their impacts, growth, productivity, number of citations, and network collaboration. We have 2295 conferences, Table 4 shows the most productive journals according to the number of publications (NP), number of citations (TC) and impact (*h*-index, or Hirsch index). We can note that, H-index gives the number of publications by the author has received at least *h* citations. When we focus on the NP metric, we can see that the most relevant, and productive source is the IEEE Access conference with a score of 143 (6%) papers. This journal publishes scientific papers related to electrical engineering, electronics, and computer technology. In the 2nd, and 3rd rank, we have respectively the Sensors applied Sciences-Basel (119), and Remote Sensing journals (47). However, we can note that the most productive conference is not necessarily the most cited, and vice-versa. For example, Sensors-Basel is most cited than IEEE Access, even though, it is less productive than Measurement journal. Regarding the source network collaboration, we consider only conferences with more than 5 publications. Finally, the Fig. 6 represents the network visualization for the most productive journal, this network showed 113 conferences distributed in 16 clusters.

6.4 Most productive authors, universities and countries

To answer the question **RQ3**, we exploit several axes of research, and we perform analyses to describe some elements, such as authors, references, universities, countries and continents.

6.4.1 Most productive and highly cited authors

Table 5 shows the most cited authors based on the TC index, Bellini, Filippetti, and Tassoni (694 total citations) are the most cited authors with the same score although they have published only one article. They have received remarkable

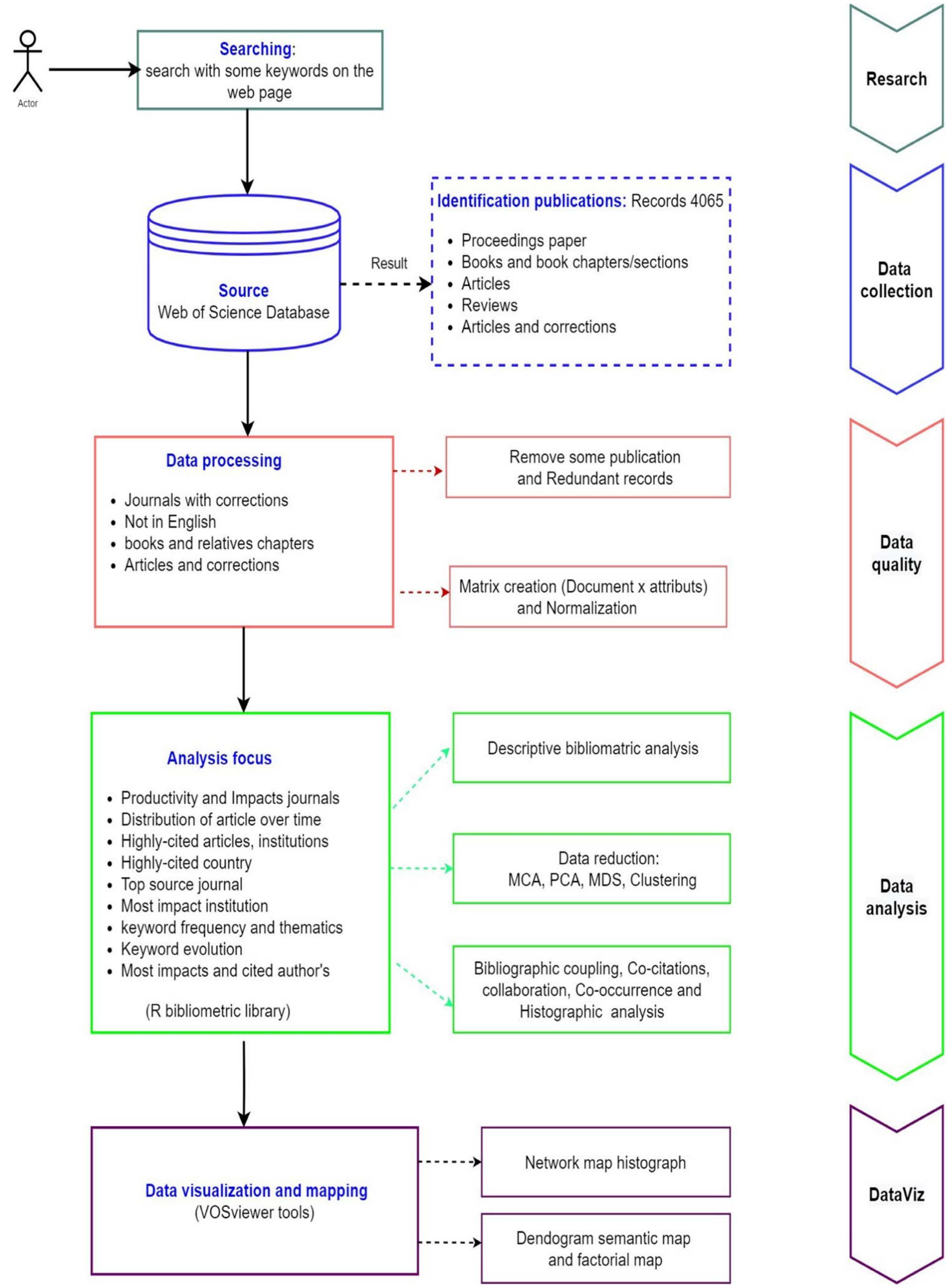

Fig. 3 Methodology framework for bibliometric analysis. Each color corresponds to a step of methodology. The different steps represent the methods or strategies used to perform this study

attention from the community for their publication. However, if we focus on the TC index (Table 5), we can note the most cited authors are not necessarily the most productive. Furthermore, Fig. 7 represents the network collaboration between the authors. In this network, the distance between two authors indicates the relationship between them in terms of co- citation links. Also additionally, the link is stronger when the distance is high, or the relationship is strong. The spheres dimension is proportional to the frequency of collaboration, and the connections indicate the presence of collaboration. We have 24 clusters, 1st cluster (Gupta, Naizi and Varma), 2nd cluster (Massaro, and Galiano), and 3rd cluster (He, Tiwari, and Wang). Ultimately, the most co-cited authors are respectively Lecun (343), Breiman (250), He (218), Hinton (171), Lee (168), and Hochreiber (165).

6.4.2 Most productive and cited affiliations

Table 6 shows the list of the most productive institutions. In the 1st rank, we have the University of Illinois with 81 publications, in the 2nd, and 3rd, we have respectively the University of Shanghai Jiao Tong (74 publications), and

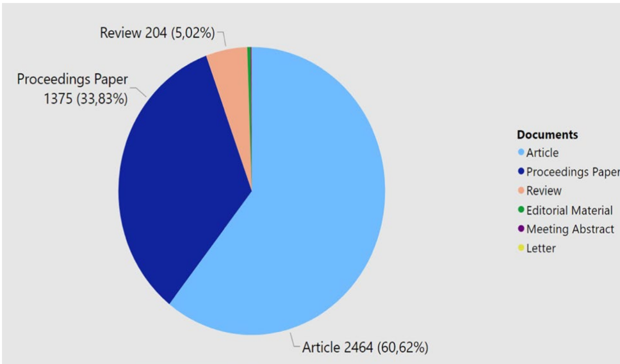

Fig. 4 Pie Chart: Types of retrieved documents over the last 20 years

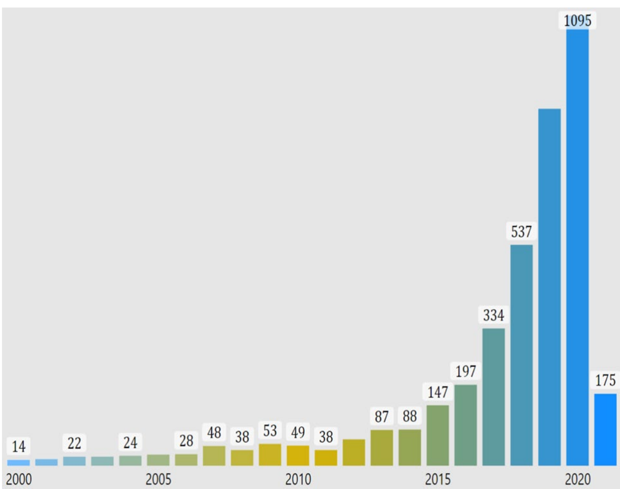

Fig. 5 Annual scientific production published in WoS journals over the last 20 years

Table 3 Productivity: Annual number of published articles between 2000–2021 on WoS

Year	ND	ND (%)	Year	ND	ND (%)
2000	14	0,33	2011	38	0,90
2001	16	0,38	2012	64	1,52
2002	22	0,52	2013	88	2,08
2003	22	0,52	2014	147	3,48
2004	24	0,57	2015	197	4,67
2005	27	0,64	2016	334	7,91
2006	28	0,66	2017	334	7,2
2007	48	1,14	2018	537	12,7
2008	38	0,90	2019	868	20,58
2009	53	1,26	2020	1095	25,95
2010	49	1,16	2021	175	4,14

D is documents, and *ND* (%) is a number of documents in percent

Table 4 Most productive journal sorted by the publication number (NP), most journal impact (h-index), most cited journals (TC)

Source journal	NP	*h*-index	TC
IEEE Access	143	14	814
Sensors	119	14	119
Applied Sciences-Basel	47	7	148
Remote Sensing	34	8	297
IEEE Sensors journal	32	8	172
Computer & Electr. in Agriculture	26	8	291
Advanced manufacturing Techno	26	6	124
Energies	21	5	142
Scientific Reports	21	6	114
Electronics	18	4	20
Measurement	18	17	303
Neural Computing and applications	17	7	70
Plos One	17	5	166
Computer in Industry	15	7	370
Expert Systems with application	15	9	227
Intelligent manufacturing	14	7	242
Computer in Industry	15	7	370
Expert Systems with Application	15	9	227
IEEE Trans.on Instr. & Measurement	113	15	41
Multimedia tools and Application	15	4	41
IEEE Internet 0F Things Journal	14	4	132
Journal of Intelligent Manufacturing	14	7	242

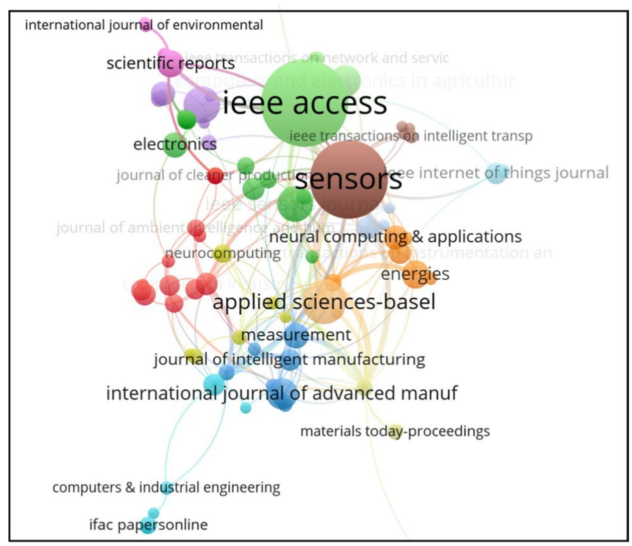

Fig. 6 Network visualization for most productive journal

California Los Angeles (71 publications). Table 7 shows the most cited organizations. Moreover, when we look at the collaborative network organization (Fig. 8), we notice that the University of Chinese academy sciences (green

Table 5 Most cited authors: authors are ordered by a total citation (TC) index

Authors	TC	NP	*h*-index	ND (%)
Bellini A	694	1	1	0.23
Filippetti F	694	1	1	0.23
Tassoni CA	694	1	1	0.23
Lin J	502	3	4	0.92
Jia F	487	4	4	0.92
Liu C	477	8	21	0.46
Xu X	467	8	15	3.44
Zheng Y	461	4	8	2.29
Lei Y	452	2	2	5.33
Dinx SX	419	1	1	0.23
Ozcana A	340	4	5	1.15
Zhang Y	390	10	18	4.20
Liu F	383	3	4	0.92
Wang C	381	4	13	3,74
Liu H	378	6	19	5.01
Li Z	363	7	18	4.20
Hsieh HP	355	1	1	0.23
Bao Z	349	1	1	0.23

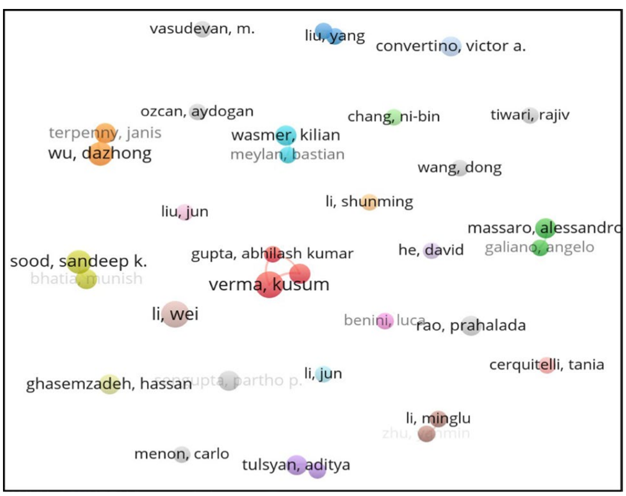

Fig. 7 Network visualization for Publication highly Co-authorship. Each cluster is represented by the color. To interpret the results, and the color of the legends in this figure, the reader can refer to the Web version of this article

cluster) is the most collaborative with 49 publications, and 684 citations, followed by Shanghai Jiao Tong University (24 publications, and 420 citations) and Georgian Institute Technology (23 publications, 410 articles). We can conclude that organizations from the USA, and China globally dominate the research in the field of study.

Table 6 Most relevant affiliations ordered by a number of articles

Affiliations	*N* Articles
University Illinois	81
Shanghai Jiao Tong University	74
University California Los Angeles	71
Nayang Technology University	66
Tsinghua University	65
Zhe Jiang University	65
Stanford University	61
Huazhong University Science and Technology	59
Xi and Jiao Tong University	51
Northwest University	47
University Michigan	46
Seoul National University	45
Yonsei University	43
King Saud University	43
University California Irvine	43
Emory University	41

Table 7 The most cited organizations ordered by TC

Organizations	Citations
University California Los Angeles	822
Xi Jiao Tong University	791
University Bologna	770
University Modena and Reggio Emilia	713
University of the Chinese Academy of Sciences	684
Stanford University	631
Georgia Institute Technology	489
London's Global University	475
University California San Diego	442
University Michigan	439
Massachusetts Institute of Technology	420
Tsinghua University	419
Northeastern University	394
University Pittsburgh	392
University Cincinnati	347
Qatar University	339
Hong Kong University	330
University of Southern Queensland	316
Los Alamos National Laboratory	309

6.4.3 Scientific productivity by country and continent

Regarding the country's scientific production, Table 8 and Fig. 9 show that the USA, and China are the most productive countries. We have already observed this trend when we study the most important institutes (subsection 6.4.2). We can, therefore, deduce that the underdeveloped countries are not representative, for example, Oceania and Africa have

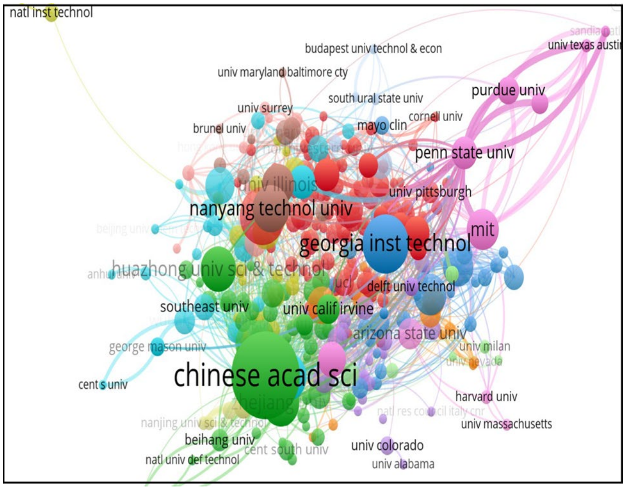

Fig. 8 Network visualization for international collaboration affiliation. Each group is represented by the color

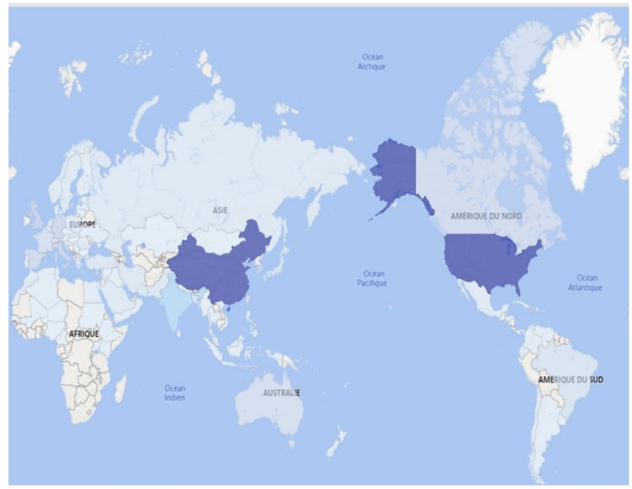

Fig. 9 World map of the country level scientific productivity, for paper collected on WoS over the last 20 years. Color scale is given by the number of articles, dark blue: high productivity, light blue: low productivity

Table 8 Most productive and cited countries ordered by the frequency publication, or productivity (years 2000–2021)

Region	Frequency	Average article citations
USA	3072	11.6
China	2977	8.1
India	1019	9.7
UK	639	8.7
South Korea	627	10.5
Italy	552	11.4
Germany	498	7.6
Spain	432	13.1
Canada	428	8.41
Australia	355	8.4
France	326	6.8
Japan	248	8.9
Brazil	243	6.6
Singapore	152	9.6
Malaysia	139	8.4
Switzerland	136	10.9

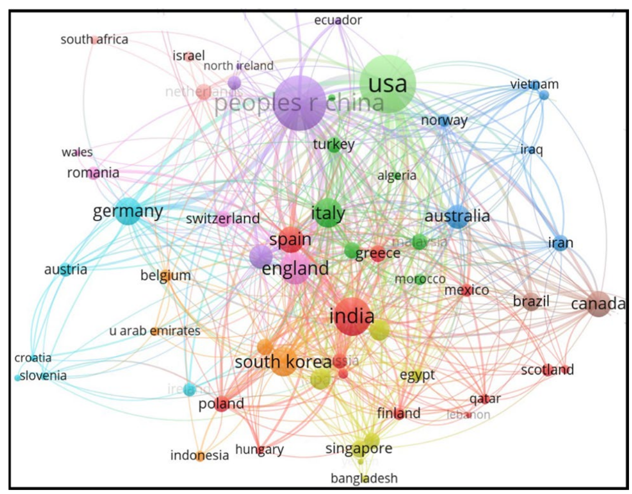

Fig. 10 Network visualization for international collaboration

respectively (399) 3%, and the African continent (212) 2% publications. This trend implies that these continents are lagging even though research activities are dispersed on a global scale. In addition, Fig. 10 illustrates the network collaboration between countries confirms that these countries are behind in research in the study. These low productivity trends of the universities, or institutions belonging to developing, Third-World countries can be partly explained by the low collaboration between authors from developing countries. Also, the lack of infrastructure, access to digital services such as the internet, electric energy, and the reputation of the institution in the scientific community are factors hindering this development.

6.4.4 Most global cited papers and references

Table 9 globally shows the most cited documents published in the WoS database over the last 20 years. In particular, paper [111] published in the IEEE Trans Ind Electron conference is the most cited (694 citations). Here, the authors are working on AI, and decision-making models are applied to the fault detection, diagnosis, and condition monitoring of electrical machines. Paper [112] has 419 citations, the

Table 9 Most global cited scientific publications

Paper	Frequency	TC/year	Year
Bellini [111]	694	49.57	2008
Lei [112]	419	69.83	2016
Zheng [113]	355	39.44	2018
Benight [114]	349	38.77	2013
Mueller [115]	253	18.20	2008
Abdeljaber [116]	249	49.80	2017
Bigio [117]	237	10.77	2000
Verrelst [118]	231	15.20	2012
Khan [119]	225	56.25	2018
Yaseen [120]	203	67.66	2010
Berg [121]	190	28.42	2015
Oresko [122]	197	16.41	2015
Jing [123]	196	39.20	2010
Gonzaga [124]	191	14.69	2009
He [125]	191	38.00	2017
Michie [126]	190	38.20	2017
Botu [127]	170	24.28	2015

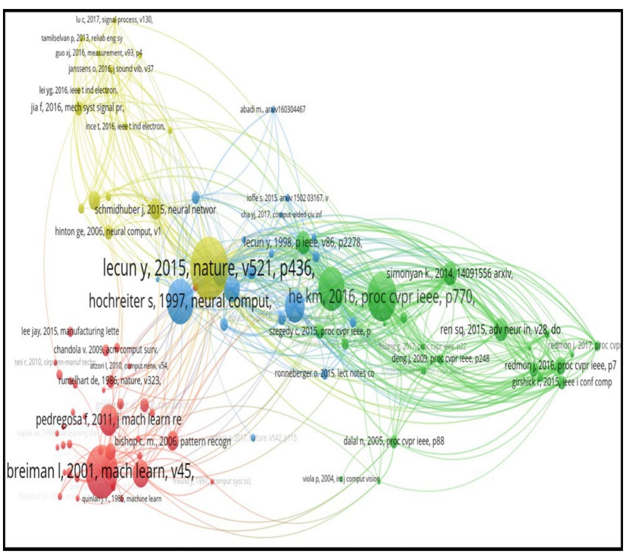

Fig. 11 Network visualization for Co-citations. Cluster 1: yellow color, cluster 2: red color, Cluster 3: green color, cluster 4: blue color (color figure online)

authors present the application of regularized sparse filtering model for intelligent fault diagnosis under large speed fluctuations. Furthermore, the scientific article [113] is cited 355 times, it published in the conference on knowledge data mining. In addition, authors Lecun [90], Breimman [83], and He [128] are the 3 most cited references (199, 169, and 156 frequency co-citations). Furthermore, Fig. 11, shows 6 clusters of co-citations network reference. Breimman, Lecun, and He are most representative for each cluster. In detail, the first cluster is formed by (Lecun, Hinton, Schmidhuber), the second, third and fourth cluster are respectively (Breiman, Bishop, Pedregosa), (Hochreiter, Kingma and Goodfellow), and (He, Ren, Redmon). In this regard, we see that LeCun (644), He (596) and Krizhevsky (591) have the highest link collaboration. Finally, we can conclude that most of these articles presented in Table 9, deal with topics related to the digitalization of industry, use of sensor data, IoT, big data, condition monitoring, anomalies detection, ML, and DL modeling applied in PdM4.0.

6.5 Most common technologies or models used in predictive maintenance

6.5.1 Most frequent keywords and co-occurrence analysis

In this section, we focus on the question **RQ4**. We first analyze the most frequent keywords, and their co-occurrence networks. The co-occurrence network keyword is a relational bibliometric metric frequency of scientific knowledge. So, the node represents a keyword, and their size is proportional to the frequency of co-occurrence of the word. While the color determines the cluster to which the element belongs. Thus, its clusters provide a global view of divergent research areas and group words according to the scientific field of research. Moreover, two keywords tend to be relatively close when they appear more frequently in the same articles. Furthermore, the distance between two nodes in the figure is determined by the density. To improve the analysis, we considered the most frequent keywords in each group, and the keywords that appear at least three times in the abstract. Table 10 presents the list of the most frequent author keywords in the publications. We can see that the most frequent author's words are machine learning with 792 occurrences followed by deep leaning, artificial intelligence, and monitoring with respectively 479, 286, 220, and 177 occurrences. Finally, the Fig. 12 shows the density of the author keywords co-occurrence, and Table 11 represents their word clustering. This illustrates the most important keywords, machine learning, fault, diagnostic, intelligent systems, data science, CNN, ANN, computer vision, network monitoring or on-line monitoring, have a great impact or importance for I4.0 and Pd4.0. In particular, DL and ML approaches have a major role in solving PdM problems in I4.0.

6.5.2 Keywords conceptual structure map

Second, we use Multiple Correspondence Analysis MCA to analyze the keyword conceptual structure Map. MCA approach is an exploratory multivariate technique for the analysis of multivariate categorical data [32, 129]. We

Table 10 Top 20 of the most author keywords

Author's keyword	*N* occurrences
Machine learning	792
Deep learning	479
Artificial intelligence	286
Monitoring	220
Artificial neutral networks	177
Machine	108
Internet of things	792
Classification	479
Fault diagnosis	88
Feature extraction	85
Sensors	83
Big data	82
CNN	78
Industry 4.0	75
IoT	71
Condition monitoring	66
Predictive maintenance	64
Anomaly detection	62
Real-time	55

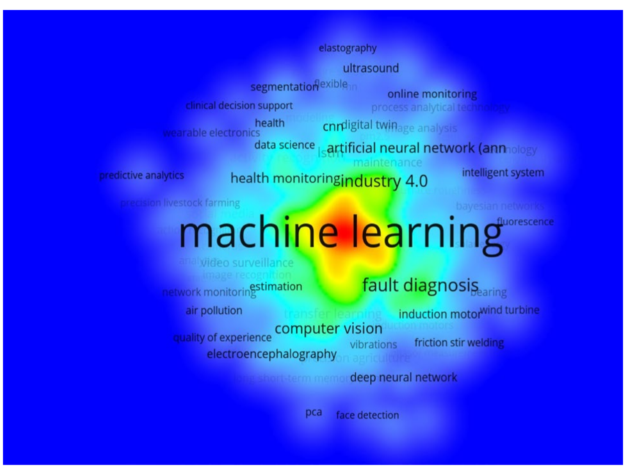

Fig. 12 Density visualization map of the most frequently related terms in retrieved articles on WoS. The frequent terms were visualized using VOSviewer

can explain the importance of the keywords in relation to their positions on the map, and on the main axes. Also, the proximity between two keywords implies that they have a similar distribution. Figure 13 shows the distribution of the most common keywords with the minimum number of documents (10) grouped into two groups. We can notice that keyword such as big data, sensors, diagnosis, systems, or simulation is located on the same plane and are very close. Furthermore, keywords such as IoT, and the internet belong to the second axis. These keywords are therefore, associated with the most common technologies applied to the PdM in I4.0. Lastly, the performance and limitations of these models have been described in section 4.2.

6.6 Research trends in industrial predictive maintenance

6.6.1 Keywords dynamics analysis and trend topics

Regarding the question *RQ5* we analyze several elements. Initially, we investigate the keyword evolution associated with the topic study (Fig. 14). From 2014, we note a real emergence of the use of approach such as AI model-based (DL, ML) applied to monitoring, diagnostic technique, and PdM4.0. When we focus on keyword plus, we have terms like classification, systems, data, real-time analysis, prediction, identification, and diagnostics that are important to develop the anomaly detection, condition monitoring and PdM4.0 systems. Furthermore, Fig. 15 shows the topic trend over the last 20 years. From 2018, we observe an increasing use of several technologies and models approaches, such as sensors, fault detection, condition, health monitoring, data, IoT, and data-based modeling, that support the rise of PdM4.0 and I4.0. Also, we have observed this evolution, and development in Fig. 14.

6.6.2 Thematic map and thematic evolution

An additional element that guides us to answer the question *RQ5* is to use a co-word for analyzing the evolution or

Table 11 Cluster of co-occurrence network author's keyword

N cluster	Node
Cluster 1	Deep learning, CNN, structural health, neutral networks, computer vision
	ANN, object detectors, LSTM, real-times monitoring, transfer learning
Cluster 2	Monitoring, fault diagnosis, sensors condition monitoring, real-time systems, signal processing, forecasting training
Cluster 3	Machine learning, IA, IoT, Big Data, data mining, pattern recognition classification, I4.0, anomalies detection, RF, predictive maintenance, Health monitoring edge computing,
Cluster 4	Support vector machine, remote sensing, neutral networks, image processing

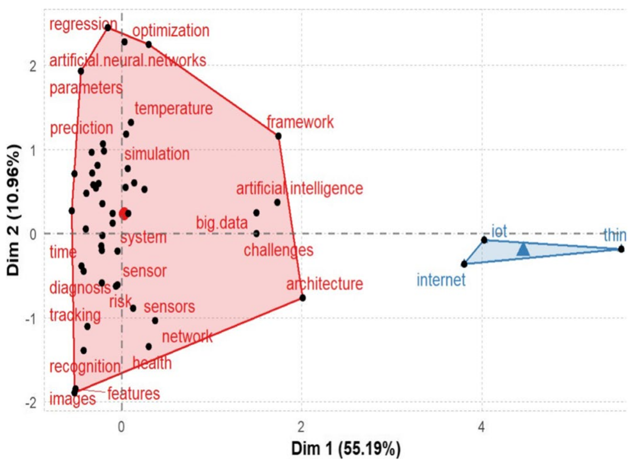

Fig. 13 Conceptual Map, and keywords clusters (minimum number of documents is 5, method: MCA, cluster 1: the red color, and cluster 2: the blue color (color figure online)

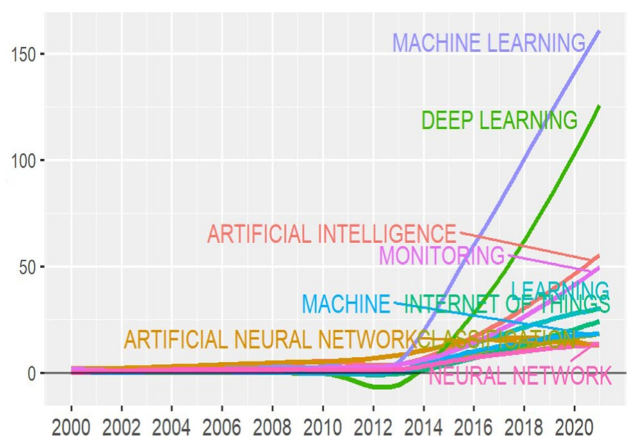

Fig. 14 Word dynamics for the keyword plus

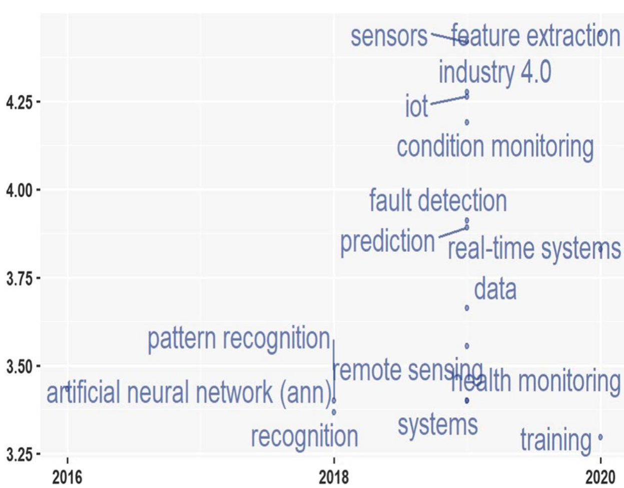

Fig. 15 Topic trend analysis over the last 20 years

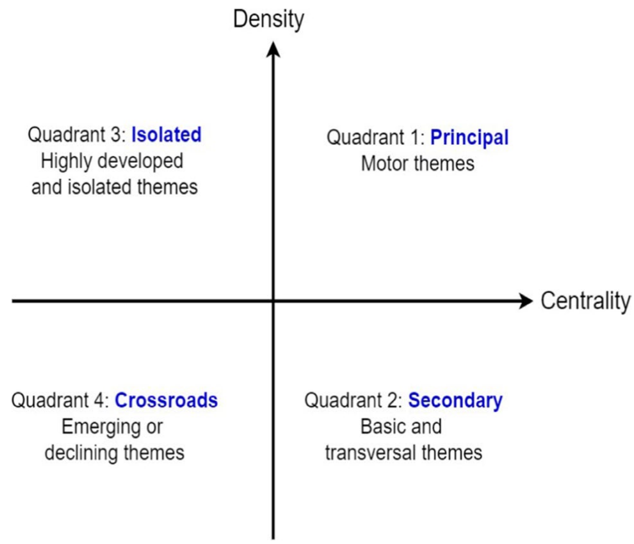

Fig. 16 The strategic diagram (adapted from [130])

trend of the most significant research thematic. Regarding the co-word analysis, each cluster represents the conceptual thematic or topic developed in the domain, and the research period. Thus, authors [130] defined a strategic diagram by Callon's centrality metric, which measures the degree of relationship, or links between each cluster. In addition, the strength, and the number of links imply a major relationship between the research problems in the scientific community. Indeed, Callon's density measures the strength of the links between the keywords of each cluster or evaluates their impact over time in the network. Lastly, the volume of the spheres is proportional to the frequency of publications associated with each research thematic. Figure 16 shows a strategy graph that represents the search sub-clusters in a bidirectional space.

Regarding the results, we consider the 200 most frequent keywords described in relation to author keywords. Figure 17 represents the strategic maps of the main thematic and trends topic. According to this figure, we have six main thematic (Industry 4.0, artificial neural networks, monitoring, deep learning, and machine learning) for the author keywords. However, when we focus on keywords Plus, we mainly have four thematic (the internet, neutral networks, system, classification, networks). Moreover, the keyword abstract or keyword titles extracted from the article's contents give three major thematic (monitoring, real-time, and data), and (times, learning, and monitoring). Furthermore, Table 12 shows the main emerging and motor topics related to the study case as well as their corresponding subgroups topics. We can deduct that I4.0 is an emerging or crossroad topic while monitoring technique, ML and DL approaches are the principals or motor topics. Finally, we can deduce that in recent years scientific

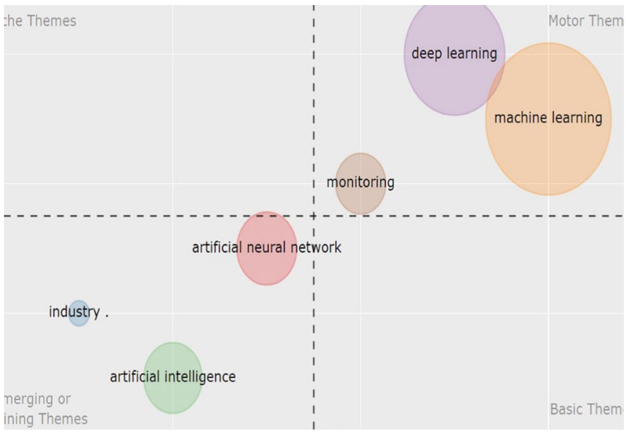

Fig. 17 Strategic map of the author's keyword. To interpret the results, and the color of the legends keyword in this figure, the reader is referred to the color

research has focused mainly on these cited subjects or topic.

6.7 Analyzes of ethical impact of the use of AI Techniques for PdM system

Ethical issues were initially not mentioned in the constitution of the initial query presented to the subsection 6.7. By performing a sub-query with the following keywords: "Ethical", "Artificial Intelligence", and "Industry 4.0" on the set of 4065 papers initially collected, we found a subset of 37 papers that address the ethics and trust based on AI models. Regarding the answering of the question **RQ6**, we exploited the results of this sub-query and the probable impact presented in the subsection 3.4. We can therefore show that AI systems and the industry's robotization will probably have several impacts on ethics, confidentiality, privacy rules, transparency, and human-robot collaboration. Furthermore, the industrialization may involve social-economic issues, particularly the increase in the unemployment rate, reduction of the workforce, and the evolution of the disparity between developed and Third-World countries.

6.8 Issues identified, key challenges and future research directions in PdM and I4.0

To answer the question **RQ7**, we showed the challenges associated with the deployment of AI Systems in I4.0 can be associated with several factors such as operational, organizational, technical, data collection or processing, cybersecurity, interpretability, trust, privacy, and ethical rules. (a) Operational and organizational: The growth and industrialization of the factory generally lead to reforms and changes in the human, operational, management and organizational levels [131]. These operators must be able to interact with professionals from other fields (multidisciplinary). Besides, companies must surround themselves with specialists (data scientist or experts) in the fields in which the solutions will be applied. (b) Machine-to-machine and human-to-machine interactions: It is essential to ensure that AI systems do not affect the functioning of other equipment or interconnected machines in the production process. Thus, industry should ensure that AI systems can interact or communicate with other devices while maintaining their behavior. Furthermore, workers must be trained or adapted to interact with these new technologies. (c) Cybersecurity and privacy: The exponential exploitation of interconnected technologies or storage systems such as IoT, sensors, databases, or big data infrastructure (local or cloud) can expose AI systems to cyberattacks notably through spamming or malicious software classification [132]. However, considerable efforts are being made to enforce ecosystem standards and guidelines such as the ISO/IEC 29180:2012 standards for sensor networks. Nowadays, there is really no standard or reliable process to ensure the security of AI models against attacks. Further questions are raised about the General Data Protection Regulation (GDPR). (d) Real-time data collection: Data are a key element in PdM, these data must be massive, secure, available, accessible, and qualitative to perform a generalizable PdM system. Data collection is therefore a major challenge for companies, since the sensors or machines do not often generate representative data on their conditions, deterioration, or configurations. In so doing, a possible solution is to label all the raw data although this

Table 12 Strategic map author's keywords. Each cluster is represented by a main thematic and its positioning in relation to the current literature

Author's keywords	Position
Cluster 1 Industry 4.0	
RF, IA, Monitoring, PdM, System, Cyber-physical object detectors, real-times Condition monitoring	Crossroads (Emerging theme)
Cluster 2 Artificial intelligent neural networks, recognition, prediction, diagnostics, image and signal processing Genetic algorithm	Crossroads (Motor theme)
Cluster 3 Monitoring Sensors, Real-time, System, Forecasting, Neutral networks	Principal (Motor theme)
Cluster 4 Deep learning LSTM, Feature extraction, Fault, Health monitoring, CNN, ANN, computer vision	Principal (Motor theme)
Cluster 5 Machine Learning IoT, Big Data, Data mining Anomalies detection, Remote, SVM, RF	Principal (Motor theme)

operation can be time-consuming, fastidious, and requires the knowledge of an expert. Furthermore, labeling operation involves risks of errors and a considerable economic and operational cost. Indeed, to address data quality issues, several approaches have been proposed, such as artificial re-sampling, interpolations techniques, semi-supervised learning, or data augmentation. Thus, when data are scarce, it is also possible to use GAN model to simulate reliable data [93], or transfer learning (different working conditions and machines) for the transfer of knowledge from one system to another [133]. Approaches cited are not systematically efficient, data simulation process is sometimes not adapted to the real conditions of a machine's operation since the imposed scenarios do not represent the complexity of the system (machine degradation or failure). (e) Adaptability of prescriptive and hybrid modeling in real time: it is important to develop prescriptive and hybrid models as a recommendation system for diagnosis, prognosis, and anomalies detection of machines in real time. Furthermore, hybrid models [134] have the benefit of integrating either both physical and numerical knowledge or constraints of the system combined with data-driven modeling. Furthermore, AI system must be able to adapt to the whole system while maintaining its performances. Adding new machine should not be an obstacle or a constraint which may impact the model's quality. (f) Using computer vision, multimodal prediction, images, videos processing, texts, and sound data in the PdM, data are collected from heterogeneous systems and are of diverse nature. The challenges of PdM are to combine all these data to perform a multimodal prediction. (g) Explainability XAI model: We have shown that some AI black box models such as CNN, or RF have a real interest in PdM problems. However, they are not easy to interpret and neither intuitive for all stakeholders. In this context explainability is an important factor for the acceptance of the AI solutions. Furthermore, a new trend of AI is focused on XAI explainability model-agnostic methods [135] such as SHapley Additive exPlanations (SHAP) [136] or Local Interpretable Model-Agnostic Explanations (LIME) [137] that are designed to explain and understand the black box model decision-making, and make it easily interpretable, comprehensible, and user-friendly for all stakeholders. Indeed, AI systems should not substitute humans, but support them in taking over low-level thinking tasks. In this regard, experts must collaborate with these new technologies to ensure productivity.

7 Discussion

This article focuses attention on a detailed bibliometric analysis based on using AI techniques for PdM in I4.0. The main objective is to highlight the evolution, impact and the current state-of-the-art of scientific research related to the exploitation of these technologies for anomalies detection, diagnosis, and PdM4.0. The results obtained give us a detailed analysis to address all the questions initially formulated. Furthermore, these results show a relative description of several metrics, including the publication trend, most productive journals, papers, authors, co-authors, references, affiliations, countries as well as network collaboration between authors or institutions. We have represented the most important keywords, conceptual, intellectual, and social structure of the research, including all past, principal, and emerging themes. Furthermore, we present the potential ethical impact rules using this AI system. Besides, we discuss the main challenges, and future research directions in AI applied to PdM4.0.

We analyzed the main information about the 4065 collected papers according to their dynamics, productivity, total citation, impact on the community and network collaborations. We have observed an exponential increase in the numbers of papers published in the last 20 years. We have 2308 source journals, however, the most productive are respectively IEEE Access (143), Sensors (119), and Applied Sciences-Basel (47). Indeed, we have shown that these journals are not necessarily the most cited. To have a high profile, and reputation in the scientific community, its journals are also open access, making it easy to view articles online. Concerning most cited, co-cited authors, as well as their collaboration network, and their impacts (subsection 6.4.1), we noted that Bellini, Filippetti, and Tassoni (694) are the most cited authors with the same TC metric. Furthermore, we analyzed affiliations, based on their productivity, impact, and collaboration. The most productive universities are respectively Illinois (81), Shanghai Jiao Tong (74) and California Los Angeles (71) university. As far as international collaboration between authors, or institutions, is concerned, institutions belonging to developing countries is not representative, notably Oceania with 399 articles (3%), and the African continent with 212 (2%) of the published articles. This trend of low productivity of universities, or institutions in Third-World countries can be partly explained by the fact that they are lagging in scientific research due to the lack of infrastructure, access to digital services such as the internet, energy, and the reputation of institutions in the scientific community. We can conclude that the institution in the USA, and China globally dominate the research on the topic.

Regarding the most cited, and productive articles, we have the following papers: [80, 111, 112, 138]. In particular, article [111] has been cited 694 times, the authors exploit AI models for monitoring, detection of electrical and mechanical defects. Article [112] which has been cited 419 times, exploits sensor data, and DL technique for the intelligent diagnosis of failures via regularized neural networks. Author [138] uses wavelet analysis and ANN models to predict the weld quality in friction stir welding. Paper [80], presents a hybrid model (RF combined to LSTM) for real-time monitoring, and corrective adjustment. In fact, the topic described in these papers deal with subject related to the digitization of industry using IA system. Moreover, the most cited references include [90, 128, 139–141]. To identify the most discussed topics and common technologies used in PdM, we extracted more than 11268 authors' keywords. Thus, we have analyzed these words including their co-occurrence networks. We can deduce that words such as ML, DL, monitoring, artificial neutral networks, data science, I4.0, IoT, sensors, big data, fault diagnosis, feature extraction, CNN, condition monitoring, predictive maintenance, anomaly detection, and real-time are the most frequent keywords. Moreover, the evolution of its keywords is associated with the main thematic of the study. We mainly have 6 topics; however, the principal and emerging themes are respectively monitoring, I4.0, DL and ML techniques. We have a real emergence of these approaches applied to monitoring, diagnostic technique, and PdM4.0. Indeed, the analyses suggest that the heterogeneity, and the link between these keywords reflect the importance of AI techniques in PdM4.0.

Additionally, we have observed that the industrialization and automated systems probably have an impact on the whole ecosystem of the industry. AI models applied to PdM systems some benefit, maintenance cost, and energy consumption reduction. Furthermore, they help to improve quality, to optimize, increase the efficiency or flexibility of production processes. On the human side, these models' impacts can be organizational, operational, security, trust, socio-economic, or legal. Indeed, authors [142] show that AI system will increase the workload of employees and create a need for adaptation and dependence on new technologies, particularly the challenge concerning transparency, human–robot collaboration raises ethical issues. In fact, [143] demonstrated that the transition to robotization of manufacturing systems is going to involve social-economic problems (generalized social exclusion) and reduction of the human workforce (massive unemployment). Also, the operators will have to re-adapt, be formed, and specialize with these new challenges or operational changes. Therefore, if the public authority does not adopt actions, the industry automation and AI system use will contribute to increasing the gap disparity between technologically advanced countries and under-developed countries.

8 Conclusion, limitation and future works orientations

The contribution of this paper is to provide a useful state-of-the-art basis for the literature search on the use of AI techniques applied to PdM in Industry 4.0. To address the research problem (RQ1–RQ7), we have performed a bibliometrics analysis using Biblioshiny, VOSviewer, and Power BI tools. This detailed analysis is based on 4096 scientific documents collected between 2000 and 2021 from the WoS Database. We focus on some metrics, including the publication trend, most productive sources, papers, authors, co-authors, references, affiliations, countries as well as network collaboration between authors or institutions. Furthermore, we analyze the most important keywords, and the principal or motor thematic associated with this study. We also analyze the benefits of AI models in the industry, their particularities, applications, impacts, and major results or performances. Particularly, we were also interested in ethical, trust, transparency and socio-economic impacts that could be caused when using these models. We give our definition of trustful AI for I4.0 and its effects. Finally, the potential limitations, key challenges and future research directions of AI systems are presented.

The results obtained showed a progressive increase in the frequency of publications over the last 20 years. Regarding the sources, we have shown that the IEEE access is the most productive and cited journal. Moreover, the most productive universities are respectively Illinois, and Shanghai Jiao Tong University. The USA and China are the countries with a major impact on scientific research related to the study topics. Indeed, the collaboration between developed and Third-World countries is very weak, while the international collaboration among developed countries is strengthened. For the author's analysis, we have observed that the most cited author, and reference are respectively Bellini and Lei. Furthermore, the analysis of collaboration network shows that some authors tend to work in small groups (three collaborators by group), which implies the large number of groups or cluster of authors. According to the author's keyword analysis, we show that the most important theoretical knowledge and research thematic on PdM4.0 are mainly in the areas of machine learning, and deep learning, including their sub-models. Moreover, we have 6 main topics among which the emerging themes are DL, ML, and monitoring. These different results clearly show that there is a wide field of applications (monitoring, diagnosis, prognostic, anomalies detection) or different situations, especially for supervised, unsupervised, or semi-supervised learning problems.

We have described the most common predictive models used in PdM 4.0. Despite their performance and application in many industrial cases, in practice, we have shown that some predictive models have several limitations, especially on their instability and overfitting in a situation for missing or noisy data, high volume, complex and unbalanced classes. In addition, they can have complex architectures, resulting in a significant requirement for GPU, and computing time, in the estimation of these parameters. Real-time or on-line analyses can become very complicated due to the high computing time and complexity of these models. Moreover, these models can interfere with the correct functioning of the system. Besides, most of these black box models are not explainable, i.e., the algorithm decision-making process is unknown. This can be a real issue for their generalization in industry. Finally, using AI technologies in the industry can also be confronted with some challenges, such as operational, organizational, adaptability, machine-to-machine, human-to-machine interactions, cybersecurity (risk attacks), analysis online, real-time data collection and data quality. Also, we have challenges concerning prescriptive, hybrid and multimodal modeling, visual reasoning, socio-economic, explainability XAI, interpretability, trust, privacy, GDPR data protection statements and ethical impacts.

8.1 Limitations

Concerning the main limitations, we performed a search with selected keywords according to the study context. However, we cannot guarantee that these keywords, and the scientific documents collected cover the whole research area. Moreover, we use an open-access journal WoS database which does not contain all the publications. Also, the scope of this research is limited to English papers collected from WoS and we used a traditional bibliometric approach to perform analyses, therefore, by combining the different methods we can considerably improve the results.

8.2 Future work orientation

In order to improve the results, we can refine the query by including more or accurate keywords. The exploitation of several bibliographic databases such as Scopus, Springer, Google scholar, Science Direct, and IEEE, as well as the selection of documents supplementary by including books, notes, and thesis will also be applied to retrieve all documents covering the field of study and improve the quality of analysis. We will also consider contributions written in languages, such as French, Chinese, Italian, Spanish, or German. Another area of improvement is to use a combination of several bibliometric analysis methods to strengthen the result.

Declarations

Conflict of interest There is no conflict of interest that might bias the described work: no financial, no personal and no other relationships with other people or organizations within three years of beginning the work submitted. We verify that the data contained in the manuscript being submitted have not been previously published, have not been submitted elsewhere and will not be submitted elsewhere while under consideration at AI and Ethics—Journal. We verify that all authors have reviewed the contents of the manuscript being submitted, approve of its contents, and validate the accuracy of the data. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Xu, L.D., Xu, E.L., Li, L.: Industry 4.0: state of the art and future trends. Int. J. Prod. Res. **56**(8), 2941–2962 (2018)
2. Rußmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., Harnisch, M.: Industry 4.0: the future of productivity and growth in manufacturing industries. Boston Consulting Group **9**(1), 54–89 (2015)
3. Hermann, M., Pentek, T., Otto, B.: Design principles for industrie 4.0 scenarios. In: 2016 49th Hawaii international conference on system sciences (HICSS), IEEE, pp. 3928–3937 (2016)
4. Orellana, F., Torres, R.: From legacy-based factories to smart factories level 2 according to the industry 40. Int. J. Comput. Integr. Manuf. **32**(4–5), 441–451 (2019)
5. Brettel, M., Klein, M., Friederichsen, N.: The relevance of manufacturing flexibility in the context of industrie 40. Proc. Cirp **41**, 105–110 (2016)
6. Szymanska, O., Adamczak, M., Cyplik, P.: Logistics 40-a new paradigm or set of known solutions? Res. Logist. Prod. **7**, 29–310 (2017)
7. Fischer, J., Obst, B., Lee, B.: Integrating material flow simulation tools in a service-oriented industrial context. In: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), IEEE, pp. 1135–1140 (2017)
8. Anderl, R.: Industrie 40–technological approaches, use cases, and implementation. Automatisierungstechnik **63**(10), 753–765 (2015)
9. Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffmann, M.: Industry 4.0. Bus. Inf. Syst. Eng. **6**(4), 239–242 (2014)
10. Sony, M., Naik, S.: Critical factors for the successful implementation of industry 4.0: a review and future research direction. Prod. Plan Control **31**(10), 799–815 (2020)
11. Adedoyin, F.F., Bekun, F.V., Driha, O.M., Balsalobre-Lorente, D.: The effects of air transportation, energy, ICT and FDI on economic growth in the industry 4.0 era: evidence from the united states. Technol. Forecast. Soc. Chang. **160**, 120297 (2020)
12. Wan, J., Cai, H., Zhou, K.: Industrie 4.0: enabling technologies. In: Proceedings of 2015 international conference on intelligent computing and internet of things, IEEE, pp. 135–140 (2015)
13. Radanliev, P., De Roure, J., Nurse, R., Nicolescu, R., Huth, M., Cannady, S., Montalvo, R.M.: New developments in cyber physical systems, the internet of things and the digital economy–discussion on future developments in the industrial internet of things and industry 4.0 (2019)
14. Oztemel, E., Gursev, S.: Literature review of industry 4.0 and related technologies. J. Intell. Manuf. **31**(1), 127–182 (2020)
15. Lund, D., MacGillivray, C., Turner, V., Morales, M.: Worldwide and regional internet of things (iot) 2014–2020 forecast: a virtuous circle of proven value and demand. Int. Data Corp. (IDC) Tech Rep **1**, 9 (2014)
16. Sezer, E., Romero, D., Guedea, F., Macchi, M., Emmanouilidis, C.: An industry 40-enabled low cost predictive maintenance approach for smes. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), IEEE, pp. 1–8 (2018)
17. Zonta, T., da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S., Li, G.P.: Predictive maintenance in the industry 4.0: A systematic literature review. Comput. Ind. Eng. (2020) 106889

18. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. Harv. Bus. Rev. **92**(11), 64–88 (2014)
19. Sajid, S., Haleem, A., Bahl, S., Javaid, M., Goyal, T., Mittal, M.: Data science applications for predictive maintenance and materials science in context to industry 4.0. Mater Today Proc (2021)
20. Hossain, M.S., Muhammad, G.: Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring. Comput. Netw. **101**, 192–202 (2016)
21. Trotta, D., Garengo, P.: Industry 4.0 key research topics: a bibliometricreview. In: 2018 7th international conference on industrial technology and management (ICITM), IEEE, pp. 113–117 (2018)
22. Souza, M.L.H., da Costa, C.A., de Oliveira Ramos, G., Righi, R.: A survey on decision-making based on system reliability in the context of industry 4.0. J. Manuf. Syst. **56**, 133–156 (2020)
23. Muhuri, P.K., Shukla, A.K., Abraham, A.: Industry 4.0: a bibliometric analysis and detailed overview. Eng. Appl. Artif. Intell. **78**, 218–235 (2019)
24. Mariani, M., Borghi, M.: Industry 4.0: a bibliometric review of its managerial intellectual structure and potential evolution in the service industries. Technol. Forecast. Soc. Chang. **149**, 119752 (2019)
25. Cezanne, C., Lorenz, E., Saglietto, L.: Exploring the economic and social impacts of industry 4.0. Rev. Deconomie Ind. **1**, 11–35 (2020)
26. Ejsmont, K., Gladysz, B., Kluczek, A.: Impact of industry 4.0 on sustainability—bibliometric literature review. Sustainability **12**(14), 5650 (2020)
27. Grubisic, V., Aguiar, J., Simeu-Abazi, Z.: A review on intelligentpredictive maintenance: bibliometric analysis and new research directions. In: 2020 International Conference on Control, Automation and Diagnosis (ICCAD), IEEE, 2020, pp. 1–6
28. Noman, M.A., Nasr, E.S.A., Al-Shayea, A., Kaid, H.: Overview of predictive condition based maintenance research using bibliometric indicators. J. King Saud Univ. Eng. Sci. **31**(4), 355–367 (2019)
29. Yu, D., Xu, Z., Fujita, H.: Bibliometric analysis on the evolution of applied intelligence. Appl. Intell. **49**(2), 449–462 (2019)
30. Bertolini, M., Mezzogori, D., Neroni, M., Zammori, F.: Machine learning for industrial applications: a comprehensive literature review. Expert Syst. Appl. (2021) 114820
31. Kamat, P., Sugandhi, R.: Bibliometric analysis of bearing fault detection using artificial intelligence. Library Philosophy and Practice (2020) 1–21
32. Aria, M., Cuccurullo, C.: bibliometrix: an r-tool for comprehensive science mapping analysis. J. Informet. **11**(4), 959–975 (2017)
33. Van Eck, N.J., Waltman, L.: Software survey: Vosviewer, a computer program for bibliometric mapping. Scientometrics **84**(2), 523–538 (2010)
34. Haleem, A., Javaid, M.: Additive manufacturing applications in industry 4.0: a review. J. Ind. Integr. Manag. **4**(4), 1930001 (2019)
35. Javaid, M., Haleem, A.: Impact of industry 40 to create advancements in orthopaedics. J. Clin. Orthop. Trauma **11**, S491–S499 (2020)
36. Strange, R., Zucchella, A.: Industry 4.0, global value chains and international business. Multinatl. Bus. Rev. (2017)
37. Lu, Y.: Cyber physical system (cps)-based industry 4.0: a survey. J. Ind. Integr. Manag. **2**(03), 1750014 (2017)
38. Stenstro¨m, C., Parida, A., Kumar, U., Galar, D.: Performance indicators and terminology for value driven maintenance. J. Qual. Maint. Eng. (2013)
39. Wang, K.-S., Li, Z., Braaten, J., Yu, Q.: Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using anns. Adv. Manuf. **3**(2), 97–104 (2015)
40. Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: 2012 IEEE conference on computer vision and pattern recognition, IEEE, pp. 2518–2525 (2012)
41. Ayad, S., Terrissa, L.S., Zerhouni, N.: An IOT approach for a smart maintenance. In: 2018 International conference on advanced systems and electric technologies (IC ASET), IEEE, pp. 210–214 (2018)
42. Haarman, M., de Klerk, P., Decaigny, P., Mulders, M., Vassiliadis, C., Sijtsema, H., Gallo, I.: Predictive maintenance 4.0-beyond the hype: Pdm 4.0 delivers results. PricewaterhouseCoopers and Mannovation (2018)
43. De Visser, E.J., Pak, R., Shaw, T.H.: From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction. Ergonomics **61**(10), 1409–1427 (2018)
44. Mcconnell, T.: Moral dilemmas [online]. USA: Stanford university: Center for the study of language and information (2014)
45. H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, Q. Yang, Building ethics into artificial intelligence, arXiv preprint arXiv:1812.02953 (2018)
46. Cointe, N., Bonnet, G., Boissier, O.:Ethical judgment of agents' behaviors in multi-agent systems. In: Proceedings of the 2016 international conference on autonomous agents & multiagent systems, pp. 1106–1114 (2016)
47. Saßmannshausen, T., Burggra¨f, P., Wagner, J., Hassenzahl, M., Heupel, T., Steinberg, F.: Trust in artificial intelligence within production management–an exploration of antecedents. Ergonomics (2021) 1–18
48. Toro, C., Sanin, C., Vaquero, J., Posada, J., Szczerbicki, E.: Knowledge based industrial maintenance using portable devices and augmented reality. In: International conference on knowledge-based and intelligent information and engineering systems, Springer, pp. 295–302 (2007)
49. Mirhosseyni, S.H.L., Webb, P.: A hybrid fuzzy knowledge-based expert system and genetic algorithm for efficient selection and assignment of material handling equipment. Expert Syst. Appl. **36**(9), 11875–11887 (2009)
50. Sa´rfi, R.J., Solo, A., Wilmington, D.: The application of fuzzy logic in a hybrid fuzzy knowledge-based system for multiobjective optimization of power distribution system operations. In: Proceedings of the 2005 International Conference on Information and Knowledge Engineering, pp. 3–9 (2005)
51. Welte, T.: A rule-based approach for establishing states in a markov process applied to maintenance modelling. Proc. Inst. Mech. Eng. Part O J. Risk Reliab. **223**(1), 1–12 (2009)
52. Yu, R., Iung, B., Panetto, H.: A multi-agents based e-maintenance system with case-based reasoning decision support. Eng. Appl. Artif. Intell. **16**(4), 321–333 (2003)
53. Sergaki, A., Kalaitzakis, K.: A fuzzy knowledge based method for maintenance planning in a power system. Reliab. Eng. Syst. Saf. **77**(1), 19–30 (2002)
54. Ruppert, T., Abonyi, J.: Software sensor for activity-time monitoring and fault detection in production lines. Sensors **18**(7), 2346 (2018)
55. Ka¨ßmeyer, M., Berndt, R., Bazan, P., German, R.: Product line fault tree analysis by means of multi-valued decision diagrams. In: International GI/ITG conference on measurement, modelling, and evaluation of computing systems and dependability and fault tolerance, Springer, pp. 122–136 (2016)
56. Tinga, T.: Application of physical failure models to enable usage and load based maintenance. Reliab. Eng. Syst. Saf. **95**(10), 1061–1075 (2010)
57. Qin, H., Xu, Z., Li, P., Yu, S.: A physical model approach to nonlinear vertical accelerations and mooring loads of an offshore

aquaculture cage induced by wave-structure interactions. Ocean Eng. **197**, 106904 (2020)
58. Coscia, J.L.O., Crasso, M., Mateos, C., Zunino, A., Misra, S.: Predicting web service maintainability via object-oriented metrics: a statistics-based approach. In: International conference on computational science and its applications, Springer, pp. 29–39 (2012)
59. Namuduri, S., Narayanan, B.N., Davuluru, V.S.P., Burton, L., Bhansali, S.: Deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors. J. Electrochem. Soc. **167**(3), 37552 (2020)
60. Simeu-Abazi, Z., Sassine, C.: Maintenance integration in manufacturing systems by using stochastic petri nets. Int. J. Prod. Res. **37**(17), 3927–3940 (1999)
61. Accorsi, R., Manzini, R., Pascarella, P., Patella, M., Sassi, S.: Data mining and machine learning for condition-based maintenance. Proc. Manuf. **11**, 1153–1161 (2017)
62. Dowdeswell, B., Sinha, R., MacDonell, S.G.: Finding faults: a scoping study of fault diagnostics for industrial cyber–physical systems. J. Syst. Softw. **168**, 110638 (2020)
63. Raissi, M., Karniadakis, G.E.: Hidden physics models: Machine learning of nonlinear partial differential equations. J. Comput. Phys. **357**, 125–141 (2018)
64. Chen, Y.-G.: Applications of bayesian network in fault diagnosis of braking system. In: 2011 Third international conference on intelligent human-machine systems and cybernetics, Vol. 1, IEEE, pp. 234–237 (2011)
65. Aivaliotis, P., Georgoulias, K., Alexopoulos, K.: Using digital twin for maintenance applications in manufacturing: State of the art and gap analysis. In: 2019 IEEE international conference on engineering, technology and innovation (ICE/ITMC), IEEE, pp. 1–5 (2019)
66. Luo, W., Hu, T., Ye, Y., Zhang, C., Wei, Y.: A hybrid predictive maintenance approach for cnc machine tool driven by digital twin. Robot. Comput. Integr. Manuf. **65**, 101974 (2020)
67. Vathoopan, M., Johny, M., Zoitl, A., Knoll, A.: Modular fault ascription and corrective maintenance using a digital twin. IFAC-Papers Online **51**(11), 1041–1046 (2018)
68. Heirung, T.A.N., Mesbah, A.: Input design for active fault diagnosis. Annu. Rev. Control. **47**, 35–50 (2019)
69. Zhou, G., Feng, W., Zhao, Q., Zhao, H.: State tracking and fault diagnosis for dynamic systems using labeled uncertainty graph. Sensors **15**(11), 28031–28051 (2015)
70. Benowitz, E.: The curiosity mars rover's fault protection engine. In: 2014 IEEE international conference on space mission challenges for information technology, IEEE, 2014, pp. 62–66
71. Schulte, P.Z.: A state machine architecture for aerospace vehicle fault protection, Ph.D. thesis, Georgia Institute of Technology (2018)
72. Novikov, D., Yampolskiy, R.V., Reznik, L.: Anomaly detection based intrusion detection. In: Third international conference on information technology: new generations (ITNG'06), IEEE, pp. 420–425 (2006)
73. Reif, M., Goldstein, M., Stahl, A., Breuel, T. M.: Anomaly detection by combining decision trees and parametric densities. In: 2008 19th international conference on pattern recognition, IEEE, 2008, pp. 1–4
74. Dhanalakshmi, Y., Babu, I.R.: Intrusion detection using data mining along fuzzy logic and genetic algorithms. Int. J. Comput. Sci. Netw. Secur. **8**(2), 27–32 (2008)
75. Cakir, M., Guvenc, M.A., Mistikoglu, S.: The experimental application of popular machine learning algorithms on predictive maintenance and the design of iiot based condition monitoring system. Comput Ind Eng **151**, 106948 (2021)
76. Agrawal, S., Agrawal, J.: Survey on anomaly detection using data mining techniques. Proc. Comput. Sci. **60**, 708–713 (2015)
77. Wang, Q., Jiao, W., Wang, P., Zhang, Y.: A tutorial on deep learning-based data analytics in manufacturing through a welding case study. J. Manuf. Process. **63**, 2–13 (2021)
78. Long, J., Mou, J., Zhang, L., Zhang, S., Li, C.: Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots. J. Manuf. Syst. (2020)
79. Shakya, S., Sigdel, S.: An approach to develop a hybrid algorithm basedon support vector machine and naive bayes for anomaly detection. In: 2017 International Conference on Computing, Communication and Automation (ICCCA), IEEE, pp. 323–327 (2017)
80. Chiu, M.-C., Tsai, C.-D., Li, T.-L.: An integrative machine learning method to improve fault detection and productivity performance in a cyber-physical system. J. Comput. Inf. Sci. Eng. 20(2) (2020)
81. Jung, M., Niculita, O., Skaf, Z.: Comparison of different classification algorithms for fault detection and fault isolation in complex systems. Proc. Manuf. **19**, 111–118 (2018)
82. Giannetti, C., Ransing, R.S.: Risk based uncertainty quantification to improve robustness of manufacturing operations. Comput. Ind. Eng. **101**, 70–80 (2016)
83. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
84. Kim, J.S., Kim, J., Lee, J.Y.: Die-casting defect prediction and diagnosis system using process condition data. Proc. Manuf. **51**, 359–364 (2020)
85. Ayvaz, S., Alpay, K.: Predictive maintenance system for production lines in manufacturing: a machine learning approach using iot data in real-time. Expert Syst. Appl. **173**, 114598 (2021)
86. Vapnik, V.: The nature of statistical learning theory, Springer science & business media (1999)
87. Gryllias, K.C., Antoniadis, I.A.: A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. Eng. Appl. Artif. Intell. **25**(2), 326–344 (2012)
88. Salcedo-Sanz, S., Rojo-Alvarez, J.L., Martinez-Ramon, M., Campus-Valls, G.: Support vector machines in engineering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov **4**(3), 234–267 (2014)
89. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 175–185 (1992)
90. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(75), 436–444 (2015)
91. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European conference on computer vision, Springer, pp. 818–833 (2014)
92. Zheng, J., Dai, Y., Liang, Y., Liao, Q., Zhang, H.: An online real-time estimation tool of leakage parameters for hazardous liquid pipelines. Int. J. Crit. Infrastruct. Prot. **31**, 100389 (2020)
93. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan, arXiv preprint arXiv: 1701.07875 (2017)
94. Koskinen, J., Isohanni, M., Paajala, H., Jaaskelainen, E., Nieminen, P., Koponen, H., Tienari, P., Miettunen, J.: How to use bibliometric methods in evaluation of scientific research? an example from finnish schizophrenia research. Nordic J Psychiatry **62**(2), 136–143 (2008)
95. Amin, M.T., Khan, F., Zuo, M.J.: A bibliometric analysis of process system failure and reliability literature. Eng. Fail. Anal. **106**, 104152 (2019)
96. Garfield, E.: Keywords plus-isi's breakthrough retrieval method 1 expanding your searching power on current-contents on diskette. Curr. Cont. **32**, 5–9 (1990)
97. Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., Duan, Z.: Comparing keywords plus of wos and author keywords: a case study of patient adherence research. J. Am. Soc. Inf. Sci. **67**(4), 967–972 (2016)

98. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Science mapping software tools: review, analysis, and cooperative study among tools. J. Am. Soc. Inf. Sci. Technol. **62**(7), 1382–1402 (2011)
99. Atzeni, G., Vignali, G., Tebaldi, L., Bottani, E.: A bibliometric analysis on collaborative robots in logistics 4.0 environments. Proc. Comput. Sci. **180**, 686–695 (2021)
100. Wang, X., Xu, Z., Skare, M.: A bibliometric analysis of economic research-ekonomska istrazˇivanja (2007–2019). Econ Res Ekonomska Istrazˇivanja **33**(1), 865–886 (2020)
101. Cobo, M.J., Jurgens, B., Herrero-Solana, V., Martinez, M.A., Herrera-Viedma, E.: Industry 4.0: a perspective based on bibliometric analysis. Proc. Comput. Sci. **139**, 364–371 (2018)
102. Riahi, Y., Saikouk, T., Gunasekaran, A., Badraoui, I.: Artificial intelligence applications in supply chain: a descriptive bibliometric analysis and future research directions. Expert Syst. Appl. (2021) 114702
103. Zupic, I., Cater, T.: Bibliometric methods in management and organization. Organ. Res. Methods **18**(3), 429–472 (2015)
104. Ab Razak, M.F., Anuar, N.B., Salleh, R., Firdaus, A.: The rise of "malware": Bibliometric analysis of malware study. J Netw Comput Appl **75**, 58–76 (2016)
105. Ale Ebrahim, N., Salehi, H., Embi, M.A., Habibi, F., Gholizadeh, H., Motahar, S.M., Ordi, A.: Effective strategies for increasing citation frequency. Int. Educ. Stud. **6**(11), 93–99 (2013)
106. Mongeon, P., Paul-Hus, A.: The journal coverage of bibliometric databases: a comparison of scopus and web of science. J Cover Web Sci Scopus Comp Anal 10 (2014)
107. Choi, J., Yi, S., Lee, K.C.: Analysis of keyword networks in mis research and implications for predicting knowledge evolution. Inf. Manag. **48**(8), 371–381 (2011)
108. Chen, G., Xiao, L., Hu, C.-P., Zhao, X.-Q.: Identifying the research focus of library and information science institutions in china with institution specific keywords. Scientometrics **103**(2), 707–724 (2015)
109. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. **24**(5), 513–523 (1988)
110. Chen, G., Xiao, L.: Selecting publication keywords for domain analysis in bibliometrics: a comparison of three methods. J. Informet. **10**(1), 212–223 (2016)
111. Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.-A.: Advances in diagnostic techniques for induction machines. IEEE Trans. Industr. Electron. **55**(12), 4109–4126 (2008)
112. Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S.X.: An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans. Industr. Electron. **63**(5), 3137–3147 (2016)
113. Zheng, Y., Liu, F., Hsieh, H.-P.: U-air: When urban air quality inference meets big data proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining (kdd'13), ACM, New York, NY, USA (2013)
114. Windmiller, J.R., Wang, J.: Wearable electrochemical sensors and biosensors: a review. Electroanalysis **25**(1), 29–46 (2013)
115. Muller, K., Dornhege, M., Krauledat, G., Curio, M., Blankertz, G.: Machine learning for real-time single-trial EEG-analysis: From Brain-computer interfacing to mental state monitoring. J Neuro Sci Methods **167**(1), 82–90 (2008)
116. Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., Inman, D.J.: Real- time vibration-based structural damage detection using one-dimensional convolutional neural networks. J. Sound Vib. **388**, 154–170 (2017)
117. Bigio, I.J., Bown, S.G., Briggs, G.M., Kelley, C., Lakhani, S., Pickard, D.C.O., Ripley, P.M., Rose, I., Saunders, C.: Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. J. Biomed. Opt. **5**(2), 221–228 (2000)
118. Verrelst, J., Munoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J.: Machine learning regression algorithms for biophysical parameter retrieval: opportunities for sentinel-2 and-3. Rem Sens Environ **118**, 127–139 (2012)
119. Khan, S., Yairi, T.: A review on the application of deep learning in system health management. Mech. Syst. Signal Process. **107**, 241–265 (2018)
120. Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.-W.: An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. **569**, 387–408 (2019)
121. Berg, B., Cortazar, B., Tseng, D., Ozkan, H., Feng, S., Wei, Q., Chan, R.Y.-L., Burbano, J., Farooqui, Q., Lewinski, M., et al.: Cellphone-based hand-held microplate reader for point-of-care testing of enzyme-linked immunosorbent assays. ACS Nano **9**(8), 7857–7866 (2015)
122. Jin, Z., Sun, Y., Cheng, A.C.: Predicting cardiovascular disease from real-time electrocardiographic monitoring: an adaptive machine learning approach on a cell phone. In: 2009 Annual international conference of the IEEE engineering in medicine and biology society, IEEE, pp. 6889–6892 (2009)
123. Jing, L., Zhao, M., Li, P., Xu, X.: A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement **111**, 1–10 (2017)
124. Gonzaga, J., Meleiro, L.A.C., Kiang, C., Maciel Filho, R.: Ann-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. Comput. Chem. Eng. **33**(1), 43–49 (2009)
125. He, J., Gong, S., Yu, Y., Yu, L., Wu, L., Mao, H., Song, C., Zhao, S., Liu, H., Li, X., et al.: Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major chinese cities. Environ. Pollut. **223**, 484–496 (2017)
126. Michie, S., Yardley, L., West, R., Patrick, K., Greaves, F.: Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J. Med. Internet Res. **19**(6), 232 (2017)
127. Botu, V., Ramprasad, R.: Adaptive machine learning framework to accelerate ab initio molecular dynamics. Int. J. Quantum Chem. **115**(16), 1074–1083 (2015)
128. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015), arXiv preprint arXiv: 1512.03385 (2016)
129. Greenacre, M., Blasius, J.: Multiple correspondence analysis and related methods, CRC press (2006)
130. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. J. Informetr. **5**(1), 146–166 (2011)
131. Ivanov, D., Tang, C.S., Dolgui, A., Battini, D., Das, A.: Researchers' perspectives on industry 4.0: multi-disciplinary analysis and opportunities for operations management. Int. J. Prod. Res. **59**(7), 2055–2078 (2021)
132. Tuptuk, N., Hailes, S.: Security of smart manufacturing systems. J. Manuf. Syst. **47**, 93–106 (2018)
133. Zhu, J., Chen, N., Shen, C.: A new deep transfer learning method for bearing fault diagnosis under different working conditions. IEEE Sens. J. **20**(15), 8394–8402 (2019)
134. Wang, T.: Hybrid decision making: When interpretable models collaborate with black-box models, CoRR, Arxiv abs/1802.04346 (2018)
135. Slack, D., Hilgard, S., Jia, E., Singh, S., Lakkaraju, H.: Fooling lime and shap: Adversarial attacks on post hoc explanation methods. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 180–186 (2020)

136. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I.: From local explanations to global understanding with explainable ai for trees. Nat. Mach. Intell. **2**(1), 56–67 (2020)
137. Ribeiro, M.T., Singh, S., Guestrin, C.: Model-agnostic interpretability of machine learning, arXiv preprint arXiv: 1606.05386 (2016)
138. Das, B., Pal, S., Bag, S.: Weld quality prediction in friction stir welding using wavelet analysis. Int. J. Adv. Manuf. Technol. **89**(1–4), 711–725 (2017)
139. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
140. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
141. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
142. Wang, K., Shu, Q., Tu, Q.: Technostress under different organizational environments: an empirical investigation. Comput. Hum. Behav. **24**(6), 3002–3013 (2008)
143. Kaplan, A., Haenlein, M.: Digital transformation and disruption: On bigdata, blockchain, artificial intelligence, and other things (2019)