Leoni, Leonardo; BahooToroody, Ahmad; Abaei, Mohammad Mahdi; Cantini, Alessandra; BahooToroody, Farshad; De Carlo, Filippo

Machine learning and deep learning for safety applications: Investigating the intellectual structure and the temporal evolution

Published in:
Safety Science

DOI:
10.1016/j.ssci.2023.106363

Published: 01/02/2024

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Review

Machine learning and deep learning for safety applications: Investigating the intellectual structure and the temporal evolution

Leonardo Leoni a, Ahmad BahooToroody b, Mohammad Mahdi Abaei c, Alessandra Cantinia, d, Farshad BahooToroody e, Filippo De Carlo a

a Department of Industrial Engineering (DIEF), University of Florence, Italy
b Marine and Arctic Technology Group, Department of Mechanical Engineering, Aalto University, Finland
c Department of Geography and Geology, University of Turku, Turun Yliopisto, Finland
d Department of management, economics and industrial engineering, Politecnico di Milano, Via Lambruschini 4B, 20156, Milan, Italy
e Priority Research Centre for Geotechnical Science and Engineering, The University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Safety engineering
Prognostics and health management
Systematic bibliometric analysis

ABSTRACT

Over the last decades, safety requirements have become of primary concern. In the context of safety, several strategies could be pursued in many engineering fields. Moreover, many techniques have been proposed to deal with safety, risk, and reliability matters, such as Machine Learning (ML) and Deep Learning (DL). ML and DL are characterised by a high variety of algorithms, adaptable for different purposes. This generated wide and fragmented literature on ML and DL for safety purposes, moreover, literature review and bibliometric studies of the past years mainly focus on a single research area or application field. Thus, this paper aims to provide a holistic understanding of the research on this topic through a Systematic Bibliometric Analysis (SBA), along with proposing a viable option to conduct SBAs. The focus is on investigating the main research areas, application fields, relevant authors and studies, and temporal evolution. It emerged that rotating equipment, structural health monitoring, batteries, aeroengines, and turbines are popular fields. Moreover, the results depicted an increase in popularity of DL along with new approaches such as deep reinforcement learning through the past four years. The proposed workflow for SBA has the potential to benefit researchers from multiple disciplines, beyond safety science.

1. Introduction

During the past decades, safety requirements have progressively become more important (Soltanali et al., 2020). This trend could be related to the introduction of increasingly more complex systems, leading to the inadequacy of traditional approaches (Luther et al., 2023). The former trend could have also been facilitated by the advances in Artificial Intelligence, which has impacted many aspects of our society (Dimiduk et al., 2018), such as engineering processes. Due to the former aspects, reliability has acquired a more pivotal role (Huang et al., 2020). Indeed, sound reliability and risk analyses are essential to reduce the risk arising from undesired events (e.g., failures) and ensure the safety of the operations. In this context, different strategies could be followed such as conducting fault diagnosis (Han and Li, 2022), health management (BahooToroody et al., 2022), predicting future conditions or predicting the Remaining Useful Life (RUL) (Hu et al., 2022), crack or damage detection (Wu et al., 2019), or planning and optimising maintenance activities (Arena et al., 2022). Moreover, many different concepts and analyses have been introduced such as availability, resilience, vulnerability, accident modelling, and precursor accident data. All these aspects generated fragmented literature spread across different research areas, which are shared by multiple engineering branches and application fields. Examples include rotating machineries (Wang et al., 2020), railways (Fecarotti et al., 2021), batteries (Schismenos et al., 2021), ships (Chen et al., 2019), and aeroengines (Chen et al., 2023).

In the former application fields and for the aforementioned purposes, different tools and approaches have been adopted such as Fault Tree (FT) (Andrews and Tolo, 2023), Bayesian Network (BN) (Leon et al., 2019), and Failure Mode and Effect Analysis (FMEA) (Catelani et al., 2018). Among the proposed approaches, Machine Learning (ML) and Deep Learning (DL) are prominent. Indeed, they present several advantages such as enabling better decision making-process thanks to the...

a Correspondent author.
E-mail address: filippo.decarlo@unifi.it (F. De Carlo).

https://doi.org/10.1016/j.ssci.2023.106363
Received 19 June 2023; Received in revised form 22 September 2023; Accepted 18 October 2023
Available online 6 November 2023
0925-7535/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Bibliometric analyses are very suitable for extended fields, which may be usually able to provide more objective results (Huang et al., 2020). Bibliometric and scientometric analyses, nevertheless, the latter are different approaches. Literature reviews are more popular compared to considering just ML or DL, or they tackle a single field of application (e.g., specific research areas (e.g., risk assessment or fault diagnosis), or they edge on safety matters. However, most of the review works focus on monitoring of wind turbines. Zhang et al. (2020) and Hoang and Kang studied the application of ML for engineering risk assessment, while safety, risk, and reliability. For instance, Hegde and Rokseth (2020) regarding the adoption of ML or DL for different purposes related to capability applications. DL is developed from traditional NN. Moreover, thanks to a set of two or more non-linear transformation, DL could overcome typical limitations arising from the analysis of huge datasets (Ahmed, 2023). Other than CNN and RNN, the following DL techniques could also be found: deep belief networks (Che et al., 2020), Generative Adversarial Networks (GANs) (Che et al., 2020), stacked auto-encoders (Cui et al., 2020), graph neural networks (Li et al., 2022), deep reinforcement learning (Qian and Liu, 2022), and convolutional long short-term memory neural networks (Hao et al., 2020). It follows that ML and DL show high variety. The variety is also increased by the different areas characterising safety, risk, and reliability applications. Over the past years, different literature reviews were published regarding the adoption of ML or DL for different purposes related to safety, risk, and reliability. For instance, Hegde and Rokseth (2020) studied the application of ML for engineering risk assessment, while Stetcu (2019) investigated the application of ML for the condition monitoring of wind turbines. Zhang et al. (2020) and Hoang and Kang (2019) reviewed the application of DL for the failure diagnosis of bearings, while Flah et al. (2021) addressed the adoption of ML in the context of civil SHM. Recently, a more generic treatment of ML for reliability and safety applications is proposed by Xu and Saleh (2021), but only a brief discussion on DL is presented. Similarly, bibliometric and scientometric analyses have also been carried out during the last decade. Examples include bibliometric analyses related to fault diagnosis of bearings (Kamat and Sugandhi, 2020), DL for crack detection (Ali et al., 2022), or Prognostic and Health Management (PHM) of hydrogen fuel engine (Wang et al., 2022). All the individual review studies allow defining research gaps, future opportunities, and guidelines. Accordingly, scholars could exploit them to expand their knowledge on safety matters. However, most of the review works focus on specific research areas (e.g., risk assessment or fault diagnosis), or they consider just ML or DL, or they tackle a single field of application (e.g., bearing or maritime). Thus, a holistic vision of the application of both ML and DL for safety, risk, and reliability purposes is still missing. Indeed, summarising and classifying the literature could be deemed as a difficult task due to its fragmented nature, the variety of topics, and the different approaches. Literature reviews are more popular compared to bibliometric and scientometric analyses, nevertheless, the latter are usually able to provide more objectives results (Huang et al., 2020). Bibliometric analyses are very suitable for extend fields, which may be difficult to summarize through classic approaches (Belter and Seidel, 2013). Bibliometric and scientometric analyses have gained attention to discover emerging patterns (Haghanian and Bliemer, 2020) or identify relevant research themes and bibliometric attributes (Kumar et al., 2021). To this end, they often involved co-words analysis (Chellappandi and Vijayakumar, 2018). This is usually achieved by analysing titles, abstracts, and keywords of scientific documents. In this context, a simple term-based strategy could result in a high number of false positives, while missing relevant studies (Haghanian et al., 2021). Thus, elaborate and systematic search approaches could be useful to maximising the coverage, while reducing the number of false positives.

Based on the former considerations, this paper aims to provide a holistic view of the literature related to ML and DL applications in the context of safety, risk, and reliability. To this end, a bibliometric analysis is adopted to investigate the intellectual structure of the field, along with analysing its temporal trends. Accordingly, this work studies the (1) primary research areas, fields of application, and adopted algorithms; (2) the trend of publications over time; (3) the most relevant journals, authors, and studies; and (4) the temporal evolution to grasp possible past, current, and future trends. A possible approach to conduct a Systematic Bibliometric Analysis (SBA) is proposed with the objective to maximise the coverage and minimise the number of false positives.

As a reminder, the rest of the paper is organised as follows: in Section 2, the research methodology adopted to conduct the SBA is described, while in Section 3 the results related to topics, research areas, main authors and studies, and temporal evolution are presented. Finally, the discussion of the results is provided in Section 4, while Section 5 draws the conclusions and the limitations of the present study.

2. Research methodology: Systematic bibliometric analysis

The literature on ML and DL for safety purposes is vast, and it includes many fields and disciplines scattered across several specialty journals. Thus, gathering and categorising information could be deemed as a difficult task. This work aims to propose a viable search option to assure a high coverage while reducing the number of false positives which could alter the finding of the bibliometric analysis. The search strategy proposed in this study is referred to as SBA, and it is composed of six steps as shown in Fig. 1. The search was conducted through the Web of Science (WoS) Core Collection, which includes more than 21,000 journals and over 87 million records from 1900 up to present days.

Step 1. Following the approach proposed by Haghanian et al. (2021), five journals that deal with safety, risk, and reliability were chosen to extract a first set of keywords to include and exclude. The selected journals are Reliability Engineering and System Safety (RESS), Safety Science (SS), Process Safety and Environmental Protection (PSEP), Journal of Loss Prevention in the Process Industries (JLPPI), and Engineering Failure Analysis (EFA). Since the former journals include also studies unrelated to ML or DL, the following preliminary string was used: TS = (“machine learning” OR “deep learning” OR “statistical learning”). The string resulted in 359 papers, whose titles, abstracts, and keywords were analysed to select the keywords to include and exclude depending on the topic treated in the related papers. In case there were doubts about the consistency of a paper, the full text was retrieved for further confirmation. For instance, a category of papers published in PSEP deal with environmental protection, which is outside of the scope of this research was excluded. Crash safety and occupational safety were also excluded from this study since they could also treat psychological or behavioural aspects. Subsequently, the keywords were listed in five major categories: (i) method-related keywords (e.g., “support vector machine”), (ii) sector, field, or context related keywords (e.g., “engine”, “chemical process”, or “railway”), (iii) general purpose (e.g., “reliability analysis” or “risk evaluation”), (iv) standalone (e.g., “flight safety” or “structural health monitoring”), and (v) keywords to
exclude. The first category is related to tools, algorithms, or methods that are included in the ML or DL field. The second group of keywords is related to the context of application since it is important to include the desired context and exclude the undesired ones (e.g., “occupational safety”). The third group is related to the purpose of the ML or DL application. Finally, the fourth class of keywords is a mix of the second and third classes since it includes both the context and the purpose. This division is pivotal for the subsequent search in the entire WoS database. Indeed, the search string is formed: (TS = “method keywords”) AND ((TS = “sector/field/context keywords” AND TS = “general purpose”) OR TS = “standalone”) NOT TS = (“keywords to exclude”). Each category includes a list of keywords separated by an OR logic operator. It is worth mentioning that during this phase and step 4, the keywords were sometimes slightly manipulated to be more inclusive or exclusive. For instance, “wear behaviour prediction” became “wear behaviour predict*” or “psychology” became “psycholog*”. British and American English were included (e.g., “wear behaviour predict*” and “wear behavior predict*”). Finally, for the keywords to exclude, also similar words were added. For instance, if the term “molecular” was found, the term “molecule” was added. The developed search string can be found in the Appendix.

Fig. 1. Developed framework for conducting a SBA.

Fig. 2. Steps required to obtain preliminary research string.
Step 2. The search string is extended to the entire WoS database, considering only studies written in English and up to 2022. A visual summary of the first two steps is provided in Fig. 2.

Step 3. Ten samples of 50 random papers are extracted from the results. The papers are studied in terms of title, abstract, and keywords to evaluate the coherence with the topic of the current work. In case a paper was marked as a false positive, unrelated terms were selected to be considered in the keywords to exclude. The process is iterated until each sample has a maximum number of false positives equal to two. The screening process to identify false positive is conducted manually by checking abstracts, titles, and keywords of papers. In case of unclear papers, the full-texts were consulted when available.

Step 4. The terms evaluated as non-coherent with ML and DL for safety purposes are added to the list of keywords to exclude. Simultaneously, new relevant terms were added as keywords to be included. However, it is worth mentioning that most of the keywords to include were found through the first step. In this step, terms to exclude are predominant. This step is required to minimize the occurrence of false positives.

Step 5. When all the random samples contain two or less false positives, the iteration is stopped. During this phase, unrelated WoS categories are studied to identify additional false positives and refine the analysis. It is important to conduct this step at the end to avoid removing apparent false positives which are inserted into unrelated categories. For instance, the category “mineralogy” could apparently be unrelated to the considered topic. However, it could still include risk or reliability analyses related to mine or mining equipment.

Step 6. The relevant pieces of information related to the remaining papers are extracted. The required information includes, but is not limited to, authors, year of publication, journal, title, abstract, keywords, and reference list.

It is worth mentioning that most of the analyses are performed through VOSviewer (Van Eck and Waltman, 2010), which is a renowned software for constructing bibliometric maps through a method called Visualization of Similarities (VoS). Some of the principles and maths behind VOSviewer could be found in (Haghani et al., 2021; Van Eck and Waltman, 2010; Waltman et al., 2010).

3. Findings

The application methodology presented in Section 2 resulted in 18,509 documents (date of search: February 2023), among which 12,767 are tagged as “article” (i.e., almost 70 %), while 5571 are regarded as “proceeding paper”. Among the extracted documents, about 30 % is published as All Open Access (5691 documents). The former papers are analysed considering the following categories:

- General statistics (Section 3.1).
- Clustering of literature (Section 3.2).
- Clustering of literature based on bibliographic coupling (Section 3.3).
- Authorship and co-authorship analysis (Section 3.4).
- Temporal evolution (Section 3.5).
- Influential studies (Section 3.6).

3.1. General statistics

The documents arising from the search belong to several WoS categories, among which over 30 are characterised by at least 200 documents. “Engineering Electrical Electronic” (number of publications equal to 5453) is by far the most common WoS category, followed by “Instruments Instrumentation” (number of publications equal to 2599) and “Computer Science Artificial Intelligence” (number of publications equal to 2593). Based on Fig. 3, it is possible to state that a particular interest is devoted to automation and control systems, energy fuels, telecommunications, material science, and civil engineering.

Another bibliometric perspective allows to identify the most relevant journals in the field of ML and DL application for safety purposes. The journal with the highest number of publications is IEEE Access, which presents 594 documents. It is followed by Sensors and Measurements characterised by 546 and 394 documents, respectively. 306 and 301 documents are published in Mechanical Systems and Signal Processing and IEEE Transactions on Instrumentation and Measurement, respectively. The preceding figures demonstrate the prevalent utilization of ML and DL in processing sensor-acquired or measured data to classify the condition of an asset or predict a forthcoming state. Furthermore, high interest is devoted to both the energy and the civil sector, denoted by the 231 and 207 documents of Energies and Structural Health Monitoring an International Journal, respectively. There are also other relevant journals which deal with broader topics that are worth mentioning such as Applied Sciences (341 documents) and Reliability Engineering and System Safety (196 documents). A list of the ten most relevant journals is summarised in Fig. 4.

Another interesting analysis regards the number of documents published in the last ten years (see Fig. 5). According to Fig. 5, the
application of ML and DL in safety, risk, reliability, resilience, and maintenance analysis has progressively increased, and it is still increasing. The documents published in 2012 were only 336, while 3967 documents were published in 2022. The advances of technologies in different field of engineering processes (e.g., maintenance planning for the turbofan of airplanes), required sound methods to handle the complexity of the process. ML and DL are capable to deal with the former complexity, thus, they progressively received more attention. The former fact is also related to the abilities of ML and DL since they could provide proper performance and accuracy. Moreover, the advances in sensors and digital technologies and the increase of computational power and efficiency of computers could have facilitated the widespread adoption of ML and DL techniques. In this context, it is worth mentioning that between 2012 and 2022, 16,385 documents were published, which accounts for about 88% of the total documents retrieved through the search.

Finally, it could be interesting to study the most prolific countries and the collaborations among them. Through VOSviewer, by imposing a minimum number of documents for a given country equal to 15, Fig. 6 is obtained. Specifically, 70 countries produced at least 15 documents. In Fig. 6, the colour represents the average publication year, while the size of each node and the links identify the number of documents and the co-authorship of each country. People’s Republic of China (PRC) and United States of America (USA) are the leading countries regarding number of documents, number of citations, and links. Considering the number of documents, other than PRC and USA, other countries produced over 300 documents: India, England, South Korea, Canada, Italy, Germany, Iran, France, Australia, Spain, Taiwan, Brasil, and Japan. The former countries can be considered as the most prolific ones in ML and DL for safety purposes.

3.2. Clustering of literature based on keywords’ co-occurrence

To cluster the collected documents, a semantic analysis of the keywords is performed through the VoS methodology as explained by Van Eck and Waltman (2007). VoS allows to aggregate the documents based on their similarities regarding different factors. First, the co-occurrence of keywords (i.e., the number of times two keywords are mentioned together) is investigated through VOSviewer software developed by Van Eck and Waltman (2010). The former software allows to conduct a VoS, and, therefore, visualise the objects in a network where each node represents an object and the distances among different objects denote their similarities (Haghi et al., 2022).

The process is carried out after the specification of a thesaurus, which excluded undesired terms and aggregate identical terms written differently (e.g., “remaining useful life” and “remaining useful lifetime”). Only terms with at least 30 occurrences were considered. The application of VoS to the keywords’ co-occurrence results in six clusters which are summarised in Fig. 7 and Table 1. In Fig. 7, the average number of citations and the average publication year are also shown. The size of each node represents the number of occurrences of the associated term. Table 1 lists the most frequent terms which are used to assign a name to the cluster, the selected prominent youngest terms (i.e., the ones with the highest average publication year), and the selected most cited on average terms (i.e., the ones with the highest number of citations). The six clusters are identified: (I) wind turbine, power system, cyber-physical systems, and maintenance, (II) Structural and civil engineering, (III) Rotating machinery and signal processing, (IV) Clustering and data reduction, (V) Battery and aero engine prognostics, and (VI) Rotating machinery, deep learning, and data manipulation.

First, it is worth mentioning that each term could belong to only one cluster, but it could be linked to all the others. This peculiar feature allows to study the relationships among different clusters. For instance, even though “neural network” belongs to the second cluster, it is linked to all the others denoting an adoption of neural networks in all the clusters.

The name of each cluster is based on the following considerations. Cluster I deal with mixed topics since it contains terms such as “wind turbine”, “cyber physical system”, “oil”, and “power system”. However, it is possible to state that it is strongly related to maintenance and related activities. It includes the terms “maintenance”, “preventive maintenance”, “condition-based maintenance”, “predictive maintenance”, and “maintenance engineering”. Cluster II is mainly related to structural and civil engineering due to the presence of “structural health monitoring”, “bridge”, “structural damage detection”, “beam”, and “structural reliability”. Cluster III contains both terms related to rotating equipment such as “rolling bearing”, “induction motor”, “rolling element bearing”, and “gearbox” and terms related to signal processing such as “empirical mode decomposition”, “wavelet transform”, “ensemble empirical mode decomposition”, and “discrete wavelet transform”. The term “clustering”
is included in Cluster I, but it presents links with the other cluster such as Cluster IV, which is identified as “Clustering and data reduction”. It presents terms such as “gaussian mixture model”, “k-nearest neighbours”, “fuzzy clustering”, and “fuzzy c-means”. It includes typical algorithms of data reduction such as “principal component analysis”, “independent component analysis”, “kernel component analysis”, and “linear discriminant analysis”. Cluster V is predominantly linked to the prognostics of batteries due to the presence of terms such as “remaining useful life”, “prognostics and health management”, “lithium-ion battery”, “battery capacity”, “health estimation”, “state of health”, and “prognostics”. Finally, Cluster VI contains terms such as “bearing”, “rotating machine”, and “planetary gearbox” which are strongly related to rotating equipment. It is possible to find “deep neural network”, “generative adversarial network”, “autoencoder”, “feature fusion”, and “feature learning”.

From a general perspective, the most common ML algorithm is the “neural network” since this term is the most frequently used. It is followed by “support vector machine”. Considering DL, the most occurred term is “convolutional neural network”, which could be considered as the most popular DL algorithm. The terms, “fault diagnosis”, “fault detection”, and “feature extraction” co-occurs with terms from all the other clusters, denoting their diffusion in all the identified divisions. Both Cluster III and Cluster VI are related to rotating equipment, as they have a common border and many co-occurred terms. For instance, “empirical mode decomposition” is linked to terms belonging to the two former clusters. A similar consideration can be made for the term “variational mode decomposition”. “Acoustic” is a term associated with Cluster II, and it co-occurs with “vibration”, which is a term of Cluster III. It co-occurs with “rolling bearing”, while “vibration” has a link to
"structural health monitoring". Both acoustic and vibration signals are used for the analysis of structural or rolling elements even though they present significant differences. The term “image” and “image processing” belonging to Cluster II co-occur with the term “convolutional neural network” of Cluster VI, denoting the popularity of CNN within the context of image analysis. Furthermore, “image processing” co-occurs with “fault diagnosis” and “classification” of Cluster III, possibly highlighting the adoption of image processing techniques for diagnosis of fault and their classification (e.g., different failure modes).

A more detailed analysis of each cluster is provided below:

**Cluster I.** Cluster I presents the term “wind turbine”, which is related to the terms “scada” and “scada data”, possibly denoting the employment of data acquired through SCADA for the diagnosis and fault detection of wind turbines. Indeed, “scada” data is also linked to the terms “diagnosis” and “fault detection” of Cluster IV. This finding is aligned with previous works that mentioned that data coming from SCADA are effective to identify wind turbine performance at low costs (Leahy et al., 2016; Kim et al., 2011). The terms “cyber-physical system” and
“cybersecurity” co-occur with the term “anomaly detection”. In this cluster, the terms “railway”, “earthquake”, “landslide”, and “landslide susceptibility” also appear, but they are characterised by a lower number of occurrences, and they are quite isolated compared to the other terms. Considering the most prominent youngest terms of Cluster I (see Table 1), it is possible to state that the focus is shifting towards resilience analysis. Moreover, “remote sensing” and “digital twin” seem to be recently integrated with the safety framework. Considering ML and DL tools, “extreme gradient boosting”, “federated learning”, and “deep reinforcement learning” appear to be popular recent approaches.

Cluster II. In this cluster, the terms “concrete” and “beam” have links with “damage detection” and “crack detection”. Moreover, “beam” has a link with “natural frequency”, while “concrete” is characterised by co-occurrence with the term “acoustic”. The term “bridge” also co-occurs with “damage detection”, but also “modal identification”. In this cluster, it is interesting to see a co-occurrence of the term “computer vision” with the following terms: “bridge”, “crack detection”, “inspection”, and “damage detection”. Computer vision is employed for inspection related to the detection of cracks and damages. It emerged to be as a prominent recent term, denoting a modern new solution to evaluate the health of a structure. In this context, also the term “image” and “vision” co-occur with “damage detection” and “crack detection”, underlying how the images and their related processing could be a common solution to assess cracks and damages.

Cluster III. This cluster mainly covers the fault diagnosis of rotating equipment, denoting the importance of signal processing techniques such as “wavelet transform” and “empirical mode decomposition”. It emerged that a prominent young technique is “variational mode decomposition”, which is employed in fault diagnosis of rotating equipment. It is also worth mentioning that among the rotating machinery falling in this group, the “reciprocating compressor” is the one which is less frequently considered since it has a lower number of occurrences compared to “rolling bearing”, “induction motor”, or “centrifugal pump”.

Cluster IV. Other than dimensionality reduction and clustering techniques, this cluster includes different fields such as “turbine”, “high-speed train”, “engine”, “chemical process”, and “nuclear power plant”. The analysis depicted “variational autoencoder” and “denoising autoencoder” as prominent youngest terms, which overlaps with Cluster VI. Another interesting finding is that “high-speed train” also resulted as a prominent youngest term, possibly denoting current and future area of development.

Cluster V. Among the most frequent terms, “remaining useful life” and “long short-term memory neural network” emerges, denoting the popularity of Long Short-Term Memory NN (LSTMNN) for estimating the RUL of batteries. The term “electric vehicle” falls in this cluster as well due to its relationships with batteries. This cluster also reports the terms “aero engine” and “turbofan engine”, which have links with “prognostics and health management” and “remaining useful life”, denoting the popularity of predictive models for aero engines. In this cluster the term “mathematical model” appears as one of the most prominent. It co-occurs with the term “degradation”, possibly identifying the recent construction of degradation mathematical models. Furthermore, “mathematical model” co-occurs with “convolutional neural network”, “neural network”, and “support vector machine”, which could mean that mathematical models have been recently integrated with ML and DL algorithms.

Cluster VI. In this cluster, the terms “small sample” and “few-shot learning” appears, possibly highlighting the need to diagnose failure under sparse data. Furthermore, “few-shot learning” emerged as a prominent youngest term, which could denote the recent need to train models with limited data. Another prominent youngest term is “graph neural network”, which presents co-occurrence with “anomaly detection” (Cluster II). The former aspect could represent a recent trend in the adoption of a “graph neural network” to detect anomaly conditions. “Meta learning” is another interesting prominent youngest term, which shows a connection with “fault diagnosis”. Thus, it is possible that recently “meta learning” could have been exploited for diagnosing

**Table 1**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Topic</th>
<th>Selected most frequent terms</th>
<th>Selected prominent youngest terms</th>
<th>Selected most cited on average</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (red)</td>
<td>Wind turbine, power systems, cyber physical systems, and maintenance</td>
<td>System, anomaly detection, optimisation, artificial intelligence, predictive maintenance, reliability, wind turbine, maintenance, random forest, decision tree, risk assessment, condition-based maintenance</td>
<td>Explainable artificial intelligence, digital twin, federated learning, deep reinforcement learning, extreme gradient boosting, maintenance engineering, internet, remote sensing, resilience</td>
<td>Implementation, fuzzy, fuzzy logic, frequency ration, Supervisory control and data acquisition (SCADA) data, hazard</td>
</tr>
<tr>
<td>II (green)</td>
<td>Structural and civil engineering</td>
<td>Neural network, structural health monitoring, identification, damage detection, genetic algorithm, performance, sensor, acoustic, pattern recognition, crack detection, bridge</td>
<td>Semantic segmentation, deformation, mechanism, surrogate model, reconstruction, computer vision, unmanned aerial vehicle</td>
<td>Response surface, vision, health monitoring, structural reliability, slope stability, response surface method, image segmentation, modal identification, cable-stayed bridge</td>
</tr>
<tr>
<td>III (blue)</td>
<td>Rotating machinery and signal processing</td>
<td>Fault diagnosis, support vector machine, classification, machine, vibration, rolling bearing, condition monitoring, empirical mode decomposition, wavelet transform, induction motor</td>
<td>Circuit faults, mode decomposition, variational mode decomposition, fuzzy entropy</td>
<td>Self-organizing map, sensor fusion, Fisher discriminant analysis, Bayesian inference, independent component analysis, Bayesian classifier, statistical process control</td>
</tr>
<tr>
<td>IV (yellow)</td>
<td>Clustering and data reduction</td>
<td>Diagnosis, fault detection, principal component analysis, selection, diagnostics, gaussian mixture model, k-nearest neighbour, dimensionality reduction</td>
<td>Variational autoencoder, denoising autoencoder, high-speed train, data driven, sparse autoencoder</td>
<td>Relevance vector machine, pack, combination, state of charge, health estimation, battery management system, electric vehicle</td>
</tr>
<tr>
<td>V (violet)</td>
<td>Battery and aero engine prognostics</td>
<td>Prognostics, remaining useful life, long term-memory neural network, remaining useful life estimation, lithium-ion battery, regression, recurrent neural network, battery, aero engine</td>
<td>Predictive model, computational modeling, mathematical model, health prognostics, online state</td>
<td>Intelligent fault diagnosis, feature learning, stacked denoising autoencoder, sparse filtering, intelligent diagnosis, planetary gearbox, deep convolutional network</td>
</tr>
</tbody>
</table>
| VI (light blue) | Rotating machinery, deep learning, and data manipulation | Convolutional neural network, feature extraction, rotating machine, bearing, transfer learning, autoencoder | Graph neural network, meta learning, adaptation models, data models, employee welfare, task analysis, informatics, attention mechanism, few-shot learning | }
faults. Finally, “employee welfare” and “task analysis” were depicted as prominent youngest terms, possibly representing an increased attention towards the abilities, skills, and welfare of operators during fault diagnosis activities mainly related to rotating equipment.

3.3. Clustering of literature based on bibliographic coupling

Through VOSviewer, an additional cluster analysis is performed by considering the reference lists of different documents. The former analysis allows to evaluate the similarity among different documents based on the references they cite. Then, the documents are aggregated based on the authors to identify similarities among different authors. In this study, authors with at least twenty documents were considered. The results arising from the analysis are shown in Fig. 8 and Table 2. In Fig. 8, each node represents a different author, and its size denotes the number of documents the author published. A link connects given authors to others with a similar reference list. Considering Table 2, the topic associated with each cluster is chosen based on the Google Scholar pages of the top cited researchers, followed by a screened of the Scopus page in case the Google Scholar page was not present or did not provide any description.

The bibliographic coupling resulted in five clusters. Each author can be associated with a single cluster, but he/she could be characterized by links with other clusters. This is very common for authors belonging to Cluster 1, Cluster 2, Cluster 3, and Cluster 4, which present wide overlapping areas. The previous scenario possibly denotes a multidisciplinary nature of the field and leads to difficulties in determining specific topics for each cluster. Indeed, the former clusters share similar topics which are not reported in Table 2 for identifying peculiar prominent topics. As a matter of fact, “rolling bearing”, “rotating machinery”, and “failure analysis” are key terms that are listed for most of the most cited authors of Cluster 1, Cluster 2, Cluster 3, and Cluster 4. Based on the former consideration, it is possible to state that many scholars have been focusing on fault detection and diagnosis of rotating machineries. This trend could be related to different factors such as the importance of rotating equipment and their complex fault behaviours (e.g., different failure modes), which has led to the adoption of ML and DL algorithms. Rotating devices such as bearings are used for several applications and systems. Finally, bearings present multiple similarities despite their eventual different sizes, applications, and usages.

Considering the single clusters, Cluster 1 contains authors that deal with RUL of batteries and critical infrastructures (e.g., nuclear plants), while Cluster 2 includes authors that treat fatigue and fracture, damage modelling, machine health monitoring and signal processing. Authors that study random vibration, model predictive control, smart materials and structures, and dynamics and vibration are assigned to Cluster 3, while Cluster 4 is associated with the following topics: control systems and artificial intelligence in condition monitoring. Cluster 5 is the most isolated one compared to the remaining networks; thus, authors belonging to Cluster 5 have quite different reference lists compared to the rest. Only two authors are associated with this cluster and the common topics are SHM and damage detection, wind turbine, and wind farm. It is also worth mentioning that Cluster 1 seems to present a sub-cluster of authors that is more distant from the remaining clusters.

3.4. Authorship and co-authorship analysis

To investigate the pattern of authorship and co-authorship, it is possible to build a collaboration network through VOSviewer. Filtering out the authors with less than 15 documents and the authors that show limited connections to the remaining network (default of VOSviewer), Fig. 9 is obtained. In Fig. 9 each node represents a distinct author, while each link identifies a collaboration between two authors. The node size is related to the number of documents realized by the corresponding

Fig. 8. Bibliographic coupling of authors with more than 20 papers.
To explore the evolution of terms used as keywords, the keywords’ co-occurrence analysis was conducted through VOSviewer for five different periods, following the procedure described in Patriarca (2020). Specifically, the analysis is carried out five different times (one for each period), and each time only the papers published during that period are considered. Furthermore, two different thresholds are considered for the minimum number of occurrences of a given term: 15 and 30. The identified period are as follows: 2003–2006 (Period I), 2007–2010 (Period II), 2011–2014 (Period III), 2015–2018 (Period IV), and 2019–2022 (Period V).

### 3.5 Temporal evolution

To explore the evolution of terms used as keywords, the keywords’ co-occurrence analysis was conducted through VOSviewer for five different periods, following the procedure described in Patriarca (2020). Specifically, the analysis is carried out five different times (one for each period), and each time only the papers published during that period are considered. Furthermore, two different thresholds are considered for the minimum number of occurrences of a given term: 15 and 30. The identified period are as follows: 2003–2006 (Period I), 2007–2010 (Period II), 2011–2014 (Period III), 2015–2018 (Period IV), and 2019–2022 (Period V).

#### 3.5.1 Period I: 2003–2006

Considering Period I, Fig. 11 shows the clusters obtained when considering 15 and 30 number of occurrences as thresholds of exclusion. Based on Fig. 11, it is possible to state that “neural network” and “support vector machine” were the most popular ML algorithms. The concepts of “fault diagnosis” and “diagnosis” were already considered. It
is worth noting that “damage detection” and “structural health monitoring” were used commonly during this period. Besides, “rotating machinery” and the related “condition monitoring” were already treated. Finally, “principal component analysis” and “wavelet transform” emerges as the most popular approaches for data pre-processing (e.g., data reduction).

3.5.2. Period II: 2007–2010

As depicted in Fig. 12, the number of relevant terms associated with Period II increases. This trend is related to the increase in popularity of ML applications for safety purposes.

Once again, “neural network” and “support vector machine” are highlighted as the most popular ML algorithms. Considering ML algorithms, this period sees the increase in popularity of “decision tree” and “clustering”. The attention of this period is also directed towards “feature selection”, “feature extraction”, “fault classification”, and “wavelet transform”. Another interesting finding is the appearance of “rolling bearing” and “induction motor”, possibly denoting a higher focus on the former devices. The term “vibration” appears among the most relevant ones, possibly highlighting the popularity of studying vibrations for fault detection and diagnosis. Finally, the terms “reliability” and “structural reliability analysis” underlines the focus also on reliability analysis other than fault diagnosis and damage detection.

3.5.3. Period III: 2011–2014

During Period III, the number of relevant terms increases once again (see Fig. 13). New algorithms and new fields of application are characterised by an increase in popularity.

Period III sees the appearance of several algorithms such as “support vector regression”, “gaussian mixture model”, “fuzzy clustering”, and “fuzzy neural network”. The last two terms, along with the term “fuzzy logic”, possibly denote a new trend, which could be identified as the integration of ML algorithms and fuzzy logic. Moreover, “empirical mode decomposition” as a data pre-processing technique has gained attention during this period. An additional finding is the appearance of “maintenance” and “optimisation” in the same cluster, possibly highlighting the employment of ML algorithms in maintenance planning and optimisation. The term “optimisation” could also refer to a higher attention towards the optimisation of ML algorithm and their related computational time. For the first time, “prognostic” and “remaining useful life” emerge as relevant terms. This could relate to the increased importance of prediction capabilities. Finally, the term “bridge” and “acoustic” are depicted as more popular compared to the previous periods. The former could mean that bridge safety was also studied through ML algorithms during Period III, while the latter could highlight the higher focus dedicated to acoustic signals in safety frameworks. In the previous periods, only vibration signals appear to be commonly considered within the proposed frameworks.

3.5.4. Period IV: 2015–2018

Period IV sees a wider adoption of DL algorithms, along with the integration of new approaches in ML and DL applications for safety purposes.

The increased attention towards DL is denoted by the appearance of relevant DL algorithms (see Fig. 14) such as “convolutional neural network”, “long short-term memory neural network”, “recurrent neural network”, and “deep neural network”. During Period IV, it is possible to state that there was a rise in the adoption of DL algorithms. The term “transfer learning” also experienced an increase in the number of occurrences. Furthermore, “condition-based maintenance”, “predictive
maintenance”, “safety”, and “risk assessment” are identified as relevant terms, possibly denoting different purposes of the ML and DL employment. In this period, terms such as “Bayesian network”, “hidden Markov model”, and “simulation” experienced an increase in popularity. This trend could highlight the need for integrating ML and DL algorithms with other approaches and tools. Besides, the following fields are depicted as more popular compared to the previous periods: “turbine”, “batteries”, “cyber-physical system”. Finally, the term “image processing” appears for the first time, possibly denoting a new trend in evaluating failures. Machine vision could improve the reliability of defect detection (Ren et al., 2022). In this context, DL is among the most popular approaches for image processing.

3.5.5. Period V: 2019–2022

Fig. 10 illustrates the results of the keywords’ co-occurrence analysis of Period V. The increase in the number of relevant terms is an indicator of the widespread adoption of both ML and DL during the past years, along with possibly denoting its multidisciplinary nature and complexity.

Regarding Period V, new fields of application are considered: “navigation”, “c-maps”, “aero-engine”, “aircraft”, “pipeline”, “corrosion”, “railway”, “road”, “slopes stability”, “tool”, “unmanned aerial vehicle”, and “cybersecurity”. This trend could represent the increase in popularity of ML and DL outside the common fields of the previous periods (e.g., rotating machines, bearings, and bridges). Furthermore, this period experienced the introduction of the concept of “resilience” and “vulnerability”, possibly denoting future trends of application. Another interesting finding is the appearance of “variational mode decomposition” for data pre-processing. Moreover, “transfer learning” and “deep transfer learning” have become more popular during Period V. Similarly, “meta-learning”, “deep reinforcement learning”, “few-shot learning”, “adversarial learning”, “attention mechanism” appear in this period. Finally, the terms related to image processing occur more times during Period V (e.g., “image classification”, “image segmentation”, “machine vision”). Nowadays, it is possible to obtain high-quality images with different devices (e.g., unmanned aerial vehicles), reducing the difficulties in reaching given places.

Fig. 10. Network of co-authorship for authors with at least 15 documents independently from their links with other authors that produced at least 15 documents.
3.6. Influential studies

A final analysis is performed considering the most cited articles. Specifically, for this step of analysis, the documents with over 300 citations are considered. Among the identified documents, 57 are characterised by several citations higher than 300. The 57 documents are clustered based on the topics reported in their title, abstract, and keywords, as shown in Table 3.

As depicted in Table 3, the health monitoring and the fault detection of rotating machinery such as bearings is among the most predominant fields (Cluster A). In this context, NN is the most common ML algorithm (Li et al., 2000; Ali et al., 2015; Bin et al., 2012; Samanta et al., 2003; Prieto et al., 2012), while considering DL algorithms, CNN is one of the most employed (Wen et al., 2017; Ince et al., 2016; Janssens, 2016; Zhang et al., 2018; Zhang et al., 2017; Shao et al., 2018; Jiang et al., 2018; Xia et al., 2017; Jia et al., 2018), eventually in a transfer learning framework (Guo et al., 2018; Yang et al., 2019). Sparse autoencoder also emerged as a popular tool adopted in the former context (Wen et al., 2017; Chen and Li, 2017; Jia et al., 2018). Documents adopting CNN are usually more recent compared to the documents where neural networks are exploited. In the aforementioned context, DL is also used for automatic feature extraction. The former findings are aligned with the ones described in the previous section. It is worth mentioning that Cluster A includes also studies related to rotating machines of wind turbines such as the gearbox (Jiang et al., 2018) or motors of autonomous aerial vehicles (Lu et al., 2017), or bearings of locomotives (Yang et al., 2019). Furthermore, in (Wen et al., 2017) an axial piston is also considered.

Considering Cluster B, different review studies were developed during the past years. Most of the former documents treat the adoption of ML or DL in health management or failure diagnosis (Isermann, 2005; Liu et al., 2018; Bellini et al., 2008; Lei et al., 2020; Lei et al., 2020; Khan and Yairi, 2018). Among the former works, there are studies also related to rotating machineries such as induction motors (Liu et al., 2018; Bellini et al., 2008), denoting the popularity of this field.

Cluster C includes documents that treat the RUL prediction of rotating machines and tools. In this context, NN (Huang et al., 2007; Gebrael et al., 2004) (i.e., ML) and RNN (Guo et al., 2017; Zhao et al., 2017) (i.e., DL) are the most popular algorithms, while relevance vector machine is considered in (Wang et al., 2018). Once again, the documents adopting DL are more recent compared to the ones considering ML.

In Cluster D, the works related to the health monitoring and the prediction of RUL for batteries and aeroengine are included. LSTMNN appears a common solution also for batteries (Zhang et al., 2018) and aircraft engine (Wu et al., 2018).

Finally, Cluster E contains work on SHM and damage detection (Abdeljaber et al., 2017; Cha et al., 2018; Lin et al., 2017), and landslide (Stumpf and Kerle, 2011; Gomez and Kavzoglu, 2005). Thus, it could be related to civil and structural engineering. In this cluster, CNN emerged as a popular tool once again (Cha et al., 2018; Abdeljaber et al., 2017; Lin et al., 2017), denoting its popularity. This trend could be related to the advantages of CNNs and their ability to autonomously extract relevant features.

4. Discussion

The findings of the present study could be related to three major areas: i) fields of application, ii) popular algorithms and tools, and iii) past, present, and possible future trends. The former areas are discussed in the following sections.

4.1. Fields of application

This study highlights the flexibility of ML and DL through the identified different fields of application. In this context, it is worth noting that the detected clusters usually present wide overlapping areas, possibly denoting similarities. Among the fields of application, the failure diagnosis, detection, and classification of rotating machineries (e.g., rolling bearings, motors) emerged as the most popular. The former aspect is underlined by the keywords’ co-occurrence analysis, the bibliographic coupling of authors, and the most influential studies with regard to the number of citations. This trend could be related to the diffusion of rotating machines, such as bearings for a wide variety of equipment (Liu and Fan, 2022). Furthermore, the abilities of ML and DL are well aligned with the complex nature of rotating machineries. Indeed, AI-based tools usually show a higher accuracy compared to other condition monitoring approaches (Manikandan and Duraiselvam, 2021).

Another common field of application is the structural and civil engineering, especially in SHM and damage detection. Despite the different nature, there are some similarities between this field and the previous one. Indeed, it appears that "vibration" and "acoustic" are exploited for the health monitoring of both the fields.

Finally, another popular field is related to the health monitoring and the prognostic of batteries. Indeed, batteries have a pivotal role in many aspects of daily life (Tang and Yuan, 2022) and serve numerous technologies, such as electric vehicles (Ma et al., 2023). Furthermore, they are an important enabler of the transition from fossil fuels to green energy (Nagulapati et al., 2021).

ML and DL are also used in other fields such as maritime (Zhang et al., 2022; Abarai et al., 2021) nuclear power plant (Worrell et al., 2019), and aeroengine (Zhuang et al., 2023), however, their degree of employment appear to be still lower compared to the former fields. Thus, it could be worth examining more in detail other fields of application thanks to the generalisability of ML and DL algorithms.
Fig. 12. Keywords’ co-occurrence of terms published between 2007 and 2010: (a) 15 as threshold and (b) 30 as threshold.
Fig. 13. Keywords’ co-occurrence of terms published between 2011 and 2014: (a) 15 as threshold and (b) 30 as threshold.
Fig. 14. Keywords’ co-occurrence of terms published between 2015 and 2018: (a) 15 as threshold and (b) 30 as threshold.
Fig. 15. Keywords’ co-occurrence of terms published between 2019 and 2022: (a) 15 as threshold and (b) 30 as threshold.
4.2. Popular algorithms and tools

Considering the ML algorithms, the most popular one is the NN as depicted by both the keywords’ co-occurrence analysis and the analysis of the most influential studies. This factor could be related to the ability of NNs to extract complex patterns for different applications (Fink et al., 2014). On the other hand, the most popular DL algorithm is the CNN. CNN has the ability of big data processing (Zhu et al., 2017) and it is well established in the context of computer vision tasks (Yamashita et al., 2018), such as image or object recognition. Another popular DL algorithm is the RNN, which is extensively employed for the RUL prediction of different devices such as batteries or aeroengines. Indeed, RNN can store and process complex signals, along with predicting future sequences (Salehinejad et al., 1801). RNN are also diffused in different variants, such as Gated Recurrent Unit (GRU) (Chen et al., 2019) and LSTM (Shi and Chehade, 2021).

Since ML and DL usually require processing a signal, they are often integrated with other tools that could be adopted for pre-processing the signal (i.e., feature extraction, feature selection, and noise removal). Different approaches emerged as popular for the data pre-processing phase. Among them, Principal Component Analysis (PCA) (Gu et al., 2018), Empirical Mode Decomposition (EMD) or Ensemble EMD (EEMD) (Zhang et al., 2020); Variational Mode Decomposition (VMD) (Habcheuche et al., 2021), and Wavelet Transform (WT) (Kankar et al., 2011) are worth mentioning.

Finally, it appears that other tools are integrated with ML and DL such as Bayesian inference or Hidden Markov Model. For instance, ML and DL algorithms require to tune hyperparameters, which is an optimisation problem. In this context Bayesian optimisation is usually more efficient compared to other optimisation algorithms (Wu et al., 2019). On the other hand, Hidden Markov Model could be integrated with RNN for prediction purposes (Zhuang et al., 2023).

4.3. Past, present, and possible future trends

Based on the keywords’ co-occurrence analysis presented in Sections 3.2 and 3.5, it is possible to state that the present field of study has seen a deep evolution during the past two decades. The former aspect is underlined by the progressive increase in the number of documents published each year.

During the first years of the past decades, only ML algorithms were adopted. Specifically, NNs and SVM were mainly employed in diagnosis of rotating equipment, damage detection, and SHM. This trend possibly identifies the adoption of ML for classification purposes. In this period, one of techniques that appear to be commonly integrated with ML is PCA. PCA is indeed a common dimensionality reduction algorithm, which identifies directions named principal components characterised by maximum variation of the data (Ringnér, 2008). Moreover, it is worth noting that some studies consider PCA as ML thanks to its unsupervised learning capabilities (Peng et al., 2020).

Next other approaches and algorithms such as DTs and clustering were considered in different frameworks. The former fact possibly denotes the diffusion of unsupervised ML. Moreover, other tools such as WTs were included in ML-based frameworks. WTs could have been preferred to other approaches such as Fourier Transform (FT) since it is able to deal with non-stationary time series (Slezia et al., 2015). Since “reliability” is also a term that appears to be more relevant during this period, it is possible to state that more importance was given to reliability analysis other than failure diagnosis and damage detection.

The period between 2011 and 2014 has seen ML reaching its maturity since new algorithms (e.g., gaussian mixture model) and new application fields (e.g., bridges) emerged as relevant. Moreover, the appearance of prognostic and RUL possibly denotes an increase in attention towards regression problems other than the already common classification ones. Stated differently, the studies also concentrated on predicting future conditions to facilitate planning of maintenance activities. The aforementioned period is also characterised by an increase in popularity of other pre-processing techniques such as EMD, possibly to its ability to deal with both non-stationary and non-linear signals (Leoni et al., 2023). However, to experience an increase in popularity of DL algorithms, it is required to consider the next four-year period (2015–2018). This finding is aligned with other studies which state how DL has become popular after 2014 (Li et al., 2020) or how CNN has seen a recent burst of citations (Su et al., 2021). This trend could be related to both the advances in computer capabilities and ML (Hoang and Kang, 2019). The popularity of DL algorithms and their widespread adoption could be related to their advantages compared to ML algorithms. For instance, DL approaches could automatically perform feature extraction and feature selection (Zhang et al., 2020). Moreover, is deemed to be a fundamental factor to achieve human or super-human AI systems
Furthermore, DL is very suited to process images, which emerged as a popular topic between 2015 and 2018. Thus, the focus during this period was also directed towards the adoption of images to perform different safety tasks such as fault diagnosis or damage detection. Another important finding is that both risk and safety are treated more in detail since they appear as more relevant during this period. Accordingly, this trend could denote that researchers also focused on risk and safety management and not only reliability analysis.

Finally, during the past four years (2018–2022), ML and DL have seen an increase in the number of application fields, other than an increase in the number of documents related to already treated field such as batteries. This period is also characterised by an increase in popularity of Variational Mode Decomposition (VMD), possibly due to its ability to deal with non-stationary and non-linear signals, along with being more robust than EMD under noise and sampling errors (Nazari and Sakhaei, 2020). Moreover, both transfer learning and deep transfer learning appear as more popular during the last four years compared to the previous period. Transfer learning, by considering the transfer of knowledge across domains, can solve a common problem of ML frameworks, which is the presence of many unlabelled data (Zhuang et al., 2020). In other word, transfer learning aims to enhance learners by transferring information or data from a domain to a related one (Weiss et al., 2016). This trend could also be aligned with the identification of few-shot learning as a prominent youngest term. Few-shot learning has been recently employed to tackle small datasets (Wang et al., 2020). Indeed, few-shot learning uses prior knowledge with few available information to generalize new tasks (e.g., image classification) (Wang et al., 2020).

Similarly, generative adversarial learning and GAN appeared during the past four years. In fact, GAN could also be used under data scarcity, since they could generate virtual data based on the real observations (Yoo et al., 2021). Another prominent youngest term is federated learning. Indeed, its introduction is very recent (i.e., 2016) and it could tackle the absence of labelled data, while preserving data privacy (Li et al., 2020). In fact, data privacy is of prominent importance nowadays, confirmed also by the appearance of cybersecurity between 2018 and 2022. As a matter of fact, federated learning allows to run collaborative ML across multiple users without sharing users’ data with the service provider (Banabilah et al., 2022). Deep reinforcement learning appears as the youngest terms as well. Indeed, deep reinforcement learning has recently become popular in different engineering fields to solve problems that were inaccessible due to high complexity such as non-linearity or high dimensionality (Garnier et al., 2021).

Besides, meta-learning emerged as a prominent youngest term. It can solve an algorithm selection problem (Khan et al., 2020), whose importance has increased due to the several algorithms that are available. Moreover, meta-learning is capable of learn new tasks by acquiring knowledge from known prior tasks (Ma et al., 2022). Finally, ML and DL were also extended to the very recent concepts of vulnerability and resilience analysis as depicted by the appearance of “vulnerability” and “resilience”.

Based on the former considerations, it is possible to state that, during the next years, studies could be focused on the expansion of not only DL algorithms but also of reinforcement, federated, meta, few-shot, and transfer learning. The focus seems to be shifted towards tackling the absence of data or labelled data and assuring the privacy of data. In this context, new application fields could emerge such as high-speed trains (detected as a youngest term). Finally, it is expected to consider more frequently images and computer vision for diagnosis and prediction purposes when adopting DL algorithms. Terms related to image processing are identified as quite recent. All the findings of this section are summarised in Fig. 16.

5. Conclusions

This paper presents a SBA about the adoption of ML and DL in the context of safety analysis, including risk, reliability, resilience, vulnerability, fault detection and diagnosis, and maintenance planning. The search conducted through WoS allowed to identify 18,509 documents dealing with several industrial and non-industrial fields (e.g., bridges). ML and DL are flexible tools that can be employed for very different applications.

From a methodological perspective, this paper proposes a new approach to conduct a bibliometric analysis by exploiting at first specific journals to identify useful keywords and subsequently expand the search to multidisciplinary journals as well. From a theoretical perspective, this paper investigates the intellectual structure of research on ML and DL applications for safety purposes. Specifically, the diffusion of ML and DL in different fields is studied, along with considering the popularity of different algorithms. This has been possible thanks to the inclusion of several fields, contrary to the most common choice of focusing on a single field of application. Topics and keywords are also investigated, allowing for a clustering of the documents and the authors. Finally, the temporal evolution of keywords’ co-occurrence is evaluated to identify past trends and potential future evolutions.
Among the main findings of the present work, the following ones are worth mentioning:

(1) The most predominant fields are rotating machinery, batteries, and SHM, followed by aeroengines and turbines.
(2) The most popular ML algorithm is NN, while the most popular DL algorithm is CNN.
(3) The authors’ bibliographic coupling and co-authorship network highlight the highly multidisciplinary nature of ML and DL since many relevant authors treat different fields.
(4) Health monitoring has become increasingly coupled with RUL prediction, with RNN being a particularly popular tool for this purpose.
(5) DL has reached a maturity level during the past four years thanks to the widespread adoption.
(6) The temporal trend shows that ML algorithms were integrated at first with feature extraction, feature extraction, and dimensionality reduction tools (e.g., PCA), while DL has been adopted in the past decade to automatically extract and select features.
(7) Recently, the focus is on the treatment of unlabelled or few data, while preserving data privacy due to the increase in popularity of few-shot learning, federated learning, deep reinforcement learning, generative adversarial learning, and meta-learning.
(8) At first only failure diagnosis and damage detection appear to be considered, during the years, also more general reliability, risk, and safety analyses were treated through ML and DL, while very recently ML and DL have been seen application also in resilience and vulnerability.

It is also worth mentioning that new fields of application have progressively been introduced. Thus, it could be worthy to continue following this path to exploit the potential and abilities of ML and DL for usually less considered fields. It is possible to expect a higher level of diffusion of federated learning, deep reinforcement learning, meta-learning, and few-shot learning, similar to what happened for transfer learning and DL. For the same consideration, it is possible to expect the introduction of new approaches similarly to the former ones. Finally, another interesting finding is the evolution of tools that are used with ML and DL. For instance, at the beginning, WT was very popular, then EMD has seen an increase in popularity, and last, VMD has experienced a wide diffusion. This trend could be related to the need to improve the data pre-processing techniques along with the DL and ML algorithms. Put simply, advances in ML and DL applications for safety purposes may require the development of complementary methods.

As with any other work, this study certainly presents some limitations. First, the identified documents are related to the adopted string, thus, excluding or including different terms could lead to slightly different results. Despite trying to minimise the number of false positives while maximising the coverage, false positives could still be present, while relevant works could still be excluded. It could be possible to progressively refine the keywords’ string. Furthermore, the analysis reflects the documents published during the past years, thus, it could be interesting to repeat it in the future. ML and DL are constantly evolving over time, and it could be tough to keep track of their advances. Finally, in the temporal analysis, the citation bursts of papers and keywords were not considered. It follows that it could be interesting to investigate them, along with the automated clustering arising from tools such as GiteSpace, similar to what previous studies did (Huang et al., 2020; Tavakoli et al., 2023).

CRediT authorship contribution statement

Leonardo Leoni: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Ahmad Bahoo-Toroody: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Mohammad Mahdi Abar: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Alessandra Cantini: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Farshad Bahoo-Toroody: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Filippo De Carlo: Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

(TS—(“deep learning” OR “deep-learning” OR “machine learning” OR “statistical learning” OR “supervised learning” OR “unsupervised learning” OR “semisupervised learning” OR “semi-supervised learning” OR “transfer learning” OR “cost-sensitive learning” OR “ensemble learning” OR “federated learning” OR “reinforcement learning” OR “clustering” OR “text mining” OR “rule-based learning” OR “association rule mining” OR “collaborative learning” OR “parameter optimization of classifiers” OR “parameter optimisation of classifiers” OR “hybrid learning” OR “hybrid-learning” OR “extreme learning machine” OR “supervised method” OR “unsupervised method” OR “supervised approach” OR “unsupervised approach” OR “representation learning” OR “feature learning” OR “mete-learning” OR “natural language processing”) OR AK—(“Artificial Neural Network” OR “Long-short term memory” OR “backpropagation neural network” OR “back propagation neural network” OR “feed forward neural network” OR “feedforward neural network” OR “recurrent neural network” OR “neural network” OR “Convolutional neural network” OR “convolutional network” OR “deep belief network” OR “generative adversarial network” OR “autoencoder” OR “support vector machine” OR “support vector regression” OR “relevant vector machine” OR “Random forest” OR “gradient boosting” OR “boosting tree ensemble” OR “K-nearest neighbour” OR “K-nearest neighbour” OR “deep intra-class adaptive network” OR “Naive Bayes” OR “classifier” OR “Decision Tree” OR “Gaussian mixture model” OR “Gaussian mixture regression” OR “Random process tree” OR “AdaBoost” OR “confusion matrix” OR “Fuzzy C-Means” OR “hierarchical cluster” OR “auto-regressive integrated moving average” OR “multilayer perceptron” OR “multi-layer perceptron” OR “K-Means” OR “spectral cluster” OR “extremely randomised tree” OR “extremely randomised tree” OR “K-medoid” OR “spatial cluster”)).

AND.

(TS—(“Hazard and Operability analysis” OR “Failure mode effects analysis” OR “Failure mode effects and criticality analysis” OR “Risk priority number” OR “Function resonance analysis method” OR “Fault tree” OR “Root cause analysis” OR “common cause failure” OR “Cognitive reliability and error analysis method” OR “human reliability analysis” OR “preliminary hazard analysis” OR “process hazard assessment” OR “health hazard predictor” OR “state of health estimation” OR “safety hazard predictor” OR “Event Tree” OR “Fault diagnosis” OR “Fault detection and identification” OR “Fault diagnosis” OR “Fault detection” OR “Fault identification” OR “Fault detection and diagnosis” OR “Fault predictor” OR “Fault classification” OR “fault type identification” OR “Fault assessment” OR “Fault mitigation” OR “Fault integrity” OR “Fault propagation” OR “Fault model” OR “Elimination of fault” OR “Avoidance of fault” OR “Tolerance of fault” OR “Failure

Safety Science 170 (2024) 106363
L. Leoni et al.
prognosis” OR “Failure detection and identification” OR “Failure diagnosis” OR “Failure detection” OR “Failure identification” OR “Failure detection and diagnosis” OR “Failure predict” OR “Failure classification” OR “Failure assessment” OR “Failure analysis” OR “Failure mode identification” OR “failure modes identification” OR “failure probability” OR “integrity and mitigation of failure” OR “Failure mitigation” OR “Failure integrity” OR “Analysis of failure mode” OR “failure model” OR “Elimination of failure” OR “Avoidance of failure” OR “Tolerance of failure” OR “Mean time between failure” OR “Mean time to failure” OR “Functional failure” OR “Unseen failure” OR “Failure mode” OR “Failure level” OR “Unseen fault” OR “failure event” OR “faulty scenario” OR “fault type” OR “failure probability” OR “failure frequency” OR “potential fault condition” OR “component failure” OR “sub-system failure” OR “failure rate predict” OR “failure rate estimate” OR “failure rate evaluation” OR “Degradation assessment” OR “Degradation model” OR “Degradation predict” OR “Degradation process” OR “Degradation state” OR “operational degradation” OR “Degradation indicator” OR “Degradation trend” OR “run-to-failure degradation” OR “Deterioration assessment” OR “Deterioration model” OR “Deterioration predict” OR “Deterioration process” OR “Deterioration state” OR “operational deterioration” OR “Deterioration indicator” OR “Deterioration trend” OR “Damage assessment” OR “Damage quantification” OR “Damage detection” OR “Damage identification” OR “Consequence model” OR “deformation monitoring” OR “Consequence analysis” OR “consequence evaluation” OR “Consequence assessment” OR “Failure probability evaluation” OR “Failure probability estimate” OR “Failure probability predict” OR “Failure probability update” OR “Risk probability” OR “Remaining useful lifetime” OR “Remaining useful life” OR “Remaining useful life-time” OR “Life predict” OR “Remaining life predict” OR “Remaining life estimate” OR “remaining life update” OR “remaining life evaluation” OR “Average useful life” OR “RUL estimate” OR “RUL predict” OR “Prognostic and health management” OR “Prognostics & health management” OR “Prognostics and health management” OR “Prognostics and health management” OR “health assessment” OR “health monitor” OR “health index construction” OR “health state identification” OR “state of health identification” OR “health status” OR “health state” OR “health indicator” OR “health image” OR “health condition” OR “health index” OR “Reliability Assessment” OR “Reliability analysis” OR “Reliability update” OR “Reliability evaluation” OR “Reliability estimate” OR “Reliability measure” OR “Reliability monitor” OR “Reliability predict” OR “Reliability life design” OR “updating reliability” OR “Reliability model” OR “Reliability management” OR “Reliability distribution” OR “reliability design” OR “survival analysis” OR “Risk Assessment” OR “Risk analysis” OR “Risk update” OR “Risk evaluation” OR “Risk estimate” OR “Risk measure” OR “Risk monitor” OR “Risk predict” OR “Risk management” OR “Dynamic risk assessment” OR “Probabilistic risk assessment” OR “Risk indicator” OR “Risk factor” OR “assessing evolving risk” OR “Quantitative risk assessment” OR “safety risk” OR “Risk mitigation” OR “risk and reliability analysis” OR “risk and reliability assessment” OR “upgrading risk” OR “upgrading risk” OR “risk warning” OR “risk correlation analysis” OR “risk reduction” OR “risk prevent” OR “prevent risk” OR “prevent risk” OR “predictive risk analytics” OR “risk model” OR “risk event” OR “risk source” OR “emerging risk” OR “Safety Assessment” OR “Safety analysis” OR “Safety update” OR “Safety evaluation” OR “Safety estimate” OR “Safety measure” OR “Safety monitor” OR “Safety predict” OR “Safety management” OR “Safety and reliability” OR “Safety and health” OR “Safety and probability” OR “safe and reliable” OR “upgrading safety” OR “Safety warning” OR “predictive Safety analytics” OR “Safety model” OR “safety design” OR “preventive safety” OR “safe and reliable service” OR “safe and reliable operation” OR “safe and desired behavior” OR “safe and desired behaviour” OR “safety education and training” OR “safety-management strategy” OR “safety-management policy” OR “safety performance” OR “safety standards development” OR “Resilience Assessment” OR “Resilience analysis” OR “Resilience update” OR “Resilience evaluation” OR “Resilience estimate” OR “Resilience measure” OR “Resilience monitor” OR “Resilience predict” OR “Resilience management” OR “Resilience and reliability analysis” OR “Resilience and reliability assessment” OR “upgrading resilience” OR “Resilience model” OR “resilience design” OR “management of resilience” OR “Emergency planning” OR “Emergency evaluation” OR “Predictive maintenance” OR “Preventive maintenance” OR “condition-based maintenance” OR “condition based maintenance” OR “prognostic maintenance” OR “risk-based maintenance” OR “risk-based maintenance” OR “maintenance plan” OR “maintenance decision-making” OR “operation and maintenance” OR “inspection plan” OR “risk-based inspection” OR “intelligent maintenance” OR “smart maintenance” OR “maintenance 4.0” OR “selective maintenance” OR “maintenance selection” OR “maintenance optim” OR “proactive maintenance” OR “maintenance scheduling” OR “maintenance schedule” OR “opportunistic maintenance” OR “defect inspection” OR “condition monitoring” OR “condition-based monitoring” OR “predictive monitoring” OR “intelligent monitoring” OR “Intelligent diagnosis” OR “Intelligent diagnostic” OR “Diagnosis model” OR “predictive statistic” OR “reliability parameter estimate” OR “reliability parameter evaluation” OR “reliability parameter predict” OR “anomaly detect” OR “abnormal condition detect” OR “anomaly predict” OR “abnormal condition predict” OR “vulnerability model” OR “vulnerability management” OR “integrity evalu” OR “Integrity management” OR “Availability Assessment” OR “Availability analysis” OR “Availability update” OR “Availability evaluation” OR “Availability estimate” OR “Availability measure” OR “Availability monitor” OR “Availability predict” OR “upgrading Availability” OR “Availability model”). AND.

TS=(-“Chemical process” OR “battery” OR “batteries” OR “petrochemical industry” OR “Petrochemical process” OR “Chemical industry” OR “reinforced concrete” OR “aircraft” OR “engine” OR “motor” OR “motors” OR “oil pipeline” OR “gas pipeline” OR “buried pipeline” OR “corroded oil” OR “drilling operation” OR “corroded pipeline” OR “construction site” OR “industrial plant” OR “engineering system” OR “safety-critical asset” OR “safety-critical application” OR “bearing” OR “landslide” OR “slope stability” OR “open-pit mine dump” OR “pavement” OR “lithium-ion batteries” OR “mechatronic system” OR “Electro-hydrostatic actuator” OR “multi-state system” OR “multi state system” OR “gas leak” OR “process system engineering” OR “tunnelling excavation” OR “civil engineering” OR “industrial application” OR “Infrastructure system” OR “steel rebar and concrete” OR “process operation” OR “Pipeline leak” OR “gearbox” OR “lifting operation” OR “belt conveyor” OR “braking system” OR “high-speed train” OR “Drilling and blasting operation” OR “turbine disk” OR “railway” OR “Blast-induced rockfall” OR “industrial product” OR “pump” OR “compressor” OR “power plant” OR “ship” OR “autonomous ship” OR “oil spill” OR “casting molding process” OR “Casting moulding process” OR “safety-critical operation” OR “Oil and gas industry” OR “Oil & gas industry” OR “industrial sector” OR “power distribution” OR “industrial control system” OR “bolted connection” OR “Coal and gas outburst” OR “coal miner” OR “corroded beam” OR “crude oil” OR “natural gas” OR “pump” OR “critical component” OR “aviation” OR “object strike” OR “offshore platform” OR “off-shore platform” OR “off shore platform” OR “onshore platform” OR “on-shore platform” OR “on shore platform” OR “turbine” OR “C-MAPS” OR “Nuclear plant” OR “Industrial engineering” OR “process industry” OR “process industries” OR “coal burst” OR “seismic event” OR “burst” OR “construction” OR “burst” OR “offshore” OR “construction” OR “offshore” OR “underwater” OR “buried oil pipeline” OR “subsea system” OR “mechanical equipment” OR “nuclear power plant” OR “complex physical system” OR “Passive safety system” OR “socio-technical system” OR “sociotechnical system” OR “cyberphysical system” OR “cyber-physical system” OR “multi-component system” OR “unattended machinery plant” OR “high-value asset” OR “Autonomous shipping” OR “Autonomous vessel” OR “Vessel” OR “autonomous navigation” OR “water supply network” OR “aero-propulsion system” OR “aero propulsion system” OR “offshore wind asset” OR “flip ship” OR

21
“traffic injur*” OR “crash injur*” OR “traffic crash*” OR “vehicle crash test” OR “Road safety” OR “Road user safety” OR “behavior-based safety” OR “behavioral-based safety” OR “crash risk” OR “crash prevent*” OR “crash predict*” OR “crash detect*” OR “crash prediction and detection” OR “crash mitigation” OR “traffic injury severity predict*” OR “Target tracking” OR “forest health monitoring” OR “follar sampling” OR “crown loss estimat*” OR “bioremediation” OR “toluene*” OR “fungi*” OR “biological waste” OR “CO2 solubility” OR “gas solubility” OR “Aqueous NACL solutions” OR “carbon dioxide-water” OR “vapor–liquid equilibria” OR “energy modelling” OR “energy modeling” OR “energy planning” OR “workplace accident” OR “fraud detect*” OR “role engineering” OR “role mining” OR “earth magnetic field” OR “document layout analysis” OR “saturation line forecasting” OR “workers unsafe action*” OR “marine fauna” OR “fish*” OR “sea trial*” OR “surveillance system” OR “tresspass*” OR “suicide” OR “voice recognition” OR “teaching” OR “pedagogy” OR “STEM education” OR “encrypted data” OR “compressed data” OR “data leak” OR “crowd activity” OR “crowd activities” OR “bone*” OR “knee joint*” OR “orthopaedic” OR “precipitate ligament*” OR “oxazation*” OR “Mesoporous Fe3O4/graphene oxide” OR “nanohybrid” OR “land degradation” OR “social capital” OR “degraded landscape” OR “desertification” OR “livelhoods” OR “forest fire” OR “cloud workload predict*” OR “licensed assisted access” OR “licensed-assisted access” OR “unlicensed” OR “flood risk analysis” OR “flash flood susceptibility” OR “electron” OR “electronics” OR “indoor safety” OR “indoor security” OR “price forecast*” OR “insurance price” OR “insurance pricing” OR “vehicle insurance” OR “pricing strategy” OR “pricing strategies” OR “traffic sign” OR “traffic signs” OR “price risk” OR “crude oil price” OR “apiarist” OR “apiarists” OR “bee swarming” OR “quantitative finance” OR “facial skin” OR “fever” OR “object detect*” OR “Android app” OR “Android apps” OR “logic bomb*” OR “DNA” OR “biological” OR “cytotoxicity” OR “nanoparticles” OR “submicroscopic particles” OR “sub-microscopic particles” OR “financial distress model” OR “smart city” OR “smart cities” OR “psychophysiological” OR “biosensor” OR “biosensors” OR “alien plant” OR “alien plants”).

References


