

Article

Insights into the Application of Machine Learning in Industrial Risk Assessment: A Bibliometric Mapping Analysis

Ze Wei ¹, Hui Liu ^{1,*} , Xuewen Tao ², Kai Pan ¹ , Rui Huang ¹, Wenjing Ji ¹ and Jianhai Wang ¹

¹ College of Quality and Safety Engineering, China Jiliang University, Hangzhou 310018, China; 1433381566wz@gmail.com (Z.W.); pankai9826@gmail.com (K.P.); huangrui8782@gmail.com (R.H.); jwenjing2001@gmail.com (W.J.); jianhai208@gmail.com (J.W.)

² Zhejiang Academy of Emergency Management Science, Hangzhou 310020, China; taoxuewen1@gmail.com

* Correspondence: hui.liu@cjlu.edu.cn; Tel.: +86-571-8687-5715

Abstract: Risk assessment is of great significance in industrial production and sustainable development. Great potential is attributed to machine learning in industrial risk assessment as a promising technology in the fields of computer science and the internet. To better understand the role of machine learning in this field and to investigate the current research status, we selected 3116 papers from the SCIE and SSCI databases of the WOS retrieval platform between 1991 and 2022 as our data sample. The VOSviewer, Bibliometrix R, and CiteSpace software were used to perform co-occurrence analysis, clustering analysis, and dual-map overlay analysis of keywords. The results indicate that the development trend of machine learning in industrial risk assessment can be divided into three stages: initial exploration, stable development, and high-speed development. Machine learning algorithm design, applications in biomedicine, risk monitoring in construction and machinery, and environmental protection are the knowledge base of this study. There are three research hotspots in the application of machine learning to industrial risk assessment: the study of machine learning algorithms, the risk assessment of machine learning in the Industry 4.0 system, and the application of machine learning in autonomous driving. At present, the basic theories and structural systems related to this research have been established, and there are numerous research directions and extensive frontier branches. “Random Forest”, “Industry 4.0”, “supply chain risk assessment”, and “Internet of Things” are at the forefront of the research.

Keywords: machine learning; industry; bibliometrics; knowledge mapping; risk assessment; safety



Citation: Wei, Z.; Liu, H.; Tao, X.; Pan, K.; Huang, R.; Ji, W.; Wang, J. Insights into the Application of Machine Learning in Industrial Risk Assessment: A Bibliometric Mapping Analysis. *Sustainability* **2023**, *15*, 6965. <https://doi.org/10.3390/su15086965>

Academic Editor: Paulo Santos

Received: 15 March 2023

Revised: 16 April 2023

Accepted: 19 April 2023

Published: 20 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A crucial role is played by industrial risk assessments in ensuring the safety of workers, the public, and the environment, as well as in complying with regulatory requirements. The number of risk assessment methods has grown, and the areas covered have expanded since the 1960s when U.S. chemical companies began to conduct more systematic safety risk assessments. Identifying, analyzing, and evaluating potential hazards and risks associated with industrial processes or activities are involved in these assessments. Hazard and Operability (HAZOP), Fault Tree Analysis (FTA), Safety Integrity Level (SIL), Event Tree Analysis (ETA), and Quantitative Risk Analysis (QRA) are traditional methods of risk assessment typically used. HAZOP is a structured inspection of a process or system that systematically identifies potential hazards and deviations from normal operating conditions. It is commonly used in industries such as chemistry, oil and gas, and nuclear [1]. FTA is a systematic approach used to identify and analyze the causes and consequences of system failures, frequently employed in industries such as aerospace, national defense, and nuclear [2]. Various other risk assessment methods are also widely utilized in industries such as aviation, transportation, and energy [3]. However, with the popularization and use of sensors, computers, and the Internet of Things in the industrial field, the amount of

data has increased dramatically. Traditional methods are not effective in predicting future events and processing large amounts of data. At the same time, the rapid development of technology makes industrial processes more complex, interrelated, and precise, which leads to new risks that are difficult to capture through traditional methods. In addition, the industrial environment is dynamic and constantly changing, which makes it difficult for static risk assessment to keep up with the pace of change. In view of these challenges, industrial organizations must adopt new risk assessment methods that are more suitable for the needs of modern industrial production, such as advanced machine learning technology.

Since 1991, when machine learning was first proposed as a tool for risk analysis and quality assessment [4], with new developments in machine learning research driven by advances in computer technology, artificial intelligence, and big data. This technology offers numerous advantages, including the ability to quickly and efficiently process vast amounts of data and identify potential risks before they occur, preventing accidents and reducing downtime. Additionally, machine learning can be customized to meet the specific needs of an organization, identifying unique risks to a particular industry or facility. By utilizing probability theory, statistics, and computational complexity theory, machine learning can achieve real-time and accurate risk assessment, handling all possible variables and predicting more potential risk factors than human evaluators. As an emerging computer technology, machine learning is becoming an increasingly important part of risk assessment. Examples of the use of machine learning for industrial risk assessment have increased in recent years. In the construction industry, scholars have compared the prediction effects of support vector machines, artificial neural networks, and kernel logistic regression on shallow landslides [5]. Some scholars use computational models to simulate the behavior of soil and fluid flow and evaluate the risk of failure or damage in structures built on or near soil [6,7]. In the biomedical industry, scholars have used machine learning to study the spread, diagnosis, and prevention of COVID-19 and fight the pandemic [8]. At the same time, machine learning has also been used to predict drug toxicity [9,10]. In the mechanical manufacturing industry, machine learning is used to diagnose rotating machinery faults [11], reduce the bullwhip effect in supply chains [12], implement predictive maintenance for wind turbines [13], and monitor network intrusions in transportation vehicles [14]. In the energy and chemical industries, machine learning has been used for the risk assessment of oil pipelines and the investigation of self-ignition risks during coal transportation [15–17]. In the fields of new energy vehicles and autonomous driving, some scholars propose a fault diagnosis system for new energy vehicles' electric drive systems based on improved machine learning and several typical fault detection and diagnosis methods [18].

The field of industrial risk assessment has seen significant advancements in recent years, particularly with the increasing use of machine learning techniques. However, this has also made it difficult to keep up with the evolving landscape and to identify the current status and future directions of research. To address this challenge, it is necessary to explore subjects such as (1) the integration of machine learning in industrial risk assessment, (2) the historical development of the field, (3) changes in the metaknowledge and research areas, and (4) the hot spots and trends in the field. One effective approach to gain insights into these topics is through bibliometric mapping analysis. This approach allows scholars to visualize the knowledge base, research hotspots, and development trends of a specific field, which can help researchers quickly understand the research overview of a specific area. Bibliometric mapping analysis has become increasingly popular in various research fields. For example, in the occupational accident analysis field, bibliometric methods have been applied to study the application of machine learning techniques [19], while in engineering risk assessment, they have been used to investigate the adoption of machine learning algorithms [20]. Bibliometric mapping analysis has also been utilized to examine the knowledge base and research hotspots related to the emergency management of sudden public health events [21] and determine the current research status and trends of emergency evacuation studies [22]. In this paper, we use bibliometric mapping analysis to

investigate the relevant literature on the application of machine learning in industrial risk assessment, employing software such as VOSviewer, Bibliometrix R, and CiteSpace. This paper discusses the following topics: (1) a comprehensive understanding of the historical development based on an analysis of the spatiotemporal distribution of related paper outputs; (2) research hotspots and research frontiers of machine learning in industrial risk assessment; (3) identifying potential future research directions in this field.

2. Materials and Methods

2.1. Data Collection

The literature on machine learning applied to industrial risk assessment can be retrieved from the Web of Science, which is widely considered one of the most comprehensive and highest-quality databases of English literature [23]. Science Citation Index Expanded (SCI-EXPANDED)-1900 and Social Science Citation Index (SSCI)-1900 from the Web of Science Core Collection were selected as the target databases for this paper. The search time ended on 15 December 2022, and the data retrieval process is shown in Table 1.

Table 1. List of search keywords used in the WOS.

No.	Selected Search Keywords	Number of Records	Periods	Dataset Used in Each Section
1	Machine Learning AND Industry AND (Risk OR Safety)	1620	1991–2022	Not used
2	Deep Learning AND Industry AND (Risk OR Safety)	733	1991–2022	Not used
3	Artificial Intelligence AND Industry AND (Risk OR Safety)	1004	1991–2022	Not used
4	Merge & De-duplicate	3116	1991–2022	Sections 3.1–3.4

The disciplines of artificial intelligence, machine learning, deep learning, and big data are interrelated and closely related, and the lack of standardized terms used for the fields of risk assessment and security assessment has led to different definitions of risk and security. To ensure that the literature search process captures all the relevant literature, different keywords need to be combined for the search, and this is shown in Table 1. By adding up the articles obtained from each keyword seen, a total of 3357 documents were obtained, and the same articles were de-weighted to finally obtain 3116 documents. In this paper, these 3116 articles are used as sample data for analysis.

2.2. Scientometric Methods

Due to the large number of documents analyzed, bibliometric methods are employed in this paper to quantitatively analyze and visualize the literature related to machine learning and industrial risk assessment. An effective way to summarize the research status of a particular field is provided via bibliometric mapping analysis. It can explore the development of a specific research area through co-occurrence analysis, co-citation analysis, and other methods. It can effectively analyze the quantity characteristics, structural distribution, internal quantitative relationships, and change patterns in the literature in order to determine the research trends, evaluate them, and predict the development of the discipline [17,24,25].

In this study, bibliometric mapping analysis is employed to perform co-occurrence, clustering, and overlapping analyses of the exported literature. The results were presented in a visual form using the VOSviewer, Bibliometrix R, and CiteSpace software to help understand the distribution, development process, research frontiers, and hot topics of the combination of machine learning and industrial risk assessment. The specific research process is illustrated in Figure 1.

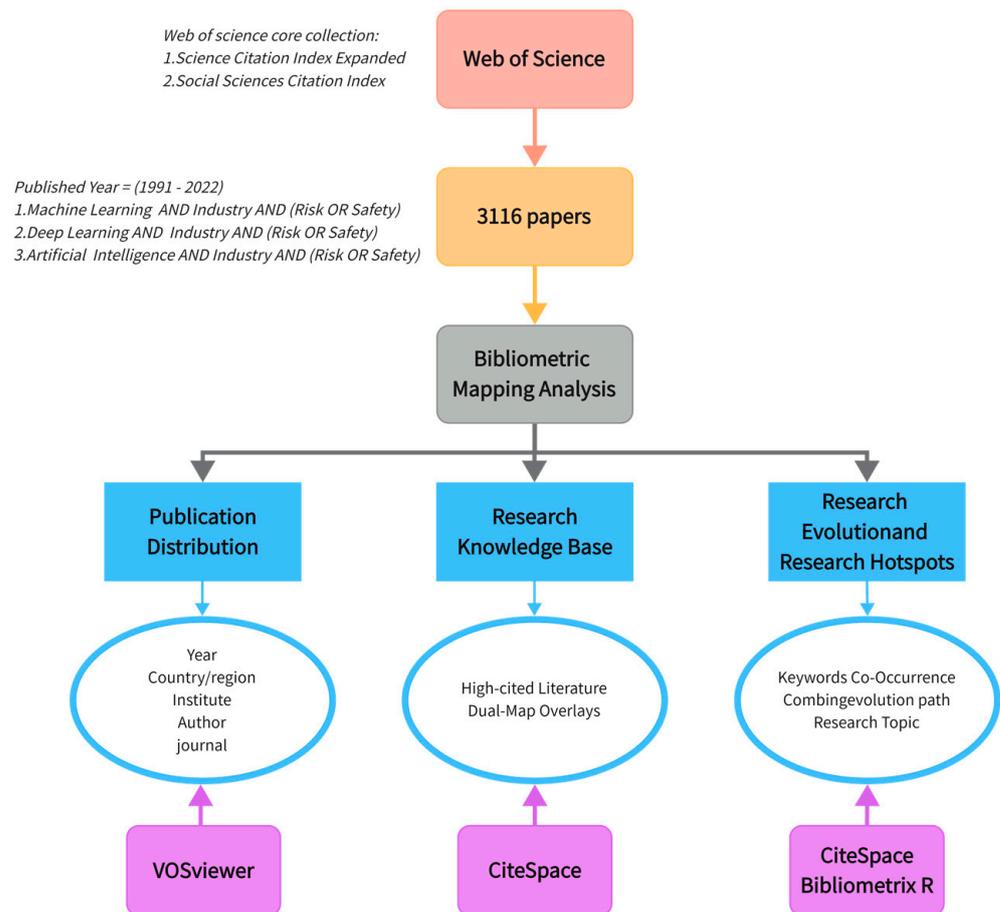


Figure 1. Procedures and methods used in research on the application of machine learning to industrial risk assessment.

VOSviewer software was employed to investigate international cooperation relationships, organization cooperation relationships, author cooperation relationships, and keyword cluster analysis of machine learning as applied to industrial risk assessment. Visualizations generated by VOSviewer helped identify key nodes in the network maps.

The thematic evolution of the literature was explored using Bibliometrix R, which computed various bibliometric indicators such as degree centrality, betweenness centrality, and closeness centrality. Degree centrality measured the number of co-authorships or citations of a node, while betweenness centrality gauged the number of shortest paths that pass through a node. Closeness centrality, on the other hand, was defined as the inverse of the average shortest path length.

CiteSpace utilized NLP techniques to extract keywords from article titles and abstracts, followed by frequency analysis and clustering algorithms to group similar keywords based on co-occurrence patterns. This clustering process identified clusters of related topics or themes in the literature. Finally, CiteSpace reviewed the keywords within each cluster to identify the common concepts or ideas they represent. This systematic approach provided insights into the topics and trends present in the analyzed literature on machine learning and industrial risk assessment.

3. Results

3.1. Temporal Distribution

In order to have a more complete and macroscopic view of the temporal distribution of the literature, the trend of the application of machine learning to industrial risk assessment research and the degree of interest in the last three decades can be visualized by the annual

publication volume and the average citations per item analysis of the retrieved sample data of the 3116 articles. Figure 2 shows the annual publication volume of the literature, the total publication volume, and the average citations per item. It can be seen from the data that the application of machine learning to industrial risk assessment research is increasing year by year, and the number of publications is rising. The overall development process can be divided into three stages: the initial exploration stage (1991–2006), the stable development stage (2006–2017), and the high-speed development stage (2017–present).

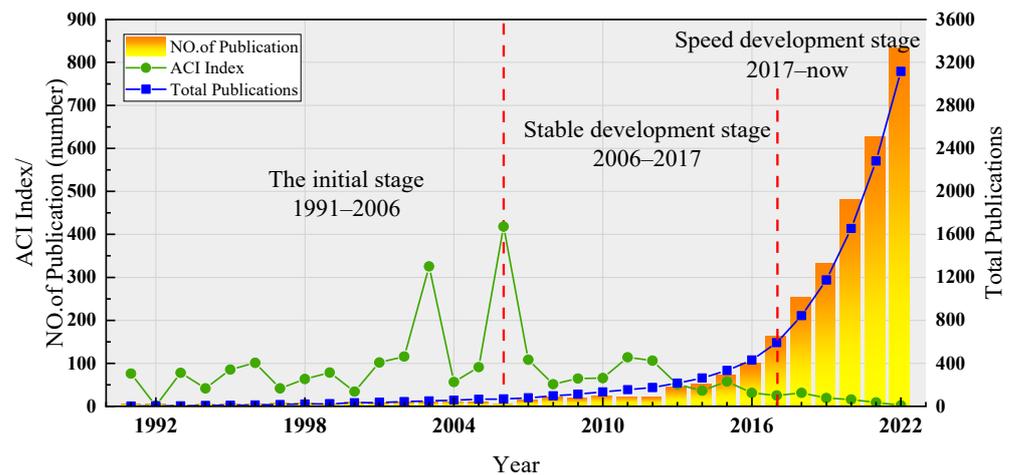


Figure 2. Publication growth trend around the world. ACI Index: Average citations per item.

During the initial exploration stage (1991–2006), little attention was given to machine learning due to limitations in computer hardware, resulting in a maximum of 10 relevant publications annually. Most of the highly cited literature in this period relates to machine learning for fault diagnosis and assisted production systems [26]. Moselhi et al. [27] pioneered the use of artificial neural networks as a management tool in construction, which compared favorably to traditional techniques such as probabilistic methods and expert systems. Fewer operators and computers are required by artificial neural networks, and they possess better decision-making and pattern-recognition capabilities. They are widely used in all construction engineering and management-level tasks. Jenkinson et al. [28] reviewed the use of artificial intelligence in fabrication yard operator support systems in the UK, highlighting its potential for real-time decision-making in complex scenarios such as planning, scheduling, and fault diagnosis. They believed that artificial intelligence had a promising commercial future with the continuous development of computer hardware and the potential to make industrial production safer and more efficient. In addition, the high citation frequency of the literature in 2003 and 2006 indicates that the literature published in these two years laid a foundation for subsequent research.

In the stable development stage (2006–2017), there was a continued rise in the number of annual publications, and a wide range of fields was beginning to be covered by highly cited literature on the application of machine learning to industrial risk assessment due to the unprecedented development of computer processing power, memory, storage, etc. Machine learning was widely used in the medical industry for assessing the risk of new drugs in clinical trials and performing image recognition to promptly assess the risk of patients suffering from cardiovascular diseases or tumors [29–31]. Risk assessment in various fields of industry, including building construction; mechanical processing; the food, energy, and chemical industries; transportation safety; and even the financial industry, was penetrated by machine learning [32–35]. Machine learning was significantly segmented in these fields, from risk identification at construction sites to risk prediction for natural disasters in building construction [36,37]. The macroscopic real-time monitoring of vehicle data flow in intelligent cities and the microscopic risk identification for autonomous driving were covered in the transportation field [38–40]. Pipeline safety evaluation for oil and gas

and advanced control systems in the power industry were covered in the energy and chemical fields [41–43]. These examples illustrate the success of machine learning in industrial risk assessment.

In the high-speed development stage (2017–present), the annual number of publications further increased, remaining above 150 per year, with a rapid rise in the annual number of publications. Due to the rise and gradual maturation of new information and communication technologies driving the industry’s shift toward intelligent monitoring, data fusion, and machine learning applications, the concept of “Industry 4.0” is more frequently mentioned during this period. A vital role is played by machine learning in the effective prediction of abnormal behavior in industrial machinery, tools, and processes to anticipate critical events and damage, prevent significant economic losses and safety issues, and reduce environmental pollution from industrial activities. Therefore, significant attention is paid to machine learning and artificial intelligence in this phase, and they are considered to be of high importance, making them a vital part of the realization of Industry 4.0 [44–46].

3.2. Spatial Distribution

3.2.1. Country/Region Distribution

The assessment of literature sources, through the analysis of their spatial distribution, provides valuable insights into the strength and influence of research on related topics in each country or region and can promote cooperation and communication among nations. The level of attention and importance given to the topic can be reflected by the geographical distribution of a field. In this study, an analysis of the publications on machine learning and industrial risk assessment was conducted, which involved contributions from 102 countries and regions. The top 10 countries were selected based on the total number of publications, as outlined in Table 2. The findings were illustrated through a world map that depicted the number of publications and collaborations (refer to Figure 3). The three leading countries in terms of the number of articles published, with their total publications accounting for over half (57.25%) of the literature analyzed, were China (856), the United States (640), and the United Kingdom (288), as emerged from the analysis of the publications on machine learning and industrial risk assessment.

Table 2. Top 10 productive countries/region, 1991–2022.

Rank	Country	Region	TP	Percentage	ACI	Total Link Strength
1	China	East Asia	856	27.47%	10.95	375
2	USA	North America	640	20.54%	20.09	447
3	England	Europe	288	9.24%	14.22	332
4	India	South Asia	252	8.09%	8.76	216
5	Australia	Australia	167	5.36%	17.70	231
6	Canada	North America	151	4.85%	13.50	140
7	Italy	Europe	144	4.62%	15.35	201
8	Germany	Europe	138	4.43%	18.18	188
9	South Korea	East Asia	132	4.24%	22.48	88
10	Spain	Europe	124	3.98%	12.12	169

Notes: TP: total publications; ACI: average citations per item.

In terms of regional distribution, the ACI of articles in Europe and the US was generally higher, which indicates that the overall quality of their publications was superior. This suggests a closer integration of machine learning techniques with industrial risk assessment and the production of more valuable and impactful research. Among the top 10 countries in terms of publication volume, South Korea (22.48), the United States (20.09), and Germany (18.18) had the highest ACI scores. This indicates that research from these countries is more influential and that the national publication volume is not necessarily indicative of the impact of the articles. Generally, the quality of articles published in developed countries

in Europe and America is higher, and the research in this field is more valuable, widely noticed, and cited.

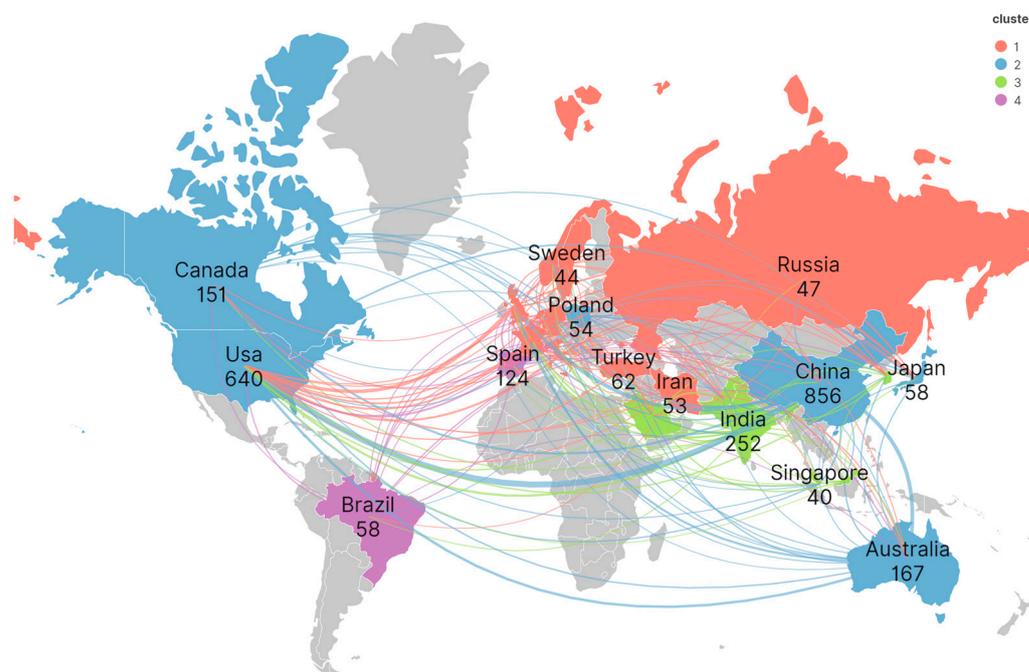


Figure 3. Network map of cooperation countries/regions.

An in-depth analysis of the level of international collaboration between countries in machine learning and industrial risk assessment research sheds light on the most active collaborators and the country's leading international research efforts in this field. Such insights are vital in fostering academic collaboration and facilitating the exchange of ideas and expertise on a global scale [47,48].

To analyze the data, the VOSviewer and Scimago Graphics software were combined to create a network diagram illustrating the country/region collaboration network (see Figure 4). The nodes are color-coded to represent clusters of countries that are collaborating, with the thickness of the connecting lines between nodes denoting the strength of the collaboration. The highest level of collaboration intensity is observed in countries with the most posts, such as China, the United States, and the United Kingdom, as demonstrated by the figure. The most extensive collaboration network is held by the United States, which links to multiple countries across all continents. Additionally, the blue clusters primarily consist of American countries and some Asian countries, while the red clusters indicate countries that are territorially connected. Factors such as geography and history significantly influence the clustering of countries.

Figure 4 presents a graphical representation of the amount of literature published in each country or region. The average years of publication in the corresponding country are denoted by the color of the nodes, while the number of publications in the respective country or region is indicated by the size of the node. China stands out with significantly higher publication rates than other countries. Meanwhile, countries such as the US, UK, France, Germany, and Spain, along with other developed countries in Europe and North America, have earlier publication years compared with their Asian and African counterparts.

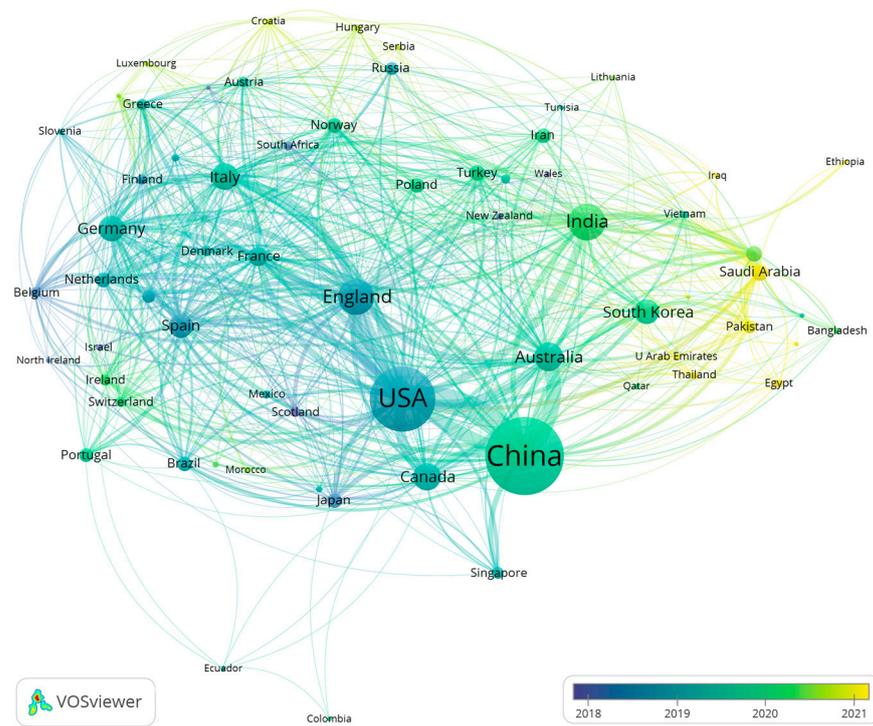


Figure 4. Collaboration network map of main countries/regions.

3.2.2. Institute Distribution

Scholars in related fields can benefit from analyzing research institution collaboration in the literature, as it provides insight into collaboration habits, research directions, and institutional results [49]. In this study, the retrieved data involved 3986 institutions, and Table 3 outlines the top 10 institutions that published the most relevant papers. Nine out of ten of these institutions are universities, highlighting that universities are the main driving force behind research in this field. Sixty percent of the institutions belonged to China, indicating that China is at the forefront of research in this field and is a significant contributor to research in this area. Among the top 10 institutions, Tsinghua University (36), the Hong Kong Polytechnic University (33), and the Chinese Academy of Sciences (31) had the highest number of publications, and Huazhong University of Science and Technology (23.50), the University of Illinois (20.00), and the Hong Kong Polytechnic University (18.76) had the highest average citations. However, the number of articles issued and the number of citations were not always correlated, as seen with the Wuhan University of Technology in China, which had the eighth-highest number of articles issued but a low number of citations, indicating low interest in its publications. Conversely, the University of Illinois had the lowest number of articles but the highest average citations, indicating that its published literature is of high value.

A collaboration graph of 59 major research institutions was generated using the VOSviewer software (see Figure 5). The size of the nodes in the graph represents the number of their published literature volumes, and the connecting line indicates the existence of cooperation between two institutions. The width of the connecting line indicates the strength of cooperation, and the color indicates the clusters formed by institutions that cooperate more closely. As shown in Figure 5, eight clusters were obtained based on the intensity of cooperation between research institutions. The Chinese Academy of Sciences and the University of the Chinese Academy of Sciences had the closest collaboration, with their research focused on machine learning in fault detection in nuclear power plants [50–52], which aligns with China's policy of investing heavily in the nuclear power field. The Hong Kong Polytechnic University and Huazhong University of Science and Technology were ranked second in the intensity of collaboration, with their research primarily focused on

using machine learning for construction risk and worker fatigue detection [53–55]. Monash University and the University of Melbourne were ranked third in collaboration intensity, with their primary research focusing on machine learning in image processing and intelligent perception for building construction [56]. Generally, related research institutions could be more aggregated, and the number of collaborative papers between institutions is low, with a higher frequency of collaboration between institutions in the same country or region. The main factors affecting inter-institutional academic communication and cooperation are the geographical area and the application direction of the research.

Table 3. Top 10 organizations in machine learning in industrial risk assessment studies, 1991–2022.

Rank	Organization	Country	TP	STC	ACI	Total Link Strength
1	Tsinghua University	China	36	456	12.67	58
2	The Hong Kong Polytechnic University	China	33	619	18.76	87
3	Chinese Academy of Sciences	China	31	362	11.68	77
4	Huazhong University of Science and Technology	China	24	564	23.50	43
5	Shanghai Jiao Tong University	China	23	219	9.52	31
6	Texas A&M University	USA	20	195	9.75	22
7	Norwegian University of Science and Technology	Norway	20	382	19.10	40
8	Wuhan University of Technology	China	17	111	6.53	21
9	The Pennsylvania State University	USA	17	155	9.12	18
10	University of Illinois	USA	17	340	20.00	36

Notes: TP: total publications; STC: sum of the times cited; ACI: average citations per item.

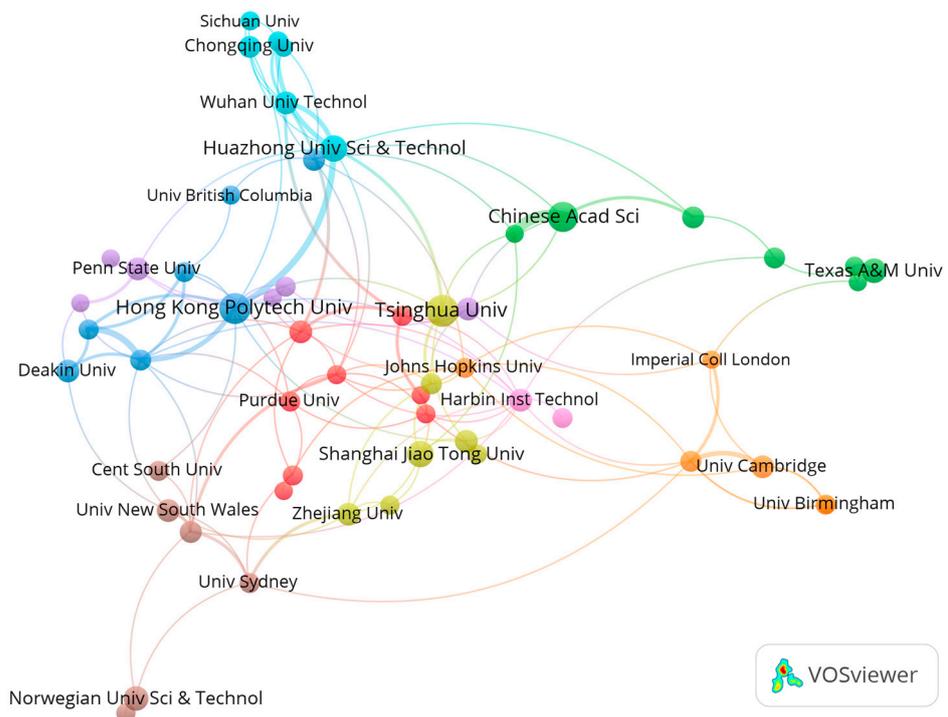


Figure 5. Network map of cooperation institutions.

3.2.3. Author Distribution

In the realm of academic research, high-output authors are widely recognized as influential leaders who make substantial contributions to the advancement of their respective fields. The research content of these authors not only embodies the substance and methodology of the entire research domain but also exerts a critical impact on the process of research development. In this study, we utilized the WOS core database to identify a total of 11,377 relevant authors. Of these, Table 4 showcases the top ten authors based on their countries and institutions, summarizing the main content areas covered by their articles

and revealing the frequency of their collaborations with other authors. Of particular note, Li Heng, with 11 publications and the highest average citation score of 31.55, emerges as the most prolific author, specializing in the use of deep learning for building construction safety inspection. Furthermore, his collaboration with other authors is the most frequent, signifying the centrality of his research within the field.

Table 4. Top 10 authors in machine learning in industrial risk assessment studies, 1991–2022.

Rank	Author	Country	Institute	Links	TP	ACI	Main Research Interests
1	Li, Heng	China	The Hong Kong Polytechnic University	22	11	31.55	Deep learning building construction safety
2	Liu, Xiang	China	The State University of New Jersey	6	7	10.14	Machine learning train derailment prediction
3	Zhao, Jinsong	China	Tsinghua University	3	6	16.00	Intelligent process fault detection
4	Wang, Lei	China	Nanjing Tech University	4	6	4.67	Machine learning-based real-time visible fatigue crack extension detection
5	Yu, Yantao	China	The Hong Kong Polytechnic University	18	6	33.00	Worker construction activity identification and monitoring
6	Umer, Waleed	England	University of Delaware	13	6	12.83	Machine learning-based body fatigue level identification and classification
7	Takahashi, Masakazu	Japan	University of Tsukuba	10	6	1.83	Validity detection for transportation decision-making research mail order industry
8	Arashpour, Mehrdad	Australia	The Hong Kong Polytechnic University	3	6	17.5	Machine learning 3D point cloud data processing for construction and infrastructure applications
9	Zhou, Jun	China	Ant Financial Services Group	11	5	8.6	Distributed learning in e-commerce
10	Zhang, Wei	China	Northeastern University	3	5	28.6	Deep learning in the fault diagnosis of rotating machinery

Notes: TP: total publications; ACI: average citations per item.

We employed VOSviewer to analyze the literature data, generating a collaboration network graph to gain insights into author collaboration patterns within the research field. After manually filtering 2400 author collaborations, we developed Figure 6, where nodes represent authors, and the node size corresponds to the number of their published works. The thickness of the node-to-node lines reflects the strength of author-to-author collaborations. Our analysis of Figure 6 reveals that authors Zhang Hao, Ma Lei, and Liu Yang, belonging to the dark blue cluster, concentrate on big data in data security and the use of deep learning in software countermeasures [47–49]. The purple cluster, including Lim, Ming, and Tseng Ming-Lang et al., focuses on the industrial Internet of Things and machine learning in Industry 4.0 [50,51]. The green cluster comprises Shariatfar, Moeid, Lee, Yong-Cheol, and others, who have studied intelligent noise identification and monitoring in factories and construction sites more intensively [52,53]. Conversely, the yellow cluster, centered on Li and Wei, concentrates on machine learning for industrial chemical toxicity prediction [54]. Lastly, the red cluster, including Zhou, Jun, and Wang, Lei, focuses on machine learning for risk prediction in the e-commerce and financial industries [55,56].

3.2.4. Journal Distribution

Academic journals serve as crucial platforms for disseminating academic knowledge, publishing research results, and promoting communication between researchers [57]. The selection of academic journals is indicative of the research areas and the literature quality of academic research.

Table 5 presents a comprehensive list of the top ten journals based on the number of publications, quantity, average citations (ACI), journal type (citation index), and impact factor (impact factor). It is noteworthy that *IEEE Access* published 98 papers, outpacing other journals in terms of publication quantity. *Automation in Construction* exhibited the second-highest average citation rate of 26.85 and an impact factor of 10.517, suggesting a significant influence on the research field. Hence, articles related to machine learning in industrial risk assessment published in *Automation in Construction* are of considerable importance and deserve the attention of scholars. Additionally, all relevant journals included

in the Science Citation Index Expanded (SCIE) indicate that the research predominantly falls within the natural sciences.

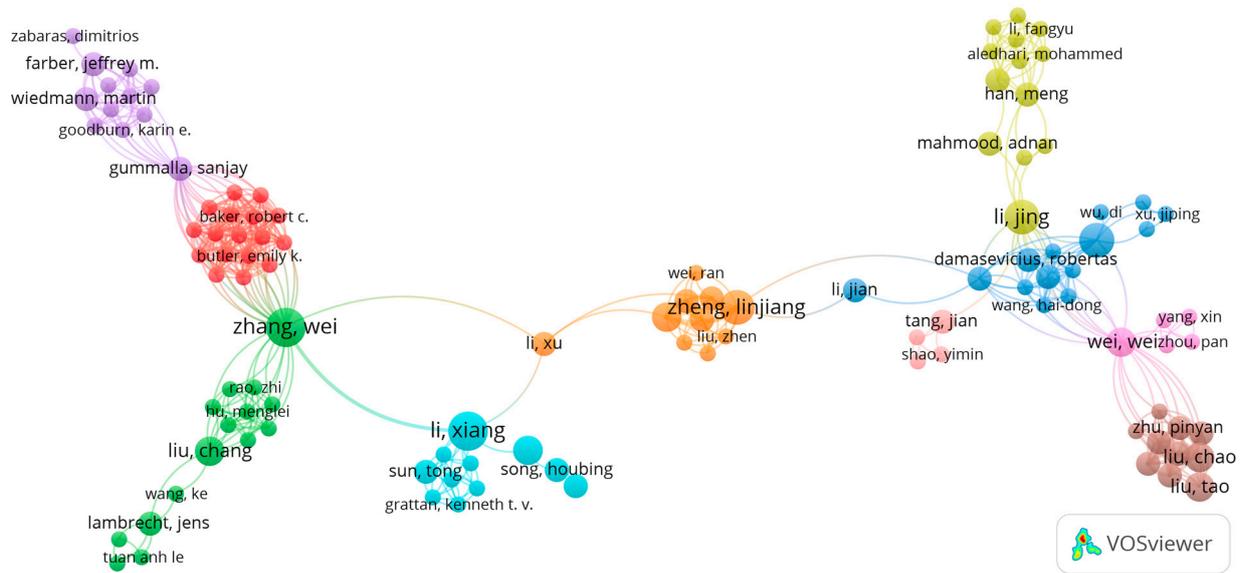


Figure 6. Network map of cooperating authors.

Table 5. Top 10 journals in machine learning in industrial risk assessment studies, 1991–2022.

Rank	Journal Title	TP	ACI	Citation Index	Impact Factor (2022)
1	IEEE Access	98	29.12	SCIE	3.476
2	Sustainability	57	8.70	SCIE/SSCI	3.889
3	Sensors	49	8.27	SCIE	3.847
4	Applied Sciences–Basel	47	6.45	SCIE	2.838
5	Automation in Construction	40	26.85	SCIE	10.517
6	Safety Science	37	23.68	SCIE	6.392
7	Expert Systems with Applications	32	19.94	SCIE	8.665
8	Computational Intelligence and Neuroscience	28	1.36	SCIE	3.12
9	Energies	23	3.78	SCIE	3.252
10	IEEE Transactions on Industrial Informatics	22	23.45	SCIE	11.648

Notes: TP: total publications; ACI: average citations per item.

We analyzed the distribution of significant source journals in the research area by using VOSviewer and retrieving data from 1680 journals. A network diagram of major journals was exported, as shown in Figure 7, where each node represents a journal, its size is proportional to its number of publications, and the width of the connecting lines between nodes represents the cross-citation intensity between journals.

The analysis reveals that IEEE Access, Sustainability, and Sensors are the top three journals in terms of the number of published articles. The inter-journal cross-citations highlight the diverse research directions and the significant research value of IEEE Access. The network diagram shows ten distinct clusters, each with its own specific research focus. The dark blue cluster primarily focuses on machine learning applications for factory safety and building construction. The Applied Sciences—Basel journal, the largest node in this cluster, closely collaborates with journals in the red, yellow, orange, green, and purple clusters, with the most substantial collaborations in the red cluster. The green cluster emphasizes integrating machine learning and the internet for control and automation in the context of building industrial risk monitoring systems through the industrial Internet of Things. IEEE Access leads the number of publications and has the closest association with the red and blue clusters. The red cluster, with the highest journal density, prioritizes computer vision, image recognition, and artificial intelligence methods, with a focus on building construction risk prediction and workers’ risk assessment. Safety Science and

Automation in Construction are the leading journals in this cluster. The yellow cluster's leading journal is Sustainability, focusing on the multidimensional and cross-cutting impact of digital intelligence on cities, transportation, Industry 4.0, and the future development outlook. The orange cluster studies machine learning algorithms for security risk prediction and optimization and explores the use of novel algorithms. This cluster needs to enhance its collaboration with other clusters. The brown cluster's leading journal is Computational Intelligence and Neuroscience Energies, focusing on the use of artificial intelligence in biomedicine, with a broad range of articles related to risk prediction in agriculture and risk assessment in personal healthcare and sports. This cluster needs more inter-cluster collaboration. The purple cluster's leading journal is Sensors, focusing on machine learning combined with various types of sensors for safety monitoring and real-time risk assessment in various areas of industry. Sensors collaborates closely with the red and dark blue clusters, as their subjects are rapidly gaining popularity in factories and construction sites. The pink cluster has fewer journals, primarily focusing on insurance-related risk assessment reviews. The leading journal in the light blue cluster is Expert Systems with Applications, which provides practical guidelines for the development and management of expert and intelligent systems in a wide range of areas.

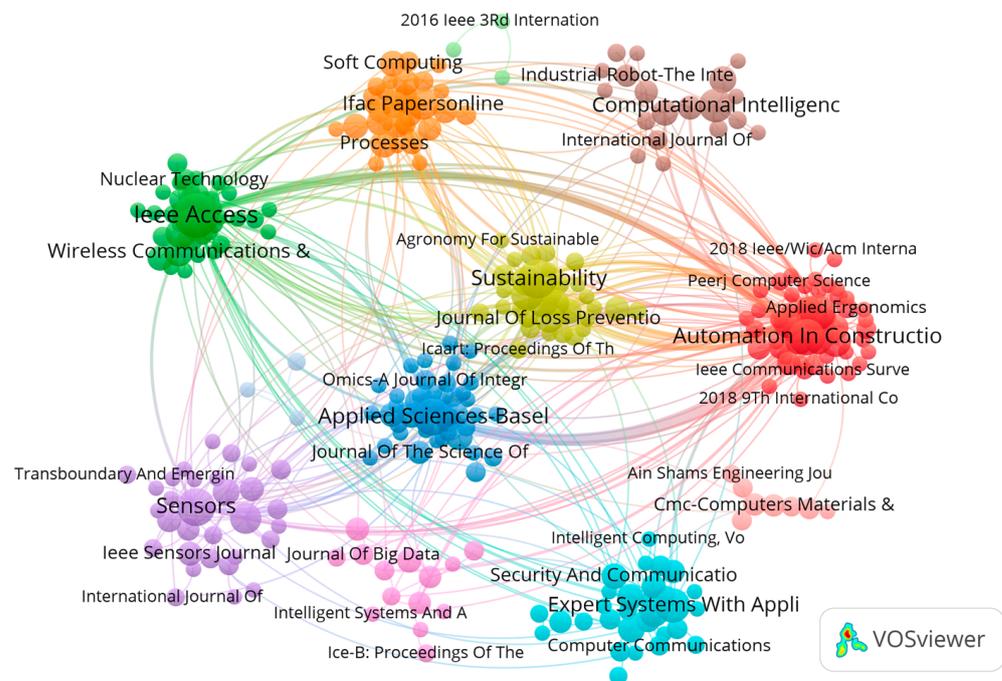


Figure 7. Network map of cooperation main research journals.

3.3. Research Knowledge Base

The knowledge base reflects the nature of the research field. There are multiple free knowledge units in different kinds of literature. When the same literature cites different kinds of literature, it represents a formal integration of multiple accessible units into a new knowledge base. As the citation network continues to evolve, a knowledge base is formed. Citation analysis of literature and journals can be an excellent way to tap into the knowledge base of a field [57].

3.3.1. High-Cited Literature Analysis

The assessment of scholarly output has evolved into a crucial component of academic inquiry. A popular tool for measuring academic impact has emerged in the form of citation analysis, which gauges the influence and quality of a publication by counting the frequency of its citations [23]. High citation rates often reflect the significance of the research content and outcomes, thus enhancing their reference value to future research

endeavors. Deliberate scrutiny and in-depth analysis of highly cited literature can provide a comprehensive understanding of the core research content in a given field.

The essence of industrial risk assessment research involving machine learning techniques is sought to be comprehended in this paper by selecting the 20 most cited related publications, as presented in Table 6. The literature's titles, authors, publication year, journal name, literature type, citations (STC), number of institutions (IN), and number of countries (CN) are itemized in the table. In co-authored works by over three authors, only the first author is listed. It is revealed by the top 20 frequently cited documents that 9 of them possess international collaborations, while 17 have institutional collaborations. This highlights the frequent intersection of institutional research directions in this field and the need for multidisciplinary teamwork to achieve effective research outcomes. A growing trend toward collective research pursuits, which significantly contribute to advancing scientific knowledge, is indicated by the prevalence of institutional collaborations.

Table 6. The top 20 papers with the most citations, 1991–2022.

Rank	Title	Journal	Type	Authors	Year	STC	IN	CN
1	A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-Based Methods	<i>Computers & Chemical Engineering</i>	Review	Venkatasubramanian et al. [26]	2003	1606	4	1
2	The Internet of Things for Health Care: A Comprehensive Survey	<i>IEEE Access</i>	Article	Islam et al. [29]	2015	1203	3	2
3	Machine Learning in Medicine	<i>Circulation</i>	Article	Deo et al. [30]	2015	1060	2	1
4	Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review	<i>Mechanical Systems and Signal Processing</i>	Review	Liu et al. [11]	2018	868	3	3
5	ProTox-II: A Webserver for The Prediction of Toxicity of Chemicals	<i>Nucleic Acids Research</i>	Article	Banerjee et al. [10]	2018	527	2	1
6	The Impact of Digital Technology and Industry 4.0 on The Ripple Effect and Supply Chain Risk Analytics	<i>International Journal of Production Research</i>	Article	Ivanov et al. [58]	2019	517	3	3
7	Sustainable Industry 4.0 Framework: A Systematic Literature Review Identifying the Current Trends and Future Perspectives	<i>Process Safety and Environmental Protection</i>	Review	Kamble et al. [44]	2018	429	2	2
8	A Digital Supply Chain Twin for Managing the Disruption Risks and Resilience in The Era of Industry 4.0	<i>Production Planning & Control</i>	Article	Ivanov et al. [59]	2021	314	2	2
9	A Critical Review of Smart Manufacturing & Industry 4.0 Maturity Models: Implications for Small and Medium-Sized Enterprises (SMEs)	<i>Journal of Manufacturing Systems</i>	Review	Mittal et al. [60]	2018	298	2	2
10	P4 Medicine: How Systems Medicine Will Transform the Healthcare Sector and Society	<i>Personalized Medicine</i>	Article	Flores et al. [31]	2013	262	2	1
11	Data Fusion and Machine Learning for Industrial Prognosis: Trends and Perspectives Towards Industry 4.0	<i>Information Fusion</i>	Article	Diez-Olivan et al. [45]	2019	236	4	2
12	Rapid and Quantitative Detection of The Microbial Spoilage of Meat by Fourier Transform Infrared Spectroscopy and Machine Learning	<i>Applied and Environmental Microbiology</i>	Article	Ellis et al. [9]	2002	217	1	1

Table 6. Cont.

Rank	Title	Journal	Type	Authors	Year	STC	IN	CN
13	A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective	<i>IEEE Access</i>	Article	Xu et al. [61]	2018	213	2	1
14	From Experience: Harnessing Tacit Knowledge to Achieve Breakthrough Innovation	<i>Journal of Product Innovation Management</i>	Article	Mascitelli et al. [62]	2000	208	1	1
15	Insights from Advanced Insights from Advanced Analytics at The Veterans Health Administration	<i>Health Affairs</i>	Article	Fihn et al. [63]	2014	194	6	1
16	Test Protocols for Evaluation of Spinal Implants	<i>Journal of Bone and Joint Surgery-American Volume</i>	Article	Goel et al. [64]	2006	193	1	1
17	Towards A Semantic Construction Digital Twin: Directions for Future Research	<i>Automation In Construction</i>	Review	Boje et al. [65]	2020	188	2	2
18	Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance	<i>IEEE Access</i>	Article	Yan et al. [66]	2017	186	2	1
19	The Neurobayes Neural Network Package	<i>Nuclear Instruments & Methods in Physics Research Section A-Accelerators Spectrometers Detectors and Associated Equipment</i>	Article	Feindt et al. [67]	2006	178	2	1
20	Occupation And Bladder Cancer Among Men in Western Europe	<i>Cancer Causes & Control</i>	Article	Kogevinas et al. [68]	2003	177	7	7

Notes: STC: sum of the times cited; IN: institute numbers; CN: country numbers.

It is revealed by the analysis that 15 of the 20 most cited articles are research articles, indicating that such studies are more representative and informative in this field. The most cited article, “A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-Based Methods” [26], has been cited 1606 times. Quantitative model-based methods in artificial intelligence for process fault detection and diagnosis are focused on in this review article, providing a valuable reference for industrial practitioners and technologists. The progression of abnormal events can be avoided and productivity losses can be reduced by the early detection and diagnosis of process faults, which is crucial for industrial risk assessment. This is helped by this review.

The article by Islam, S. M. Rizal, “The Internet of Things for Health Care: A Comprehensive Survey” [29], is the second most cited article with 1203 citations. Security and privacy concerns in the Internet of Things from a healthcare standpoint are discussed, and an intelligent collaborative security model is proposed to reduce security risks. Additionally, state-of-the-art network architectures and platforms, as well as industry trends in IoT-based healthcare solutions, are reviewed, providing avenues for future IoT-based healthcare research.

In the third most cited position, the article “Machine Learning in Medicine” [30] has been cited 1060 times. The problems in medicine that can be solved through large medical datasets and various learning algorithms are explored in this article, and the meaningful

contributions made by machine learning to clinical care over the decades are collated. Statistical learning methods are used to conclude the article and identify possible barriers to changing the practice of medicine, as well as discussions of how to overcome them. Therefore, significant insights into the potential of machine learning to transform healthcare have been provided by this article.

3.3.2. Dual-Map Overlays

A better understanding of the knowledge foundation of a research field can be provided by an analysis of the citation relationships and subject coverage of a journal, which is helpful for conducting further in-depth research [69]. Using the CiteSpace software, a dual-map overlay of journal citations and cited relationships was generated, as shown in Figure 8. The citation relationships between journals are represented by the curved lines between the left and right areas, and the strength of the citation relationship is represented by the thickness of the lines. Each cited journal in the right basic map is pointed to by a curve starting from a citing journal in the left basic map. The number of citations and references received by journals is indicated by the size of the circles representing them. The corresponding subject areas of the citations are indicated by the labels at both ends of the lines.

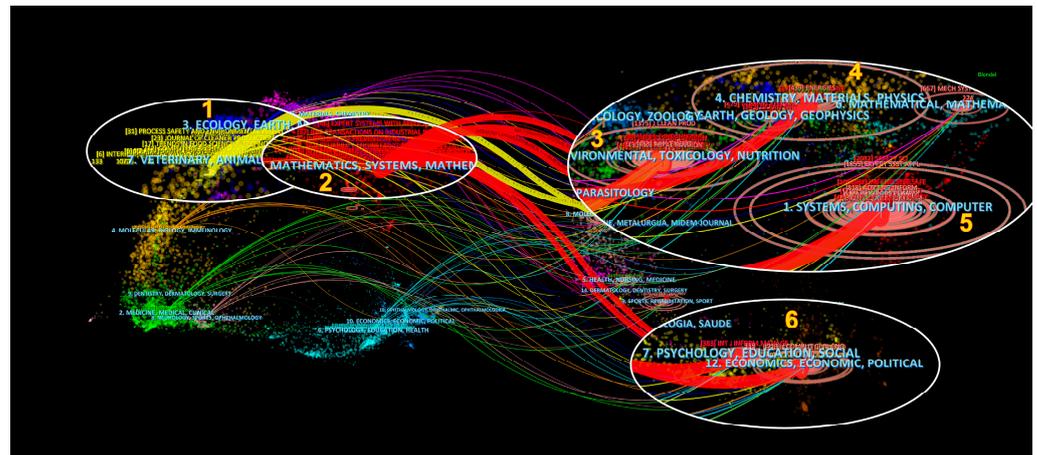


Figure 8. Citation overlays of machine learning in industrial risk assessment studies.

The citation map is divided into two halves: the left half and the right half. Journals such as *Process Safety and Environmental Protection*, the *Journal of Cleaner Production*, *Trends in Food Science & Technology*, *Computers & Chemical Engineering*, etc., which cover disciplines such as veterinary medicine, animals, and science, are contained in area 1 of the left half of the citation map. Three disciplines, mathematics, systems, and mathematics, are covered by three journals in area 2: *Expert Systems with Applications*, *IEEE Transactions on Industrial Informatics*, and the *Journal of Loss Prevention in the Process Industries*. The disciplines of environment, toxicology, and nutrition are covered by the more cited journals in Region 3 on the right half of the citation map: the *Journal of Cleaner Production* and *IEEE Transactions on Industrial Informatics*. The disciplines of chemistry, materials, and physics are covered by the *Energies*, *Sensors—Basel*, and *Ocean journals*, which encompass Region 4. The disciplines of systems, computing, and computers are covered by *Safety Science*, *Expert Systems with Applications*, and *Future Technologies*, which are included in Region 5. *Automation in Construction* and the *International Journal of Information Management* are two journals in Region 6 that are related to psychology, education, society, economics, and politics.

Seven curves that illustrate the citation relationships between the journals are showcased in the diagram presented in the figure. Of the seven curves, five are represented in red and are denoted as 2-3, 2-4, 2-5, and 2-6(2), while the remaining two are represented in yellow and correspond to 1-3 and 1-5. The highest number of cited journals is displayed in Region 2, with all five red lines directly connected to Regions 3, 4, 5, and 6. Mathematics

is the primary discipline associated with Region 2. Machine learning is widely applied in this industry, and its core algorithms continue to influence applied mathematics studies. A strong relationship with the other Regions is exhibited by Region 2, as mathematics serves as the foundation and core discipline of machine learning. Further mathematics research is necessitated by the continued development of machine learning in various fields.

Disciplines such as veterinary medicine, zoology, and science are encompassed by Region 1, and journals located in Region 3, which is focused on environmental science, toxicology, and nutrition, and Region 5, which is concentrated on computer science, are primarily cited. Integrating knowledge from toxicology, nutrition, environmental science, and computer science will facilitate the development of drugs for artificial intelligence, health monitoring, disease diagnosis, and environmental protection. Fundamental disciplines such as chemistry, materials science, and physics, which form the basis for industrial risk assessment, comprise Region 4. The incorporation of knowledge from these disciplines is required by machine learning, as a method, to achieve applications. The study and research of these fundamental disciplines are the foundation of industrial risk assessment, and integrating machine learning into these fields can facilitate its application. Continuous investment in these fundamental disciplines is necessary to deepen the application of machine learning in the industry. The most prosperous disciplines, such as psychology, education, society, economics, and politics, are encompassed by Region 6, and these are the research fields of machine learning in areas besides industry.

Based on the analysis of the above-mentioned disciplines and related journals, we can derive four knowledge bases for machine learning in industrial risk assessment: machine learning algorithm design, applications in biomedicine, risk monitoring in construction and machinery, and environmental protection.

3.4. Research Evolution and Research Hotspots Analysis

3.4.1. Keywords Co-Occurrence Analysis

Keywords can be considered a summary of the main content of an article, and the time and frequency of the occurrence of various keywords and the changes they undergo can be reflected in the research hotspots of that research area. Therefore, keyword analysis can be used to identify the evolving research frontiers related to the field of knowledge. A total of 5173 keywords were obtained from 3116 documents in the Web of Science core database. The top 20 keywords in terms of frequency of occurrence and the year of their first occurrence were taken to generate Table 7.

Table 7. The top 20 keywords of machine learning in industrial risk assessment studies, 1991–2022.

Rank	Keywords	Count	Year	Rank	Keywords	Count	Year
1	Machine Learning	603	2007	11	Blockchain	47	2018
2	Artificial Intelligence	362	1992	12	Cloud Computing	46	2010
3	Deep Learning	286	2017	13	Convolutional Neural Network	78	2016
4	Big Data	212	2013	14	Risk Assessment	41	2010
5	Industry 4.0	198	2016	15	Privacy	40	2013
6	Safety	84	1994	16	Fault Diagnosis/Detection	66	2003
7	Internet of Things	153	2015	17	Feature Extraction	39	2013
8	Security	69	2004	18	Classification	38	2004
9	Risk Management	54	2004	19	Random Forest	36	2012
10	Data Mining	53	2008	20	Automation	35	2003

Year: When the keyword first appeared.

In the current study, a keyword co-occurrence network was generated by using the CiteSpace software, as depicted in Figure 9. In this network, each circle denotes a keyword, with its size reflecting the frequency of occurrence.

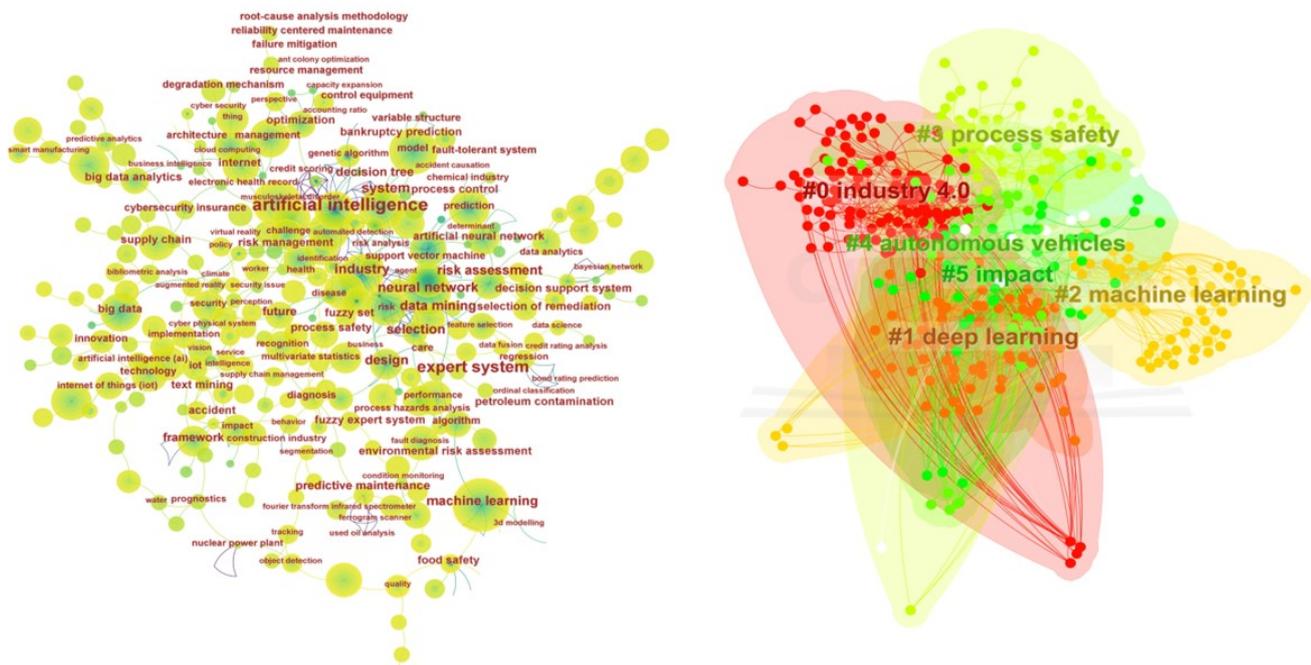


Figure 9. Network map of co-occurrence keywords.

The clustering of keywords is based on their co-occurrence, As illustrated in Figure 9, six clusters were identified, including Industry 4.0, deep learning, machine learning, process safety, autonomous vehicles, and impact. The research trends in the field are accurately reflected by the clustering topic terms, which provide guidance for future research. Of note, #0 and #2 emerged as the most frequent clusters, indicating that they are the hot topics of research in this field. Valuable insights for scholars and practitioners in the field of safety science to identify and prioritize future research directions can be provided by these findings.

#0 Industry 4.0: The characteristics of Industry 4.0 include the integration of various sensors, controllers, and technologies such as mobile communication and artificial intelligence that have sensing and monitoring capabilities. The production efficiency of the manufacturing industry will be significantly improved while the risks associated with each production step will be greatly reduced. An unparalleled upsurge in the use of machine learning algorithms has resulted from the rapid establishment of Industry 4.0, coupled with the emergence and utilization of the industrial Internet of Things. Crucial terms such as “Internet of Things”, “supply chain”, “big data”, and “implementation”, which are fundamental components of machine learning in industrial applications, comprise the keywords included in this cluster [44,45].

#1 Deep Learning: The ideal of artificial intelligence is being brought closer to realization by deep learning, a nascent research area within machine learning. The quick identification of patterns within sample data is enabled by its sophisticated algorithms, leading to the faster recognition of text, images, sounds, and more. A wider range of problems across multiple domains, including complex industrial applications, can be addressed by deep learning as a result of its potential. Clustered keywords such as “anomaly detection”, “rule”, “vision”, “fault diagnosis”, “recognition”, and “CNN” demonstrate that deep learning is primarily used in the industry for fault diagnosis, defect detection, and computer vision discrimination. These applications have proved invaluable in tackling complex problems such as risk detection and risk assessment, which require real-time, speedy identification and decision-making [70–73].

#2 Machine Learning: The discipline of machine learning is at the core of artificial intelligence, encompassing statistics, complex algorithms, and related knowledge such as the Alcoa wheel. Frequently occurring keywords, such as logistic regression, Random

Forest, neural network, and support vector machine, among others, are featured in this cluster. These algorithms are commonly utilized across a range of fields. For instance, Kibria applied logistic regression to predict the risk of secondary cardiovascular disease [74], while Liu and Tao used the Random Forest algorithm to analyze risk in financial markets [75]. Furthermore, Sresakoolchai utilized a neural network algorithm to detect DDOS attacks on bank monitoring systems [76]. “Surveillance” and “classification”, which represent the primary functions of machine learning for risk assessment through surveillance and classification, are the most commonly used keywords in this cluster.

#3 Process Safety: Supervising complex industrial processes in the energy, chemical, nuclear, and aerospace industries to prevent and manage accidental disturbances and hazardous situations during industrial production, storage, and transfer is involved in process safety. This aims to prevent harm to employees and community residents, as well as environmental damage and property loss. In the central cluster, the most frequently mentioned keywords are uncertainty, oil, abnormal situation management, and corrosion. Machine learning methods that can be used to detect abnormal risks and diagnose faults during the transportation and storage of oil, gas, and chemicals are represented by these keywords. The corrosion and rupture of pipelines, storage tanks, transportation carriers, and other facilities can be prevented by these methods, as reported in previous studies [15,77–79].

#4 Autonomous Vehicles: In recent years, autonomous driving has been a highly debated research topic. Heavy reliance is placed on the utilization of machine learning algorithms for the achievement of autonomous driving systems. Automated vehicle systems are equipped with numerous sensors, which generate data that require real-time processing. In order to detect and evaluate driving environments and risks in real-time, machine learning algorithms are indispensable in ensuring the safety of autonomous driving [80,81].

#5 Impact: The impact of accidents or risks in various fields, including society, assessment, mitigation, the economy, and the environment, is mainly elucidated by this cluster. Within this cluster, Paes, Vanessa Marques, delves into how industrialization and the development of artificial intelligence can contribute to the emergence of prejudice and discrimination in society and the increased risk of harm to humans from autonomous systems. Additionally, Santana, Julio Ariel Duenas, proposed a new economic loss index for fires and explosions using machine learning algorithms. The sharing of resettlement can be improved and guidance for accident prevention, mitigation, and risk management can be provided by this method [82–84].

3.4.2. Combing Evolution Path

In the present study, the CiteSpace software was employed to examine the timeline view analysis, and the generated keyword timeline view was subjected to appropriate editing, as illustrated in Figure 10.

The illustration of the inter-cluster relationships and the historical span of keywords in a specific cluster, showcasing the historical development of different research areas, was facilitated by the timeline view. A time axis along the horizontal axis and the various clusters of keywords along the vertical axis are featured in the keyword timeline view. A unique keyword is denoted by each node, with the frequency of the keyword indicated by the size of the node. The sudden change in the keyword from 1991 to 2022 is reflected in the color of the node, ranging from dark to light. The node outlined in purple has a greater centrality than 0.1. The earlier temporal suddenness is represented by a darker node color. The co-occurrence relationship between two nodes is represented by the line connecting them, with the level of co-occurrence indicated by its thickness. Thus, the relationship between two nodes can be quantitatively evaluated using this approach.

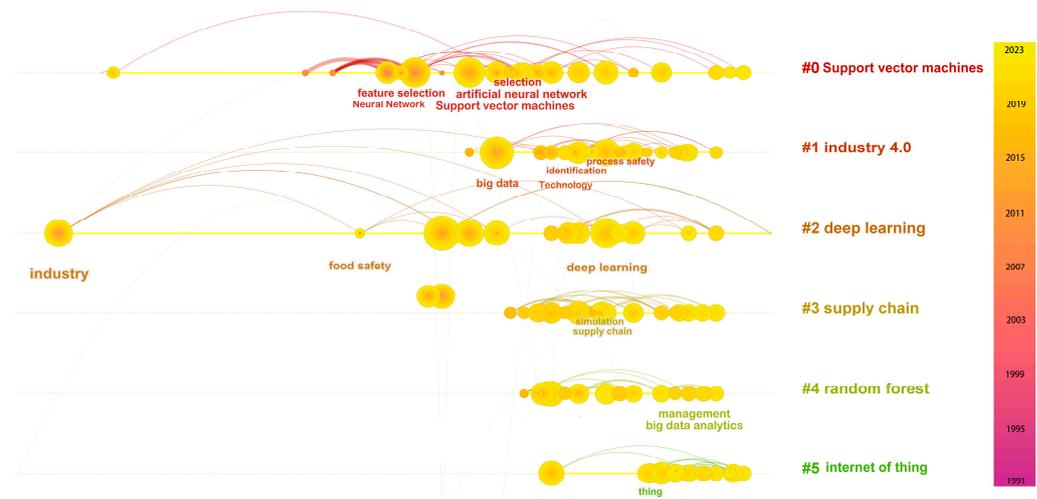


Figure 10. The keyword timeline view.

Six distinct timelines, each representing a unique research theme, can be broadly classified based on the keywords associated with machine learning in the domain of industrial risk assessment research, as depicted in Figure 10. Keywords such as “neural network”, “support vector machine”, “model”, “performance”, and “analytics” are featured in the first timeline, which spans from 1999 to 2022. This timeline is characterized by research on the application of support vector machine (SVM)-based algorithms in fault detection and diagnosis for process safety. SVM is a popular machine learning algorithm used for classification and regression tasks. It works by finding the best possible decision boundary, known as a hyperplane, that can separate data points into different classes or predict numerical values. It offers higher fault diagnosis rates and shorter diagnosis times than traditional artificial intelligence algorithms such as artificial neural networks (ANN), K-nearest neighbors (KNN), and decision trees (DT).

The second timeline, featuring keywords such as “big data”, “industry 4.0”, “cloud computing”, and “anomaly detection”, spans from 2016 to 2022 and focuses on integrating machine learning into research related to industry 4.0. However, relevant research in this area is still in its infancy, and this timeline was created relatively late.

The third timeline, ranging from 1997 to 2022, is based on keywords such as “industry”, “systems”, “machine learning”, and “convolutional neural networks”. The research content primarily revolves around the application of deep learning to the industrial sector. While research in this area was carried out early, the research direction was relatively singular initially due to the small research population for deep learning. It was not until after 2008 that the research directions related to deep learning began to diversify, the number of application industries increased, and the types of algorithms became more diverse.

The fourth timeline is characterized by keywords such as “simulation”, “challenge”, “impact”, and “behavior”, and is mainly focused on the prediction and assessment of supply chain risks in the industrial system using machine learning. It spans from 2016 to 2022.

Spanning 2017 to 2022, the fifth timeline was based on keywords such as “classification”, “big data analytics”, and “Random Forest” and focused on research on Random-Forest (RF)-based algorithms for risk prediction. RF is a type of machine learning algorithm that combines multiple decision trees to make predictions or classifications. It works by building a collection of decision trees, where each tree is trained on a random subset of data and features. The predictions from all the individual trees are then combined to produce a final prediction or classification. By comparing the first timeline, we can see that support vector machines were used earlier and are more mature in the field of risk assessment. However, as Random Forest algorithms are better suited to handling large and complex datasets, they can handle missing data and classification features well. Therefore, in the era of big data today, Random Forests are starting to become more popular.

Spanning 2017 to 2022, the sixth timeline is characterized by keywords such as “Internet of Things”, “privacy”, and “blockchain technology”. The development of machine learning in the Internet of Things is the focus of research in this area. Although the current research is still in its early stages, a promising new direction is held by the use of machine learning algorithms to realize the Internet of Things in the future.

3.4.3. Research Topic Evolution

It has been shown that clustering algorithms, when applied to co-occurrence networks, are a useful tool for identifying and highlighting different topics within a domain. The resulting output from these algorithms can be visualized in a topic graph, where the two dimensions of centrality and density are used to provide insight into the structure of the underlying dataset [85].

Based on the values of centrality and density, the topic graph can be divided into four quadrants, each representing a different level of centrality and density. The importance of a given topic in the broader field of study is reflected by the centrality measure, which increases along the horizontal axis. The strength of word links within a given topic is indicated by density, which increases along the vertical axis [86]. The higher the density of links within a topic, the more coherent and complete the research questions corresponding to that topic. The size of each domain on the graph is proportional to the number of articles containing the respective keyword. The core of the database is represented by themes located in quadrant 1 (top right), as they are both central to the public network and strongly internally connected, indicating a high degree of development. Themes that are less central but have a higher internal connection strength, suggesting that they are more mature as standalone research content, are contained in quadrant 2 (top left). A marginal area of research is indicated by themes located in quadrant 3 (bottom left), which has low linkage strength both within and outside its themes. Themes with relatively low internal linkages, which are gradually developing and still need to form a well-established research core, comprise quadrant 4 (bottom right) [85]. It is worth noting that the adjacent themes in the graph may not be linked and may only represent similar centrality and density data. Valuable insights into the structure of a dataset can be provided by clustering algorithms applied to co-occurrence networks visualized on a topic graph, and the resulting output can aid in identifying areas of research that require further development.

The research landscape of machine learning applied to industrial risk assessment is subject to temporal variability. To effectively capture and track the evolution of this field, we leveraged the Bibliometrix R package to identify topic terms that occur more than five times during three distinct phases of literature distribution. Utilizing these terms, we constructed topic maps for each period, allowing for the visualization of research hotspots and their temporal changes.

1991–2006: During the period of 1991 to 2006, the research landscape of industrial risk assessment using machine learning was characterized by a scattered research system with few related studies. As depicted in Figure 11, this topic is not located in the first quadrant, indicating a lack of a stable and long-term research system for machine learning in industrial risk assessment. Meanwhile, the research on process control and safety was becoming a coherent system with strong inter-study links, albeit in a system of independent research. Some scholars were also combining accident studies with algorithms, which are located in quadrant 2. On the other hand, artificial intelligence, expert systems, and process safety are positioned in quadrant 4, indicating that research, development, and applications in these areas were progressing but that core theories had not yet been developed.

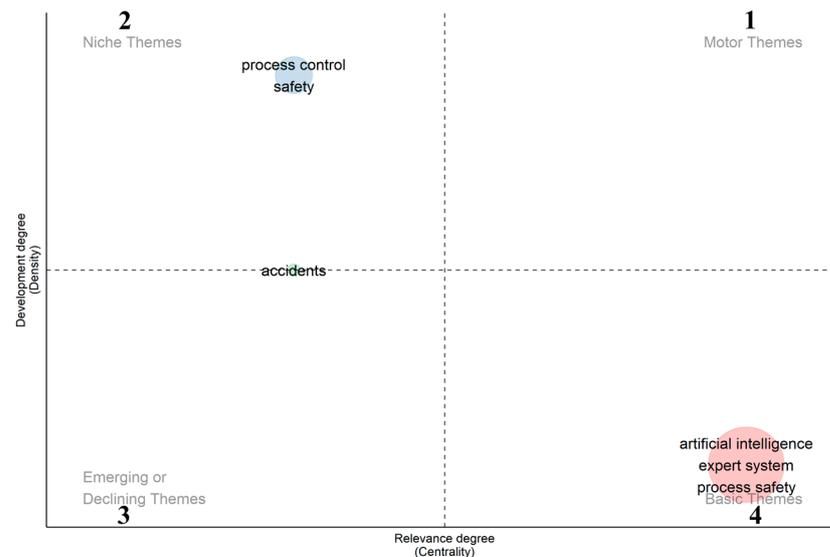


Figure 11. Thematic map based on co-word network analysis and clustering: 1991–2006.

2006–2017: In the second phase, emergent grouping themes can be observed, as shown in Figure 12. The rapid advancement of machine learning has been facilitated by the development of increasingly sophisticated computers, which has enabled companies to utilize it for managing risks efficiently, accurately, and at low cost through a holistic and systematic approach. As a result, core areas of academic research have emerged in machine learning and risk management. During this period, a new industry has emerged in e-commerce, and different algorithms to monitor the possible risks faced by this industry are being investigated through more studies. A research hotspot has emerged in the field of food safety, and the application of machine learning is increasingly required for the detection of food quality. Anomaly detection, cyber-physical systems, unsupervised learning, and construction safety are new emerging research topics located in quadrant 2, characterized by high density but low centrality and more mature yet independent research content. Technical topics such as data security, data analytics, and risk assessment are positioned at the junction of quadrants 3 and 4, with related research moving from the margins to the mainstream and related theoretical systems gradually taking shape. In quadrant 4, the richest themes are at the next stage of popularity in research, with big data, artificial intelligence, deep learning, neural networks, cloud computing, and other related subjects closely linked. The number of papers produced by the related themes in this stage is the largest in terms of size, reflecting the popularity of this stage of research. However, it should be noted that the Herring theory has yet to be fully developed and still holds great potential for further development. It is evident that digitalization and information technology are booming, and machine learning is becoming increasingly integrated into the industrial field.

2017–2022: As depicted in Figure 13, machine learning and risk assessment, which experienced rapid development in the previous phase, have become the focal point of research in this phase. The independence of risk assessment and machine learning, which was relatively established in the previous phases, is now closely intertwined. Research in related fields has matured, and the use of machine learning techniques for industry risk assessment has become the primary research objective. With the rapid development of the new energy vehicle industry, new research topics such as autonomous driving, predictive modeling, and data modeling have emerged. These topics are closely related and relatively mature in independent research. However, the volume of articles is small, and their centrality is low, indicating a considerable development potential (quadrant 2). As machine learning becomes more intricate, algorithms such as convolutional neural networks reside in quadrant 3. With more novel algorithms, their research is relatively marginal as these

topics move closer to the center. In quadrant 4, big data, artificial intelligence, security, and risk management are still popular research areas, and the output of related articles remains high. It is worth mentioning that Industry 4.0 has become a new research hotspot in this phase, indicating the gradual realization of the industrial sector’s informatization and digitization. However, a perfect research core has not yet been formed for these topics, and they still hold excellent development potential.

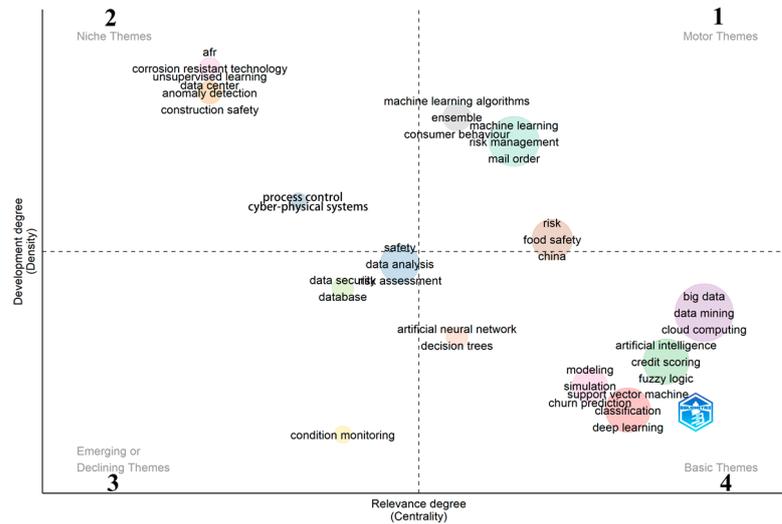


Figure 12. Thematic map based on co-word network analysis and clustering: 2006–2017.

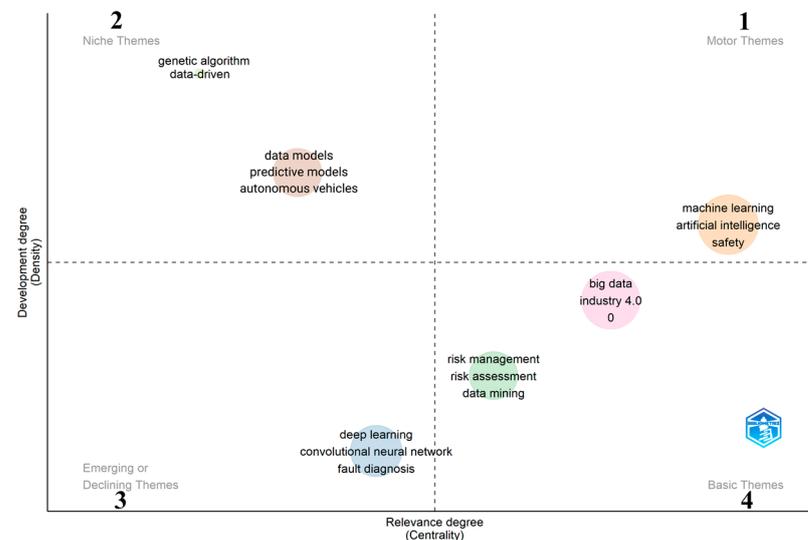


Figure 13. Thematic map based on co-word network analysis and clustering: 2017–2022.

The development of machine learning in the industrial field can be summarized as follows: from being an auxiliary tool or system for single-domain production to multi-domain risk assessment, safety management, and the informationization and digitization of Industry 4.0. The evolution of research topics in the field of machine learning, from its initial applications as an aid in single-domain production to its current role in supporting broader industrial risk management and safety, is highlighted in this overview. Additionally, the growing importance of the Industry 4.0 system, which emphasizes the use of advanced technologies to improve industrial processes, is a prominent theme in this field. The need for continued research into the application of machine learning in industrial settings, as well as its potential to support broader industrial transformation, is underscored by these developments.

4. Discussion

In this study, 3116 relevant publications were extracted from the Web of Science core database using VOSviewer, Bibliometrix R, and CiteSpace for the application of machine learning in industrial risk assessment, and the keywords were analyzed via co-occurrence analysis, cluster analysis, and dual-map overlays. An overview of the development history, topic evolution, relevant knowledge base, current research hotspots, and future research trends in the application area of machine learning in industrial risk assessment is presented. The main content is presented below:

The number of articles on machine learning in industrial risk assessment has increased year by year, indicating a growing interest in the field. The overall development process of machine learning in industrial risk assessment can be divided into three stages: the initial exploration stage (1991–2006), the stable development stage (2006–2017), and the high-speed development stage (2017–present), as shown in Figure 2 in Section 3.1. During the initial exploration stage (1991–2006), fewer than 10 relevant publications per year were produced. As machine learning was a relatively cutting-edge technology at the time and was limited by computer hardware, it did not receive widespread attention. However, an increase in the number of publications per year was observed during the stable development stage (2006–2017). With the improvement of computer processing capabilities and the arrival of the information age, machine learning was required to handle more complex learning tasks and was successfully applied to risk assessment in various industrial fields. In the high-speed development stage (2017–present), a rapid increase in the number of publications was indicated by the annual publication volume, which remained above 150 papers per year. During this period, due to the rise and gradual maturity of information and communication technologies, there is a need to establish big data prediction models, the Internet of Things, and autonomous driving and promote the Industry 4.0 model transformation toward intelligent monitoring and data empowerment, which are more complex tasks. As a result, widespread attention and high regard have once again been received by machine learning and artificial intelligence.

By selecting the top 20 most cited relevant articles, we have gained a deeper understanding of the research essence of using machine learning technology for industrial risk assessment. It was shown by the analysis results that research articles are more representative and informative in this field, with 15 of the top 20 most cited articles being research articles. The focus of the most cited articles is on model-based quantitative methods in artificial intelligence for process fault detection and diagnosis. The security and privacy issues of the Internet of Things from a healthcare perspective are discussed in the second most cited article. The potential of machine learning to solve medical problems using large media datasets and various learning algorithms, thereby transforming healthcare, is explored in the third most cited article. Various aspects within the industry are shown to be covered by research on machine learning in the world through the analysis of highly cited literature. Of the 20 articles, 9 were produced through international collaborations, and 17 were the result of multi-institutional collaborations, demonstrating the close collaborations between scholars in the field across institutions and regions, as well as the closely interdisciplinary nature of the research content.

By using CiteSpace to analyze the keywords in a dual-map overlay, we found that mathematics and computer science are the most important fields for research into how machine learning is used in industrial risk assessment. These two disciplines provide the basis for the development and optimization of a wide range of algorithms, from theory to application. In addition, this research incorporates knowledge from a number of disciplines, including chemistry, physics, materials science, nutrition, environmental science, toxicology, finance, and sociology, for its application in a variety of fields, including construction, energy, chemical engineering, and biomedicine. Based on the analysis in Section 3.4.2, we can derive four knowledge bases for machine learning in industrial risk assessment: machine learning algorithm design, applications in biomedicine, risk monitoring in construction and machinery, and environmental protection.

Using CiteSpace for keyword timeline analysis, we have obtained six timelines with two main directions: algorithms and applications for risk assessment. Three clusters were included in the algorithms for risk assessment: “support vector machines”, “Random Forests”, and “deep learning”, while three clusters were included in the applications for risk assessment: “Industry 4.0”, “supply chain risk assessment”, and “Internet of Things”, as shown in Figure 14. High keyword frequencies were observed in the period of 2017–2022 for four clusters, namely, “Random Forests”, “Industry 4.0”, “supply chain risk assessment”, and “Internet of Things”, indicating that they are currently at the forefront of research.

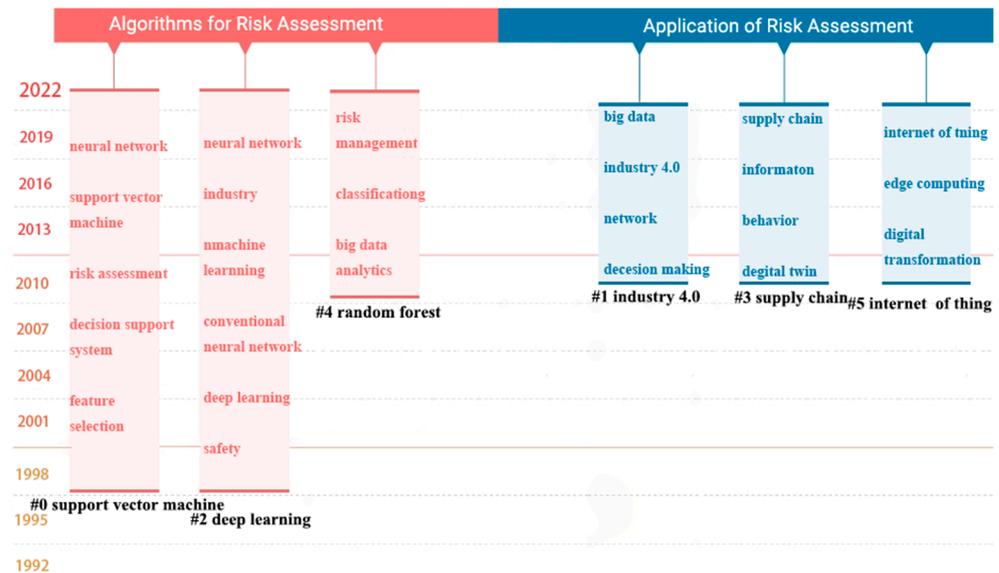


Figure 14. Keyword timeline clustering.

By analyzing the three stages of the theme using Bibliometrix R, the development of machine learning in the industrial field can be summarized as follows: from the initial use as a production assistant tool in a single field of industry to achieving real-time risk monitoring and assessment in multiple fields and then to realizing the informatization and digitization of Industry 4.0. Of these, the concept of Industry 4.0 is becoming increasingly important, emphasizing the use of advanced technologies to improve industrial processes, which is a prominent theme in this field. The necessity of continuing research on the application of machine learning in industrial environments is emphasized by these developments, as well as its potential to support broader industrial transformation.

Based on the evolution of time and topics, we have identified three current hotspots of machine learning in industrial risk assessment research: Firstly, the research on machine learning and deep learning algorithms themselves. Secondly, machine learning risk management in the Industry 4.0 system. Thirdly, the application of machine learning in the field of autonomous driving technology. In the era of the Internet of Things and big data, it is necessary to continuously explore machine learning algorithms, integrating various sensor information at the microlevel and integrating information from various fields at the macrolevel, achieving the informatization and digitization of the Industry 4.0 model. Additionally, researchers also need to use machine learning and deep learning to address more complex security issues, of which autonomous driving technology is one application.

The future research direction for machine learning applied to industrial risk assessment is expected to focus on several key areas. Firstly, further research on machine learning and deep learning algorithms themselves is anticipated, with the aim of developing more advanced and efficient algorithms for risk assessment in industrial settings. This could involve exploring new machine learning techniques, optimizing existing algorithms, and integrating different approaches to improve the accuracy and reliability of risk assessment models. Secondly, machine learning risk management in the context of Industry 4.0 is likely

to be a significant research area given the increasing emphasis on the use of advanced technologies to improve industrial processes. This could involve developing machine learning models that can effectively manage risks in complex and dynamic industrial environments, where data from various sources and sensors are integrated to enable real-time risk monitoring and assessment. Additionally, the application of machine learning in the field of autonomous driving technology is expected to be another important research direction. As autonomous vehicles become more prevalent in industrial settings, machine learning algorithms can play a crucial role in enabling these vehicles to assess and manage risks in real-time, ensuring safe and efficient operations. Overall, the future research direction for machine learning in industrial risk assessment is expected to focus on advancing algorithms, integrating technologies in the context of Industry 4.0, and addressing complex security issues in emerging applications such as autonomous driving technology.

However, there are several challenges faced in using machine learning in industrial risk assessment, although it has the potential to improve accuracy and efficiency. The reliability of machine learning models can be affected by limited data availability and quality, algorithmic and data bias, and a lack of interpretability. To overcome these challenges, issues related to data quality and bias need to be addressed and the interpretability of machine learning models needs to be ensured while human expertise is incorporated into the risk assessment process. More robust and transparent machine learning models that can be validated and interpreted by human experts should be developed in future research [87]. The interpretability challenge may be addressed by recent advancements in explainable AI techniques. The limitations of data availability and quality can be overcome by efforts to improve data collection and sharing [88]. The future of machine learning in industrial risk assessment looks promising, but overcoming the challenges and limitations associated with its use requires continued research and development.

5. Conclusions

Using bibliometric mapping analysis, the research on the application of machine learning in industrial risk assessment was reviewed in this paper, with a focus on time distribution, highly cited literature, the research knowledge base, the evolutionary path, research hotspots, and frontier areas. Based on this analysis, three main conclusions are drawn by the paper.

The research history of machine learning applied to industrial risk assessment is broadly divided into three phases: the initial exploration phase (1991–2006), the stable development phase (2006–2017), and the high development phase (2017–present). The application of machine learning in industrial risk assessment research is increasing year by year, and the number of publications is rising. The years of publication in European and North American countries are significantly earlier than those in Asian and African countries. The highest number of publications are in China, the US, and the UK, the three countries with the highest intensity of collaboration. The highest number of publications and author collaborations are from Tsinghua University and Li, Heng, respectively, and IEEE Access, the journal most cited and published within, is the primary carrier of the literature in this research area.

Based on the citation relationships in the literature, the application of machine learning to industrial risk assessment is a multidisciplinary research field that requires a foundation in mathematics and computer science. It also necessitates the integration of knowledge from various disciplines, such as chemistry, materials science, physics, environmental science, nutrition, and toxicology, depending on the application area. The key technology in this field is the monitoring and diagnosis of process failures [26]. The knowledge base in the field of applying machine learning to industrial risk assessment is machine learning algorithm design, applications in biomedicine, risk monitoring in construction and machinery, and environmental protection.

Currently, three hotspots have been formed in the industrial field by machine learning research: the study of machine learning and deep learning algorithms themselves, the

risk management of machine learning in the Industry 4.0 system, and the use of machine learning in the direction of autonomous driving technology. The four research frontiers are “Random Forests”, “Industry 4.0”, “supply chain risk assessment”, and the “Internet of Things”. The trend in research content is for the application of machine learning in industry to range from a single production aid to risk assessment in several areas to the informatization and digitization of Industry 4.0 systems.

Author Contributions: Conceptualization, H.L., Z.W. and X.T.; methodology, H.L. and Z.W.; software, Z.W. and K.P.; validation, Z.W., K.P., R.H., W.J. and J.W.; investigation, H.L., Z.W., X.T., K.P., R.H., W.J. and J.W.; writing—original draft preparation, Z.W. and H.L.; writing—H.L. and Z.W.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China (No. LY22E040001) and the Fundamental Research Funds for the Provincial Universities of Zhejiang (Nos. 2021YW92 and 2022YW92).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Dunjó, J.; Fthenakis, V.; Vílchez, J.A.; Arnaldos, J. Hazard and operability (HAZOP) analysis. A literature review. *J. Hazard. Mater.* **2010**, *173*, 19–32. [[CrossRef](#)] [[PubMed](#)]
- Vesely, W.E.; Goldberg, F.F.; Roberts, N.H.; Haasl, D.F. *Fault Tree Handbook*; Nuclear Regulatory Commission: Washington, DC, USA, 1981.
- Smith, D.J.; Simpson, K.G. *The Safety Critical Systems Handbook: A Straightforward Guide to Functional Safety: IEC 61508 (2010 Edition), IEC 61511 (2015 Edition) and Related Guidance*; Butterworth-Heinemann: Oxford, UK, 2020.
- Briand, L.C.; Basili, V.R.; Thomas, W.M. *A Pattern Recognition Approach for Software Engineering Data Analysis*; IEEE Transactions on Software Engineering: New York, NY, USA, 1991.
- Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [[CrossRef](#)]
- Xiong, H.; Zhang, Z.; Sun, X.; Yin, Z.-Y.; Chen, X. Clogging effect of fines in seepage erosion by using CFD–DEM. *Comput. Geotech.* **2022**, *152*, 105013. [[CrossRef](#)]
- Xiong, H.; Qiu, Y.; Lin, X.-T.; Chen, X.; Huang, D. Multiple arching in cohesion–friction soils: Insights from deformation behavior and failure mechanisms using FEM–SPH approach. *Comput. Geotech.* **2023**, *154*, 105146. [[CrossRef](#)]
- Laure, W.; Van Calster Ben, C.G.S.; Riley Richard, D.; Georg, H.; Ewoud, S. Prediction models for diagnosis and prognosis of COVID-19: Living systematic review and critical appraisal. *BMJ* **2020**, *369*, m1328.
- Ellis, D.I.; Broadhurst, D.; Kell, D.B.; Rowland, J.J.; Goodacre, R. Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. *Appl. Environ. Microbiol.* **2002**, *68*, 2822–2828. [[CrossRef](#)] [[PubMed](#)]
- Banerjee, P.; Eckert, A.O.; Schrey, A.K.; Preissner, R. ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* **2018**, *46*, W257–W263. [[CrossRef](#)] [[PubMed](#)]
- Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [[CrossRef](#)]
- Hsu, C.-H.; He, X.; Zhang, T.-Y.; Chang, A.-Y.; Liu, W.-L.; Lin, Z.-Q. Enhancing Supply Chain Agility with Industry 4.0 Enablers to Mitigate Ripple Effects Based on Integrated QFD–MCDM: An Empirical Study of New Energy Materials Manufacturers. *Mathematics* **2022**, *10*, 1635. [[CrossRef](#)]
- Canizo, M.; Onieva, E.; Conde, A.; Charramendieta, S.; Trujillo, S. Real-time predictive maintenance for wind turbines using Big Data frameworks. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; IEEE: New York, NY, USA, 2017; pp. 70–77.
- Song, H.M.; Woo, J.; Kim, H.K. In-vehicle network intrusion detection using deep convolutional neural network. *Veh. Commun.* **2020**, *21*, 100198. [[CrossRef](#)]
- Lu, X.; Deng, J.; Xiao, Y.; Zhai, X.; Wang, C.; Yi, X. Recent progress and perspective on thermal-kinetic, heat and mass transportation of coal spontaneous combustion hazard. *Fuel* **2022**, *308*, 121234. [[CrossRef](#)]
- Priyanka, E.; Thangavel, S.; Gao, X.-Z.; Sivakumar, N. Digital twin for oil pipeline risk estimation using prognostic and machine learning techniques. *J. Ind. Inf. Integr.* **2022**, *26*, 100272. [[CrossRef](#)]

17. Lang, Z.; Wang, D.; Liu, H.; Gou, X. Mapping the knowledge domains of research on corrosion of petrochemical equipment: An informetrics analysis-based study. *Eng. Fail. Anal.* **2021**, *129*, 105716. [[CrossRef](#)]
18. Liu, H.; Song, X.; Zhang, F. Fault diagnosis of new energy vehicles based on improved machine learning. *Soft Comput.* **2021**, *25*, 12091–12106. [[CrossRef](#)]
19. Noort, D.; McCarthy, P. The critical path to automated underground mining. In Proceedings of the First International Future Mining Conference, Sydney, Australia, 19–21 November 2008; pp. 179–182.
20. Sarkar, S.; Maiti, J. Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Saf. Sci.* **2020**, *131*, 104900. [[CrossRef](#)]
21. Chen, K.; Lin, X.; Wang, H.; Qiang, Y.; Kong, J.; Huang, R.; Wang, H.; Liu, H. Visualizing the Knowledge Base and Research Hotspot of Public Health Emergency Management: A Science Mapping Analysis-Based Study. *Sustainability* **2022**, *14*, 7389. [[CrossRef](#)]
22. Liu, H.; Chen, H.; Hong, R.; Liu, H.; You, W. Mapping knowledge structure and research trends of emergency evacuation studies. *Saf. Sci.* **2020**, *121*, 348–361. [[CrossRef](#)]
23. Li, J.; Goerlandt, F.; Reniers, G. An overview of scientometric mapping for the safety science community: Methods, tools, and framework. *Saf. Sci.* **2021**, *134*, 105093. [[CrossRef](#)]
24. Gou, X.; Liu, H.; Qiang, Y.; Lang, Z.; Wang, H.; Ye, D.; Wang, Z.; Wang, H. In-depth analysis on safety and security research based on system dynamics: A bibliometric mapping approach-based study. *Saf. Sci.* **2022**, *147*, 105617. [[CrossRef](#)]
25. Hong, R.; Liu, H.; Xiang, C.; Song, Y.; Lv, C. Visualization and analysis of mapping knowledge domain of oxidation studies of sulfide ores. *Environ. Sci. Pollut. Res.* **2020**, *27*, 5809–5824. [[CrossRef](#)]
26. Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [[CrossRef](#)]
27. Moselhi, O.; Hegazy, T.; Fazio, P. Neural networks as tools in construction. *J. Constr. Eng. Manag.* **1991**, *117*, 606–625. [[CrossRef](#)]
28. Jenkinson, J.; Shaw, R.; Andow, P. Operator support systems and artificial intelligence. *Reliab. Eng. Syst. Saf.* **1991**, *33*, 419–437. [[CrossRef](#)]
29. Islam, S.R.; Kwak, D.; Kabir, M.H.; Hossain, M.; Kwak, K.-S. The internet of things for health care: A comprehensive survey. *IEEE Access* **2015**, *3*, 678–708. [[CrossRef](#)]
30. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
31. Flores, M.; Glusman, G.; Brogaard, K.; Price, N.D.; Hood, L. P4 medicine: How systems medicine will transform the healthcare sector and society. *Pers. Med.* **2013**, *10*, 565–576. [[CrossRef](#)]
32. King, T.; Cole, M.; Farber, J.M.; Eisenbrand, G.; Zabarar, D.; Fox, E.M.; Hill, J.P. Food safety for food security: Relationship between global megatrends and developments in food safety. *Trends Food Sci. Technol.* **2017**, *68*, 160–175. [[CrossRef](#)]
33. Koyuncugil, A.S.; Ozgulbas, N. Financial early warning system model and data mining application for risk detection. *Expert Syst. Appl.* **2012**, *39*, 6238–6253. [[CrossRef](#)]
34. Fernandez, M.G.; Tokuhira, A.; Welter, K.; Wu, Q. Nuclear energy system's behavior and decision making using machine learning. *Nucl. Eng. Des.* **2017**, *324*, 27–34. [[CrossRef](#)]
35. Lang, Z.; Liu, H.; Meng, N.; Wang, H.; Wang, H.; Kong, F. Mapping the knowledge domains of research on fire safety—An informetrics analysis. *Tunn. Undergr. Space Technol.* **2021**, *108*, 103676. [[CrossRef](#)]
36. Goh, Y.M.; Chua, D. Case-based reasoning approach to construction safety hazard identification: Adaptation and utilization. *J. Constr. Eng. Manag.* **2010**, *136*, 170–178. [[CrossRef](#)]
37. Mortimore, R. *Making Sense of Chalk: A Total-Rock Approach to Its Engineering Geology*; Geological Society of London: London, UK, 2012.
38. Hussain, R.; Rezaeifar, Z.; Lee, Y.-H.; Oh, H. Secure and privacy-aware traffic information as a service in VANET-based clouds. *Pervasive Mob. Comput.* **2015**, *24*, 194–209. [[CrossRef](#)]
39. Fernández-Rodríguez, J.Y.; Álvarez-García, J.A.; Fisteus, J.A.; Luaces, M.R.; Magaña, V.C. Benchmarking real-time vehicle data streaming models for a smart city. *Inf. Syst.* **2017**, *72*, 62–76. [[CrossRef](#)]
40. Wang, H.; Liu, H.; Yao, J.; Ye, D.; Lang, Z.; Glowacz, A. Mapping the knowledge domains of new energy vehicle safety: Informetrics analysis-based studies. *J. Energy Storage* **2021**, *35*, 102275. [[CrossRef](#)]
41. Tan, K.H.; Ortiz-Gallardo, V.G.; Perrons, R.K. Using Big Data to manage safety-related risk in the upstream oil & gas industry: A research agenda. *Energy Explor. Exploit.* **2016**, *34*, 282–289.
42. Layouni, M.; Tahar, S.; Hamdi, M.S. A survey on the application of neural networks in the safety assessment of oil and gas pipelines. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES), Orlando, FL, USA, 9–12 December 2014; IEEE: New York, NY, USA, 2014; pp. 95–102.
43. Stamatescu, I.; Stamatescu, G.; Fagarasan, I.; Arghira, N.; Calofir, V.; Iliescu, S.S. ASID: Advanced system for process control towards intelligent specialization in the power engineering field. In Proceedings of the 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, Romania, 21–23 September 2017; IEEE: New York, NY, USA, 2017; pp. 475–480.
44. Kamble, S.S.; Gunasekaran, A.; Gawankar, S.A. Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. *Process Saf. Environ. Prot.* **2018**, *117*, 408–425. [[CrossRef](#)]

45. Diez-Olivan, A.; Del Ser, J.; Galar, D.; Sierra, B. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Inf. Fusion* **2019**, *50*, 92–111. [[CrossRef](#)]
46. Moktadir, M.A.; Ali, S.M.; Kusi-Sarpong, S.; Shaikh, M.A.A. Assessing challenges for implementing Industry 4.0: Implications for process safety and environmental protection. *Process Saf. Environ. Prot.* **2018**, *117*, 730–741. [[CrossRef](#)]
47. Yang, Y.; Reniers, G.; Chen, G.; Goerlandt, F. A bibliometric review of laboratory safety in universities. *Saf. Sci.* **2019**, *120*, 14–24. [[CrossRef](#)]
48. Liu, H.; Xie, Y.; Liu, Y.; Nie, R.; Li, X. Mapping the knowledge structure and research evolution of urban rail transit safety studies. *IEEE Access* **2019**, *7*, 186437–186455. [[CrossRef](#)]
49. Shi, Y.; Xue, X.; Xue, J.; Qu, Y. Fault Detection in Nuclear Power Plants using Deep Learning based Image Classification with Imaged Time-series Data. *Int. J. Comput. Commun. Control.* **2022**, *17*. [[CrossRef](#)]
50. Shi, Y.; Xue, X.; Qu, Y.; Xue, J.; Zhang, L. Machine Learning and Deep Learning Methods used in Safety Management of Nuclear Power Plants: A Survey. In Proceedings of the 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 7–10 December 2021; IEEE: New York, NY, USA, 2021; pp. 917–924.
51. Yao, Y.; Wang, J.; Long, P.; Xie, M.; Wang, J. Small-batch-size convolutional neural network based fault diagnosis system for nuclear energy production safety with big-data environment. *Int. J. Energy Res.* **2020**, *44*, 5841–5855. [[CrossRef](#)]
52. Zeng, L.; Li, R.Y.M. Construction safety and health hazard awareness in Web of Science and Weibo between 1991 and 2021. *Saf. Sci.* **2022**, *152*, 105790. [[CrossRef](#)]
53. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Rose, T.M.; An, W.; Yu, Y. A deep learning-based method for detecting non-certified work on construction sites. *Adv. Eng. Inform.* **2018**, *35*, 56–68. [[CrossRef](#)]
54. Yu, Y.; Li, H.; Yang, X.; Kong, L.; Luo, X.; Wong, A.Y. An automatic and non-invasive physical fatigue assessment method for construction workers. *Autom. Constr.* **2019**, *103*, 1–12. [[CrossRef](#)]
55. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 104440. [[CrossRef](#)]
56. Arashpour, M.; Ngo, T.; Li, H. Scene understanding in construction and buildings using image processing methods: A comprehensive review and a case study. *J. Build. Eng.* **2021**, *33*, 101672. [[CrossRef](#)]
57. Liu, H.; Hong, R.; Xiang, C.; Lv, C.; Li, H. Visualization and analysis of mapping knowledge domains for spontaneous combustion studies. *Fuel* **2020**, *262*, 116598. [[CrossRef](#)]
58. Ivanov, D.; Dolgui, A.; Sokolov, B. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *Int. J. Prod. Res.* **2019**, *57*, 829–846. [[CrossRef](#)]
59. Ivanov, D.; Dolgui, A. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Prod. Plan. Control* **2021**, *32*, 775–788. [[CrossRef](#)]
60. Mittal, S.; Khan, M.A.; Romero, D.; Wuest, T. A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). *J. Manuf. Syst.* **2018**, *49*, 194–214.
61. Xu, H.; Yu, W.; Griffith, D.; Golmie, N. A survey on industrial Internet of Things: A cyber-physical systems perspective. *IEEE Access* **2018**, *6*, 78238–78259. [[CrossRef](#)]
62. Mascitelli, R. From experience: Harnessing tacit knowledge to achieve breakthrough innovation. *J. Prod. Innov. Manag. Int. Publ. Prod. Dev. Manag. Assoc.* **2000**, *17*, 179–193. [[CrossRef](#)]
63. Fihn, S.D.; Francis, J.; Clancy, C.; Nielson, C.; Nelson, K.; Rumsfeld, J.; Cullen, T.; Bates, J.; Graham, G.L. Insights from advanced analytics at the Veterans Health Administration. *Health Aff.* **2014**, *33*, 1203–1211. [[CrossRef](#)] [[PubMed](#)]
64. Goel, V.K.; Panjabi, M.M.; Patwardhan, A.G.; Dooris, A.P.; Serhan, H. Test protocols for evaluation of spinal implants. *JBJS* **2006**, *88* (Suppl. S2), 103–109.
65. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Autom. Constr.* **2020**, *114*, 103179. [[CrossRef](#)]
66. Yan, J.; Meng, Y.; Lu, L.; Li, L. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. *IEEE Access* **2017**, *5*, 23484–23491. [[CrossRef](#)]
67. Feindt, M.; Kerzel, U. The NeuroBayes neural network package. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2006**, *559*, 190–194. [[CrossRef](#)]
68. Kogevinas, M.; Mannelje, A.T.; Cordier, S.; Ranft, U.; González, C.A.; Vineis, P.; Chang-Claude, J.; Lynge, E.; Wahrendorf, J.; Tzonou, A. Occupation and bladder cancer among men in Western Europe. *Cancer Causes Control* **2003**, *14*, 907–914. [[CrossRef](#)]
69. Chen, C.; Leydesdorff, L. Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 334–351. [[CrossRef](#)]
70. Li, X.; Zhang, W.; Ding, Q.; Sun, J.-Q. Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *J. Intell. Manuf.* **2020**, *31*, 433–452. [[CrossRef](#)]
71. Heimberger, M.; Horgan, J.; Hughes, C.; McDonald, J.; Yogamani, S. Computer vision in automated parking systems: Design, implementation and challenges. *Image Vis. Comput.* **2017**, *68*, 88–101. [[CrossRef](#)]
72. Xu, Q.; Chong, H.-Y.; Liao, P.-C. Exploring eye-tracking searching strategies for construction hazard recognition in a laboratory scene. *Saf. Sci.* **2019**, *120*, 824–832. [[CrossRef](#)]

73. Glowacz, A.; Tadeusiewicz, R.; Legutko, S.; Caesarendra, W.; Irfan, M.; Liu, H.; Brumerick, F.; Gutten, M.; Sulowicz, M.; Daviu, J.A.A. Fault diagnosis of angle grinders and electric impact drills using acoustic signals. *Appl. Acoust.* **2021**, *179*, 108070. [[CrossRef](#)]
74. Kibria, H.B.; Matin, A. The severity prediction of the binary and multi-class cardiovascular disease—A machine learning-based fusion approach. *Comput. Biol. Chem.* **2022**, *98*, 107672. [[CrossRef](#)] [[PubMed](#)]
75. Liu, T.; Yu, Z. The analysis of financial market risk based on machine learning and particle swarm optimization algorithm. *EURASIP J. Wirel. Commun. Netw.* **2022**, *2022*, 1–17. [[CrossRef](#)]
76. Kaewunruen, S.; Sresakoolchai, J.; Huang, J.; Zhu, Y.; Ngamkhanong, C.; Remennikov, A.M. Machine Learning Based Design of Railway Prestressed Concrete Sleepers. *Appl. Sci.* **2022**, *12*, 10311. [[CrossRef](#)]
77. Islam, F.B.; Lee, J.M.; Kim, D.S. Smart factory floor safety monitoring using UWB sensor. *IET Sci. Meas. Technol.* **2022**, *16*, 412–425. [[CrossRef](#)]
78. Karun, B.; VR, R.; Elayidom, S. Application of fuzzy logic and machine learning techniques to improve inherently safer design in process safety management: A brief study. *Process Saf. Prog.* **2022**, *41*, S178–S186. [[CrossRef](#)]
79. Goel, P.; Datta, A.; Mannan, M.S. Application of big data analytics in process safety and risk management. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; IEEE: New York, NY, USA, 2017; pp. 1143–1152.
80. Lugano, G. Virtual assistants and self-driving cars. In Proceedings of the 2017 15th International Conference on ITS Telecommunications (ITST), Warsaw, Poland, 29–31 May 2017; IEEE: New York, NY, USA, 2017; pp. 1–5.
81. An, D.; Liu, J.; Zhang, M.; Chen, X.; Chen, M.; Sun, H. Uncertainty modeling and runtime verification for autonomous vehicles driving control: A machine learning-based approach. *J. Syst. Softw.* **2020**, *167*, 110617. [[CrossRef](#)]
82. Paes, V.M.; Silveira, F.F.; Akkari, A.C.S. Social Impacts of Artificial Intelligence and Mitigation Recommendations: An Exploratory Study. In Proceedings of the 7th Brazilian Technology Symposium (BTSym'21) Emerging Trends in Human Smart and Sustainable Future of Cities, Campinas, Brazil, 19–21 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 521–528.
83. Liao, M.; Lan, K.; Yao, Y. Sustainability implications of artificial intelligence in the chemical industry: A conceptual framework. *J. Ind. Ecol.* **2022**, *26*, 164–182. [[CrossRef](#)]
84. Santana, J.A.D.; Arana, Y.C.; Gomez, O.G.; Furka, D.; Furka, S.; Orozco, J.L.; Di Benedetto, A.; Russo, D.; Portarapillo, M.; Febles, J.S. Fire and Explosion Economic Losses (FEEL) Index: A new approach for quantifying economic damages due to accidents in hydrocarbon storage sites. *Process Saf. Environ. Prot.* **2022**, *165*, 77–92. [[CrossRef](#)]
85. Callon, M.; Courtial, J.P.; Laville, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics* **1991**, *22*, 155–205. [[CrossRef](#)]
86. Cobo, M.J.; López-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.* **2011**, *62*, 1382–1402. [[CrossRef](#)]
87. Hegde, J.; Rokseth, B. Applications of machine learning methods for engineering risk assessment—A review. *Saf. Sci.* **2020**, *122*, 104492. [[CrossRef](#)]
88. Paltrinieri, N.; Comfort, L.; Reniers, G. Learning about risk: Machine learning for risk assessment. *Saf. Sci.* **2019**, *118*, 475–486. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.