International Journal of Information Science and Management
Vol. 21, No. 1, 2023, 17-35
DOI: 10.22034/ijism.2022.1977763.0 / https://dorl.net/dor/20.1001.1.20088302.2023.21.1.2.2

Original Research

Machine Learning in Safety and Health Research: A Scientometric Analysis

Khairul Hafezad Abdullah
Senior Science Officer, Universiti Teknologi MARA, Perlis Branch, Perlis, Malaysia.
Corresponding Author: ezadneo88@gmail.com
ORCID iD: https://orcid.org/0000-0003-3759-6541

Davi Sofyan
Lecturer, Department of Physical Education, Universitas Majalengka, West Java, Indonesia.
davisofyan@unma.ac.id
ORCID iD: https://orcid.org/0000-0001-9510-7123

Received: 22 March 2022
Accepted: 20 April 2022

Abstract

Safety and health are intricately interwoven and have become indispensable to the thriving business world and anthropology. It is concerned with ensuring employees' physical, emotional, and mental well-being. Based on the Scopus and Web of Science databases, the current study intends to analyse the global research output on machine learning in safety and health. This study utilized ScientoPy and VOSviewer to delve into the annual growth, patterns of research communication on source titles, international collaboration among countries, and authors' keyword analysis. This study found that the Web of Science database tracks the evolution of publications throughout time. PLoS One has surpassed all other source titles in terms of publishing activity. Also, this study indicated that US researchers are constantly working on machine learning in safety and health research and have developed significant collaborations with China and Australia. Between 2020 and 2021, the University of Toronto published 86% of all papers, outpacing other institutions. The keywords "machine learning", "artificial intelligence", "electronic health records", "deep learning", and "mental health" were the most popular and trending keywords in 2020 and 2021, and "artificial intelligence" appeared in most publications among others. Future researchers should conduct scoping or systematic literature reviews to elucidate the relationships between these terms. This study may entice the curiosity of practitioners and researchers to advance new knowledge in this field by being devoted to cutting-edge research in the contemporary philosophy of science, cognitive, and cultural anthropology on machine learning in safety and health research. In conclusion, this scientometric analysis demonstrates that machine learning in safety and health is a study domain that requires further refinement in future research, as this technology has the potential to significantly improve workplace safety and health through targeted applications with clear benefits.

Keywords: Machine Learning, Safety, Health, Scientometric, Scopus, Web of Science, Publication Trajectories.

Introduction

Based on an anthropological standpoint, safety and health in any business are becoming an apparent priority increasingly. One strength of the interpretive approach relevant to this context is that the researcher provides an account of other people's beliefs and actions; the exercise is effectuated with the consideration that the interpretation is motivated by historical and cultural causes (Sachs,1990). Safety specialists agree that organizations' emphasis on wellness extends

to employee safety in and out of the business operations; it's about ensuring employees' physical, emotional, and mental well-being (Abdullah, Hashim & Abd Aziz, 2020; Van Nunen, Li, Reniers & Ponnet, 2018). Likewise, managing safety and health in organizations relies heavily on preventing accidents, illnesses, and diseases (Abdullah & Abd Aziz, 2020). It is vital since managing safety and health has become an integral part of business operations (Krstić, Rađenović & Živković, 2022). Companies must undertake risk assessments to determine hazards and threats, simultaneously establishing measures to prevent worker injuries (Gul & Ak, 2018), fatalities, or even workplace mishaps that could lead to trauma.

However, although many companies have implemented safety and health programs, accidents or other risk elements still prevail. The reality that accidents often occur and occasionally fatally at factories, building sites, warehouses, and other workplaces cannot be concealed (Bas & Koseoglu, 2019; Heo et al., 2018; Shao, Hu, Liu, Chen & He, 2019). Hence, the best strategy for identifying high-risk workplaces is enforcing safety-leading indicators. However, there is a shortage of verified leading indicators capable of accurately classifying the safety risk level (Poh, Ubeynarayana & Goh, 2018). This development is beginning to transform the safety and health management paradigm by appealing to artificial intelligence (AI), the Internet of Things (IoT), and intelligent devices as the industrial 4.0 era unfolds (Badri, Boudreau-Trudel & Souissi, 2018).

Machine learning (ML), a subset of AI, has been an effective predictive tool due to its extensive applications across multiple industries and disciplines (Yang, Fidelis & Sun, 2019). ML has become a technology used to augment and enhance physicians' cognitive capabilities in the health industry as they provide treatment to increasingly complex patients (Bini, 2018). In safety-critical industries such as the petroleum and chemical industries, ML has been utilized to assess a drive-off scenario to forecast risk intensification or reduction when system circumstances change (Paltrinieri, Comfort & Reniers, 2019). The capacity of networks and computing devices to collect, store, and transport massive amounts of data has increased dramatically over the last decade, resulting in "Big Data" (Naeem et al., 2022).

In terms of data mining and analysis, AI is permeating and demonstrating remarkable capabilities in a wide range of fields of science and technology for developing practical algorithms and software for computer vision, language processing, and image recognition, among other applications (Ashok, Madan, Joha & Sivarajah, 2022). ML is a broad category of algorithms and modeling tools used for numerous data processing tasks to replicate human intelligence by observing its surroundings (El Naqa & Murphy, 2015). ML is one of the most powerful strategies in artificial intelligence, as it employs algorithms to learn from data (Jordan & Mitchell, 2015). While all ML techniques fit models to data, the specific methods are pretty varied and can appear bewildering at first glance (Greener, Kandathil, Moffat & Jones, 2022). As algorithms, the data used to train them and the models they generate become more powerful and ingrained in society; thus, safety concerns must be discoursed (Varshney, 2016).

ML algorithms aim to create a statistical model that can be used to make predictions, classifications, and estimations. For instance, researchers have used ML algorithms to treat cancer, provide supportive care, and alleviate shoulder discomfort and respiratory diseases (Trung, Huy & Le, 2021). It is designated that ML methods for extracting health indicators from events and transient data can be exploited as input parameters for predictive models to forecast future health circumstances (Sirola & Hulsund, 2021). Cancer prediction is a far cry from risk-based inspection object targeting. However, both are examples of predictive

challenges. (Liakos, Busato, Moshou, Pearson & Bochtis, 2018) clarified further that ML methods typically begin with a learning process in which the goal is to improve performance by learning from "experience" (training data). They add that a collection of attributes frequently describes a particular example, also referred to as a feature or variable. ML tasks are typically classified into broad categories based on the type of learning, either supervised or unsupervised, the learning models used to accomplish the task, or the learning models used to achieve the mission (Liakos et al., 2018).

Numerous studies have established the utility of ML approaches in safety and health research. The studies include those undertaken by Bonifazi, Corradini, Ursino, Virgili, Anceschi and De Donato (2021), Koklonis, Sarafidis, Vastardi and Koutsouris (2021), Simsekler, Rodrigues, Qazi, Ellahham & Ozonoff (2021), Tang and Golparvar-Fard (2021), Yong, Xiaoming and Alshehri (2021). Bonifazi et al. (2021) developed a new way to improve safety using Sentient Multimedia Systems and Machine Learning. These systems can detect falls, send out alarms, and coordinate rescue operations. Koklonis et al. (2021) developed an ML-based Decision Support Systems model for predicting workplace accidents. This model is based on hospital workplace incident and accident data collected by personnel and safety engineers. Preliminary findings suggest that the proposed system can help predict accidents and evaluate the effectiveness of safety and health interventions.

Simsekler et al. (2021) investigated the link between organizational factors and patient and staff safety errors. This study was the first to use ML algorithms to compare patient and staff safety drivers in this specific context. Tang and Golparvar-Fard (2021) revealed that an ML-based severity prediction model could predict the severity level based on a worker's state. An ablation study demonstrates that the complete worker state is more predictive of severity level than any worker state based on partial visual information. These findings are corroborated by a newly introduced large image dataset that includes exhaustive annotations of single-worker activity, body pose, personal protective equipment, tools, and materials. Yong et al. (2021) applied ML to analyze and detect burst signals in an intelligent healthcare system to protect public health. They discovered that differential correlation detection has excellent frequency deviation robustness, constant false alarm, and detection threshold. These features make differential correlation detection ideal for burst detection. These results also show how to set the detection threshold to meet system performance requirements under various conditions. Based on previous studies, ML methods can be used to improve the implementation of adequate safety and health policies and preventive measures. Also, the healthcare industry has profited tremendously from technological improvements where tech-enabled medical devices have greatly aided society's progress by outstandingly enhancing the quality of diagnosis, treatment, and recovery of patients.

The field of ML encompasses various methodologies and tasks, some of which have readily available software solutions, while others need extensive research (Mukhamediev, Symagulov, Kuchin, Yakunin & Yelis, 2021). It would be fascinating to examine how academics' interests have evolved and, if feasible, focus on those research areas that are currently receiving more attention. Simultaneously, the number of papers in various ML domains continues to expand. As a result, merely stating the growth in publications is insufficient. In this context, scientometric measures such as the number of escalations and publications progress are frequently employed to gauge a researcher's activity. Previously published studies focused on research trends in ML in construction management (Van & Quoc, 2021), global ML research

(Dhawan, Gupta & Singh, 2020), and comparing human and ML performance (Goh, Cai, Theseira, Ko & Khor, 2020). In safety and health, prior research has examined the use of ML in occupational accident analysis (Sarkar & Maiti, 2020) and ML to aid in hospital safety and health decision-making (Jayatilake & Ganegoda, 2021). There have been relatively few scientometric studies on ML concerning safety and health domains. Thus, the scientometric review in this study focuses on examining this subject's publishing trajectory. This should be refined because various ML technologies are being utilized to identify and analyze risks connected to safety and health.

Scientometric review is an excellent method for getting a holistic picture of previous, present, and new research topics by visualizing research trends and progress, identifying linkages across research clusters, and identifying research needs for future research. The benefits of a scientometric review include providing evidence of the impact of research, forecasting future research directions based on identified research gaps, guiding funding allocation for recognized impactful research at both the local and international levels, and facilitating collaboration with productive research networks, institutions, and countries (Martinez, del Mar Delgado, Marin & Alvarez, 2019; Mingers & Leydesdorff, 2015). The number of publications containing specified key terms is employed as an indicator for conducting scientometric analysis in this work. To address the existing body of knowledge limitations, the current study exploited scientometric analysis to ascertain the influence of journals, authors, authors' keywords, countries, and institutions. It is vital to describe global trends of ML applications in safety and health research and propose new research directions in this domain.

Materials and Methods

This section describes the methodology used to collect the dataset for this review, including the pre-processing steps and the review methodology.

Dataset collection and initial analysis

On February 24, 2022, data collection was initiated from two bibliographic databases, Web of Science (WoS) and Scopus. The search string "machine learning" AND ("safety and health" OR "health and safety" OR "health" OR "safety") appertained to this study. The retrieval process was operated through the title search to browse and select relevant publications based on the title of a journal article, conference proceeding paper, book, or book chapter. The title-based dataset retrieval is critical for highlighting specific research topics' content-related and structural development (Tullu, 2019).

The downloaded dataset was then pre-processed with ScientoPy to examine the parameters governing the evolution of trending topics and publications. ScientoPy is a Python-based scientometric analysis tool that removes bias from individual publications for accurate analysis (Ruiz-Rosero, Ramírez-González & Viveros-Delgado, 2019). The collected data yielded 2132 publications grounded on both databases from January 1, 1996, until December 31, 2021. In 1996, there was a single article indexed in the Scopus database. The article was titled "Abductive machine learning for modeling and predicting the educational score in school health surveys", written by Abdel-Aal and Mangoud (1996).

Articles, proceeding papers, conference papers, review papers, and book chapters are this study's publications. Those types of publications were set to default by ScientoPy software

(Ruiz-Rosero et al., 2019). Likewise, it is worth noting that language was not restricted as a primary analysis parameter. One possible explanation is that a scientometric review seeks to assess the impact of publications rather than reading the entire document as a scoping or systematic literature review would.

Not all documents are found to rule out the possibility of duplicate files or documents. To work around this, ScintoPy can delete all duplicate documents (Sofyan & Abdullah, 2022). The data was then harmonized during pre-processing stage by replacing the comma in the author's name with a semicolon, removing dots, commas, and unusual accents from the author's name, and removing duplicated samples with identical titles and authors. This procedure increases the reliability and precision of the datasets (Ruiz-Rosero et al., 2019). The data was then listed together in the new dataset after pre-processing.

The final sample examined in this study was represented in Table 1, which included 1148 documents after removing duplicate data. It contained 741 WOS and 407 Scopus datasets. After duplication data was removed, more publications from WoS databases were discovered because the ScientoPy pre-processing script prioritizes WoS publications. Importantly, it is critical to note that, after removing duplicates, the total number of papers meets the minimum requirement of metric analysis with greater than 300 metadata (Donthu, Kumar, Mukherjee, Pandey & Lim, 2021).

Table 1
Information on initial data analysis

Data Retrieval	Number	Percentage (%)
Loaded papers	2132	
Omitted papers by document type	289	13.60
Total papers after omitted papers removed	1843	
Loaded papers from WoS	746	40.50
Loaded papers from Scopus	1097	59.50
Duplicated removal results:		
Duplicated papers found	695	37.700
Removed duplicated papers from WoS	5	0.70
Removed duplicated papers from Scopus	690	62.90
Duplicated documents with different cited by	447	64.30
Total papers after duplicate removal	1148	
Papers from WoS	741	64.50
Papers from Scopus	407	35.50

Results

The evolution of publications

Generally, publications benefit both researchers and the journal's host institutions. Publications of scientific works are shared with other researchers with similar research interests. This facilitates the advancement of knowledge and its application by drawing the attention of similar-minded researchers and practitioners to novel ideas. Figure 1 depicts the remarkable growth of ML in safety and health research from Wos and Scopus databases. Notably, the publication trajectory in WoS has increased significantly since 2016 compared to the Scopus

database. The discovery demonstrated that ML is a revolutionary tool that enables researchers in safety and health to have a vital necessity for additional investigation. Besides, this technology keeps people safer when hospitals, businesses, and the government rely on ML to mitigate risks and actively protect people from various dangers and hazards (Siegel, 2019).

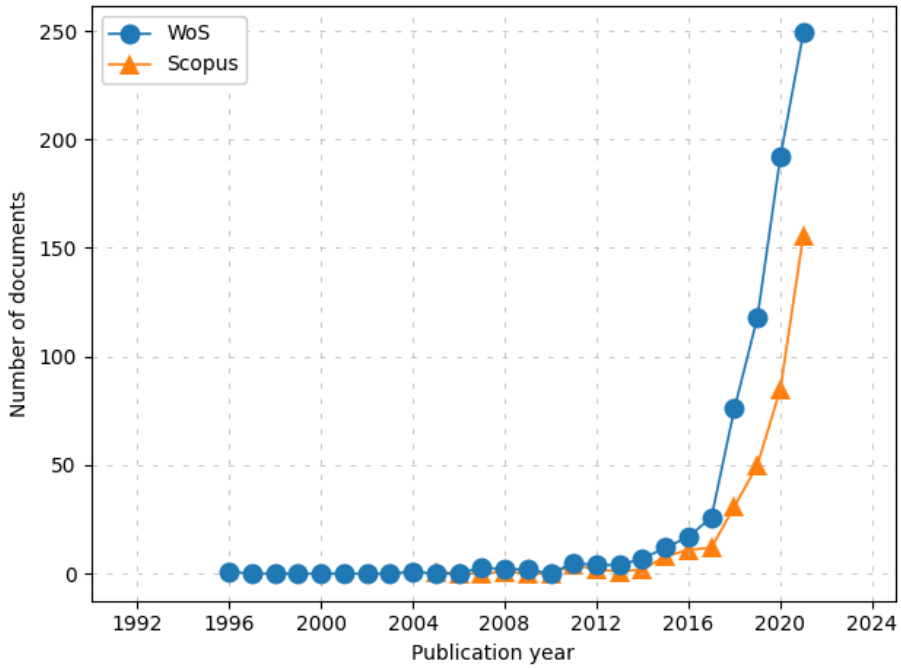

Figure 1. Database timeline

Descriptive category of analysis

The first descriptive category examined in this study was source titles. This classification showed 745 source titles that published documents concerning ML in safety and health. Table 2 lists the top ten source titles with the most articles published between 1996 and 2021. PLoS One published the most papers, with 26. The following source title, Journal of Medical Internet Research, contained 22 publications and placed at second rank, and the third-ranked, JMIR Medical Informatics, composed 18 publications.

Additionally, Table 2 illustrates the source titles with positive annual growth rates (AGR) in 2020 and 2021. The values suggest that the number of articles generated had a beneficial effect and rose gradually. This shows that the interest in publishing research on ML in safety and health, either in empirical or theoretical studies, has grown progressively. Hence, this information will assist readers, and future researchers obtain the most up-to-date information from these source titles to advance their current research using the most conversant evidence and pertinent issues.

Significantly too, PLoS One and Safety Science had a negative AGR during the same period; yet, these two source titles have the highest Hirsch index (h-index) with seven and above. The International Journal of Environmental Research and Public Health received the highest AGR in this study. A possible explanation is that the International Journal of Environmental Research and Public Health is interdisciplinary and publishes articles in environmental sciences and engineering, public health, occupational hygiene, health economics, and global health research. Thus, it has attracted previous scholars to publish their

research on ML in safety and health.

Table 2
The top ten source titles

Position	Source Title	Total	AGR	h-Index
1	PLoS One	23	-0.5	11
2	Journal of Medical Internet Research	21	4	5
3	JMIR Medical Informatics	20	5	7
4	Safety Science	19	-5.5	7
5	IEEE Access	14	0.5	8
6	International Journal of Environmental Research and Public Health	13	4.5	2
7	JAMA Network Open	10	0.5	6
8	Applied Sciences-Basel	9	1	4
9	Journal of The American Medical Informatics Association	9	-0.5	6
10	Journal of Physics: Conference Series	9	3.5	1

The second descriptive category examined in this study was country. Overall, 99 countries actively publish articles on ML in safety and health. Figure 2 depicts the top ten countries with at least 28 publications concerning the number of articles published. American scholars published over 300 articles on ML in safety and health. In this study, India became the second-largest publishing country, with 149 publications, and China was ranked third with 116 publications. These countries formed the nucleus of scientific production related to ML in safety and health research, with document production steadily increasing since 2008.

This study also shows that the United States has the highest average number of documents per year (ADY) at 108, followed by India with 19 and China with 39.5. Based on these findings, scholars from various countries have made strenuous efforts to publish more research on ML in safety and health, demonstrating the critical nature of identifying and specifying technological aspects that contribute to global safety and health. It was also vital to provide the appropriate and best technology-adopted mechanism to enhance safety and health management at the workplace to avoid injuries, loss of life, and damage or loss of property.

This study will also investigate co-authorship analysis because of the importance of research collaborations in most scientific fields. Co-authorship analysis by country was used in this study to show how multiple authors have worked together and collaborated. In the scientific community, co-authorship is a term used to describe a scientific collaboration between numerous authors to accomplish specific research goals. Researchers are also motivated to work together because it allows them to discover new knowledge and develop solutions to research problems. There is no specific number to set the threshold for conducting co-authorship analysis by country. Yet, this study applied the minimum number of publications of a country which is 20, and 0 minimum number of citations through VOSviewer software. As a result, 14 items met the threshold of 99 countries within the dataset. Figure 3 illustrates countries' co-authorship in research on ML in safety and health globally. Based on Figure 3, four clusters are signified with various colors of nodes (red, blue, green, and yellow). The size of nodes represents the number of documents per country, and the thickness of lines depicts the strength of collaboration.

Hence, this study indicated that the collaboration among researchers from the United States, China, and Australia was more favorable than in other countries.

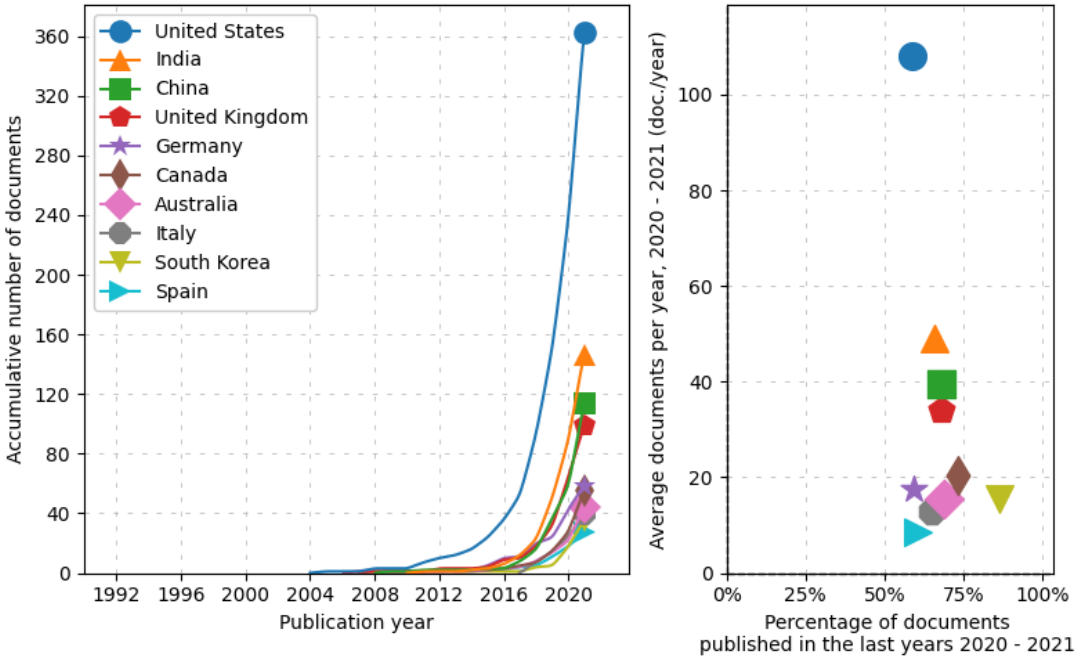

Figure 2. Ranking the most productive countries

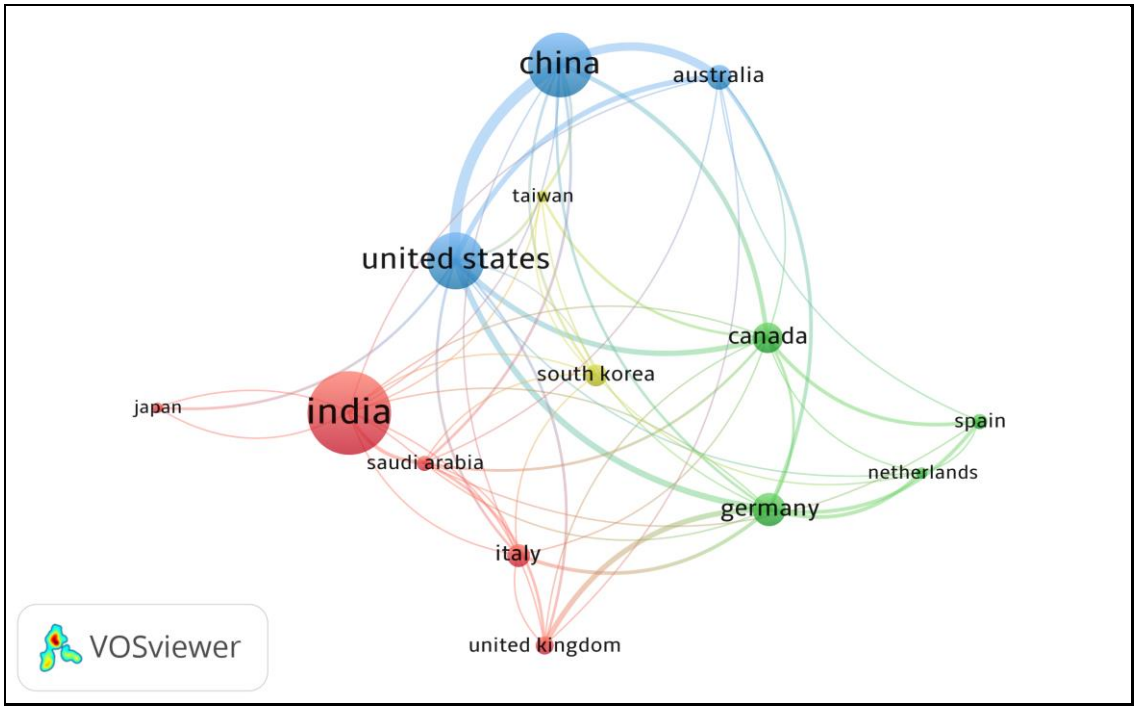

Figure 3. Network visualization of co-authorship by country

Meanwhile, articles on ML in safety and health have been published extensively by 1263 institutions. The ten institutions with the most articles published are depicted in Figure 4. The Harvard Medical School of Harvard University and Stanford University in the United States contributed the most articles published in this research perspective, totaling 19. However, the

total number of publications between 2020 and 2021 published by the Harvard Medical School is 63%, which stands out from Stanford University's with 47%. In this study, institutions' total number of publications between 2020 and 2021 designated that the respective institutions are actively researching a particular subject and have shifted their focus towards their research interests.

The University of Sydney was ranked second with five publications but accounted for 80% of the publications between 2020 and 2021. This demonstrates that the University of Sydney has prioritized ML in safety and health research over the last two years. The University of California San Francisco shares the third position with the University of Toronto, with 14 publications. Still, it's worth noting that the University of Toronto published 86% of all publications between 2020 and 2021, outperforming other universities. This study discovered that United States-based institutions dominated ML-related safety and health studies. Compared to other countries, it is verified that the United States has prioritized and focused on machine learning studies in safety and health.

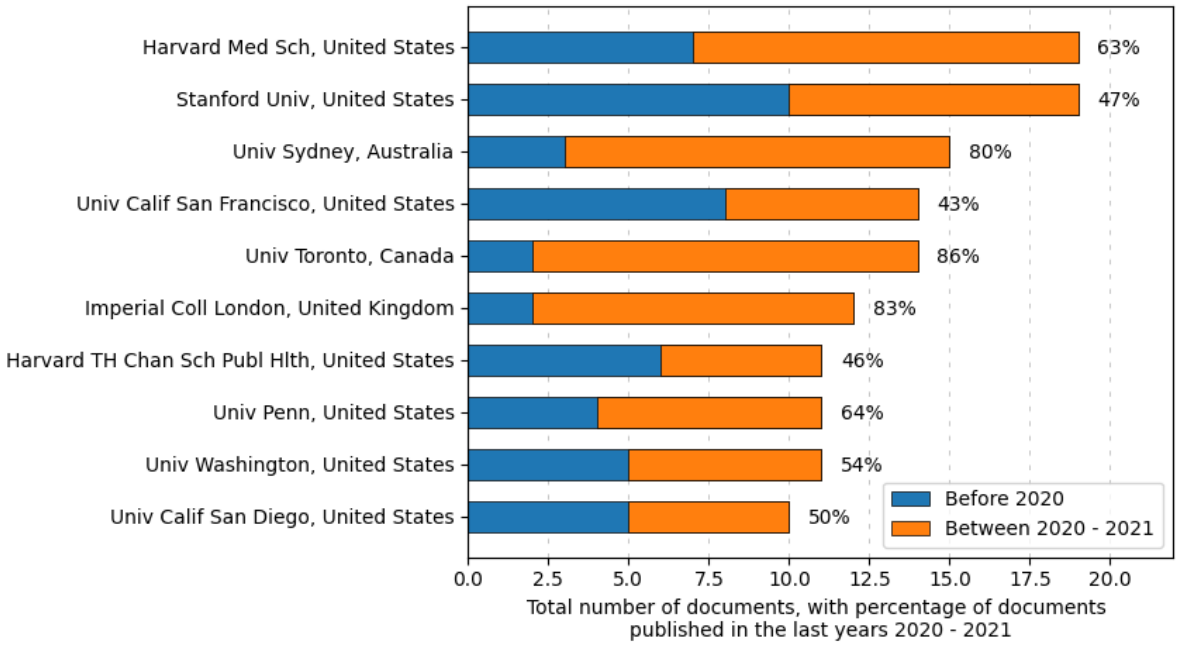

Figure 4. Ranking the most productive institutions

Analysis of the authors' keywords

The author's keywords are linked to the article's field, subfield, topic, and research concern in scholarly articles. It is beneficial for other researchers to do a subject search to locate specific publications relevant to their research interests (Abd Aziz, Abdullah & Samsudin, 2021). This study analyzed and interpreted the top ten author keywords based on their co-occurrence. Figure 5 shows the total number of documents, with the percentage published in the last two years (2020 to 2021). It is shown that the keyword "machine learning" was used in 571 previous research and accounted for 66% of those utilized in the recent two years. Other keywords, which are more than 50, are "artificial intelligence" and "electronic health records". Those keywords are recorded with 79% and 55% of the total publications during 2020 and 2021, respectively. Yet, the keyword "artificial intelligence" appeared in most publications in 2020 and 2021; and

"deep learning" was the second most appealing in 2020 and 2021, accounting for 74% of the total publications.

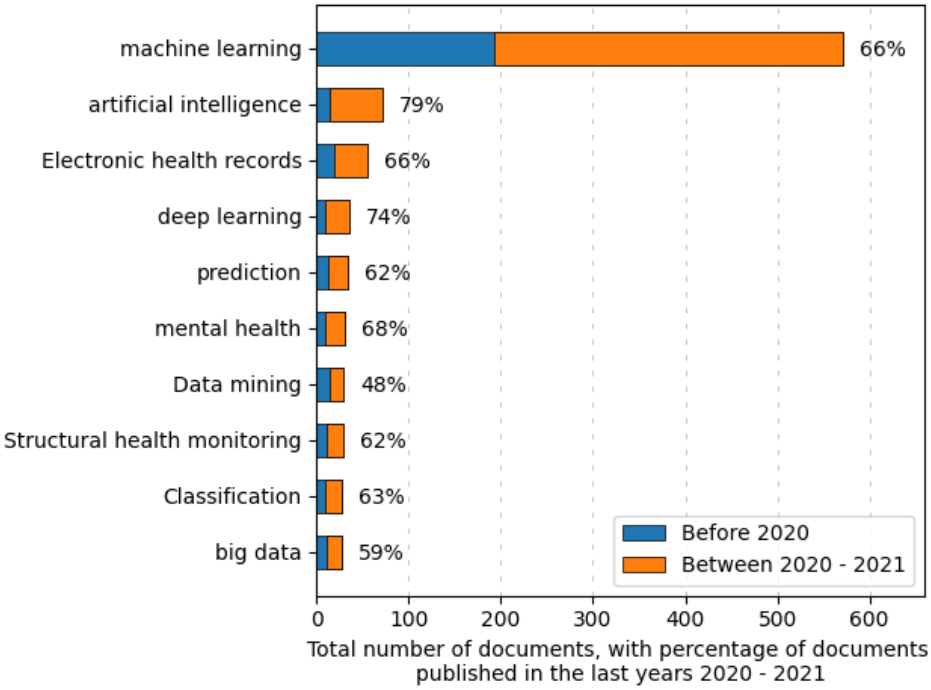

Figure 5. The Top 10 Authors' Keywords

This study has further analyzed the co-occurrence of authors' keywords using VOSviewer by generating the graphic visualization (see Figure 6). The co-occurrence of keywords refers to how they are associated or appear together within the dataset. This analysis is vital for researchers to comprehend the occurrence of keywords and interpret the results from a particular research perspective. Generally, the co-occurrence of keywords could be visualized by network visualization, overlay visualization, and density visualization.

In this study, overlay visualization is employed to depict the nexus of keywords based on the evolution of publication. To conduct this analysis, the minimum co-occurrence of selected keywords was 20 out of 2705 total. The results offered 17 keywords that met the selected criteria. The overlay visualization of those keywords is illustrated in Figure 6. Based on Figure 6, two different colours indicate the progression of authors' keywords. The keywords found in red rectangular were associated with publications from 2020 to 2021 and are currently being researched. The keywords in blue rectangular showed that those keywords had been introduced before 2020. The rectangular size is depicted as the number of occurrences, which means that the larger the rectangular size, the more the keywords have appeared in previous studies. Next, the connection or nexus between the rectangular size is related to the strength of the links or nexus where the thickness is shown in close proximity. In this analysis, the keywords "machine learning" is related more to "deep learning", "artificial intelligence", "electronic records", "mental health", "support vector machine", and "covid-19", which formed the keywords that appeared in most publications in 2020 and 2021. Nevertheless, before 2020, previous researchers gave significant attention to keywords such as "structural health monitoring", "data mining", "classification", and "prediction". This analysis revealed fascinating results indicating

that the keyword “machine learning” appeared near the keyword “covid-19” in 2021, which may indirectly help readers and future scholars understand this link to study further and precisely.

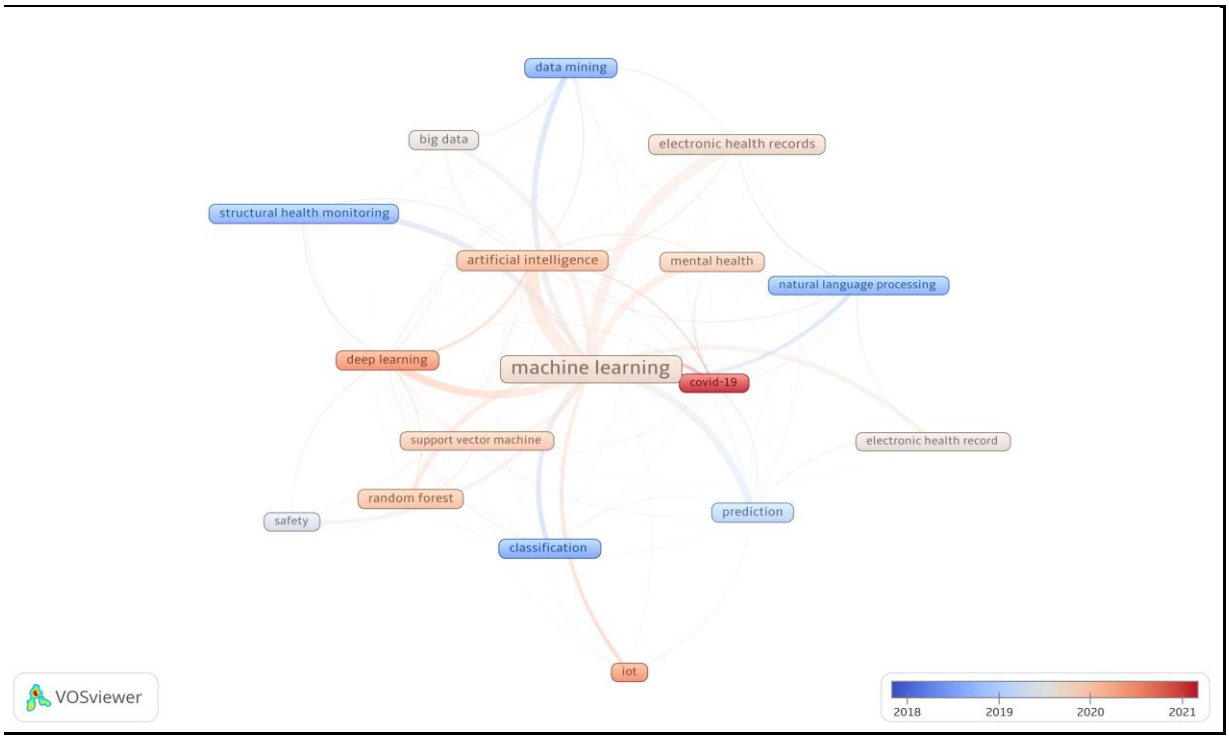

Figure 6. Overlay visualization of Authors’ Keywords

Discussion

Over the years, safety and health practitioners and researchers have used traditional approaches to address challenges relating to the physical, chemical, biological, ergonomic, and organizational environment. However, with the advent of Industry 4.0, this picture began to shift (Maheronnaghsh, Zolfagharnasab, Gorgich & Duarte, 2021). Potential research activity focused on leveraging ML approaches to improve the implementation of effective preventive measures and policies in the safety and health management framework was projected (Koklonis et al., 2021).

This study examined the history and progression of articles relating to machine learning in safety and health research indexed in the Scopus and WOS databases. ScientoPy was utilized to track publication patterns, and VOSviewer was customized to conduct evaluations and visualize all data gathered during the scrutinizing process. This study provides a brief overview of ML in safety and health research to assist readers and researchers in gaining significant knowledge for future investigations. Noteworthily, it emphasizes the importance of interdisciplinary studies involving the contemporary philosophy of ML in safety and health research and covers the significance of cognitive ML towards safety cultural anthropology. Hence, improving the current ML in safety and health is beneficial.

The evolution and progression of ML in safety and health research may be seen in the exponential growth in the number of publications since 2016. Notably, compared to the Scopus database, the publication trajectory in WoS has climbed dramatically. It’s worth mentioning

that this situation is conceivable because hospitals, businesses, and the government rely on ML to mitigate risks and actively protect people from various threats and hazards (Siegel, 2019). Thus, research into ML in safety and health has increased dramatically in recent years, demonstrating the critical role of ML in forecasting safety and health in any organization. Moreover, by examining the healthcare sector, we can see that while the potential influence of ML in healthcare deserves genuine enthusiasm, its limited acceptance in clinical care reveals that many present tactics are far from ideal, necessitating further investigation (Wiens et al., 2019). Also, big data and ML are altering almost every element of modern life. Hence, the study of ML in safety and health research comes within the interdisciplinary scope. Numerous studies on ML in safety and health have demonstrated its unmatched benefits in assisting with trend identification and prediction, resulting in significant labor, material, and financial resource savings (Jiao, Hu, Xu & Wang, 2020). Overall, its direct impact on productivity, quality of service, and especially workers' safety and security stands out conspicuously.

PLoS One published the most publications with the highest Hirsch index (h-index) on ML in safety and health research in terms of source titles. This demonstrates that interest in publishing research on ML in safety and health, whether empirical or theoretical, is increasing in PLoS One. PLoS ONE is a diverse journal community committed to promoting science for society's current and future benefits based on open-access publications. PLoS One believes that all rigorous science should be publicly accessible, discoverable, widely shared, and disseminated. This will assist readers and future researchers obtain the most recent material from these source titles to progress with their present research by utilizing the most current evidence and topics. Moreover, electronic advancements and availability nowadays facilitates the global distribution of the journals' good, boosts readership and the likelihood that an article will be cited in others' works, and enhances or maintains the journal's prestige and revenue (Vardakas, Tsopanakis, Poulopoulou & Falagas, 2015).

Based on the country's point of view, the United States has the highest average number of documents per year (ADY), followed by India and China. With the introduction of a robotic system in the United States, research on ML continues to build on improved coordination and automation concepts to improve application features, designs, and testing, which are becoming a critical component of economic growth in this highly competitive age of technology (Surya, 2016). Thus, ML has played a crucial role in shaping the technological environment in the United States, especially in safety and health. Many experts anticipate that ML will become the standard and thus will grow significantly in the forthcoming years. Meeting ambitious global health goals while operating on a shoestring budget requires India to take a precision public health strategy, which may be accomplished by integrating data collection optimization, traditional analytics, and causal ML (Huang et al., 2020). This demonstrates how leveraging data to identify key behavioral drivers in conjunction with conventional and ML analytics can aid in developing a public health strategy that maximizes the impact of limited resources, thereby spurring additional research on ML in public health in India. In China, construction mishaps can easily result in enormous casualties and property losses (Zhu, Hu, Hou & Li, 2021). ML techniques are used to study crucial components and determine the effect of various factors on predicting construction accidents' severity (Zhu, Kim, Yan, Kim & Qi, 2021). This characteristic may boost the expertise of safety specialists in the construction field, consequently increasing the efficiency of safety intervention measures, and the use of ML in safety and health research has grown significantly.

Another intriguing component of the country is the collaborative research conducted by researchers from the United States, China, and Australia. This study discovered that co-authorship was more prevalent among active scholars in the United States, China, and Australia who had published more papers. China's economy has grown at a breakneck pace in recent decades. China has engaged with the US during this process, establishing collaborations in various fields. Accordingly, China and the United States have progressively improved their cooperation in scientific research, which may signify the two countries' intention to continue collaborating in their respective fields of study (Zhu, Kim, Yan, Kim & Qi, 2021). Another appealing reason for research collaboration is that researchers are mobile, and as they advance in their careers, they develop connections that contribute to the formation and expansion of networks (Woolley, Turpin, Marceau & Hill, 2008). These networks enable researchers to pool resources for research and broaden their international reach. Based on Figure 3, China contributes more collaboration publications, which could be explained as a reflection of the nation's growing dominance as a key player in science and the number of co-authored articles published. This dominance should be recognized as a reflection of the nation's contribution to the global science arena and a likeness of the nation's strengthened links in science worldwide (Niu, 2014).

It is noteworthy that over the last 25 years, authors from Harvard Medical School of Harvard University and Stanford University in the United States have contributed the most articles published in this research perspective. The most cited publication by Julia L. Marcus from Harvard Medical School is "Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modeling study" (Marcus, Hurley, Krakower, Alexeeff, Silverberg, & Volk, 2019). The article received 31 citations to date (February 24, 2022). The paper aims to investigate the limits of existing HIV risk prediction systems, which have become barriers to pre-exposure prophylaxis (PrEP) deployment. As a result, the researchers created and validated an HIV prediction model to identify possible PrEP patients in a comprehensive healthcare system. The most cited paper corresponded by an author from Stanford University is entitled "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data" (Gianfrancesco, Tamang, Yazdany & Schmajuk, 2018). This publication obtained 234 citations until February 24, 2022. According to the paper, ML algorithms can potentially improve medical treatment by forecasting various outcomes measured in the electronic health record and offering clinical decision support based on these predictions.

By scrutinizing authors' keywords, this study found that "machine learning", "artificial intelligence", "deep learning", and "electronic health records" appeared in most publications in 2020 and 2021. Also, the keyword "machine learning" is the most frequent keyword that nexus closely to some different popular keywords such as "deep learning", "artificial intelligence", "electronic records", "mental health", "support vector machine", and "covid-19", where those keywords appeared on most publications in 2020 and 2021. The keyword "machine learning" is synonymous with "artificial intelligence", a broad term that describes how computers simulate intelligent behavior with minimal human intervention. It is widely believed to have begun with the development of robots (Hamet & Tremblay, 2017). Thus, the legal, social, and ethical dimensions of balancing the economic benefits of AI should be refined as human rights principles and legislation play a critical role in addressing AI-related ethics to contribute to human flourishing (Stahl et al., 2021).

The significance of "deep learning", "artificial intelligence", "electronic records", "mental health", "support vector machine", and "covid-19" have all been highlighted as strengths in previous health studies. Also, this study revealed fascinating results indicating that the keyword "machine learning" appeared near the keyword "covid-19" in 2021 (see Figure 6), which may indirectly help readers and future scholars understand this link to study further and precisely. A likely reason is that researchers, clinicians, and healthcare professionals worldwide are constantly looking for new technologies to aid in the fight against the Covid-19 epidemic. According to Lalmuanawma, Hussain and Chhakchhuak (2020), the evidence for using ML and Al in previous epidemics encourages researchers by providing a new angle for combating the novel Coronavirus outbreak. Also, during a Coronavirus outbreak, it may be possible to identify and manage people's emotional, psychological, and social well-being by identifying individual-level risk factors. Jha, Awasthi, Kumar, Kumar, and Sethi (2021) discovered that using explainable AI to predict mental health at the population level by exploiting survey data broadened its applicability. The study balanced knowledge discovery and 80% predictive accuracy, establishing a baseline under a novel scenario. This situation best illustrates how algorithms can identify individuals needing assistance and how additional measurements and features may improve prediction accuracy. As a result, the collocational relationship between the "deep learning", "artificial intelligence", "electronic records", "mental health", "support vector machine", and "covid-19" keywords are solid and bolstered by one another.

Before 2020, however, previous researchers placed a premium on the keywords "structural health monitoring", "data mining", "classification", and "prediction". Data mining technology may solve structural health monitoring to conduct deeper data analysis (Duan & Zhang, 2006). Additionally, because of its computational capability, data mining technology has been used to solve various structural health monitoring problems as a comprehensive strategy (Gordan, Ismail, Razak, Ghaedi & Ghayeb, 2020). One could conclude that machine learning provides advanced mathematical frameworks and algorithms for determining and modeling the performance and conditions of a structure based on monitoring data. Artificial intelligence is in high demand for structural health monitoring. At the same time, the most popular algorithms for detecting a malfunction include artificial neural networks, genetic algorithms, fuzzy logic, and principal component analysis (Gordan et al., 2022).

Conclusion

A scientometric method is a practical approach for mapping the published research on ML in safety and health research. Additionally, significant concepts from the literature were acknowledged, as was an understanding of the evolution and interaction of the field's research trends. In this study, essential concepts of ML in safety and health research have been scrutinized; where the most trending keywords were "machine learning", "artificial intelligence", "deep learning", and "electronic health records". Also, "deep learning", "artificial intelligence", "electronic records", "mental health", "support vector machine", and "covid-19" have all been highlighted as strengths in previous health studies. These keywords are critical for future researchers to examine in greater detail to comprehend their relationship through scoping reviews or systematic literature reviews.

This study can guide scientists, engineers, environmental health and safety professionals, and anyone concerned with safety and health issues who wish to learn and apply ML to their work and research. ML is gradually altering our way of thinking, working, and co-existing, and

it will remain a focus of future research and development in line with ML's advancements and keeping with the demands of IR 4.0. As critical subjects for ensuring and promoting personnel safety and health in academia and industry, health and safety concerns can and need to be continuously enhanced and prioritized to stay abreast with the rapid development of computer science and artificial intelligence. Importantly it must be seen as a leap forward for stepping up strategies to improve and safeguard safety and health aspects at workplaces at both micro and macro levels, which can also benefit workplace communities globally.

This study is a starting point that could be supplemented with a more in-depth examination of thematic content. The results obtained in this study may provide a theoretical perspective on the subject and a map of the state of the art and make potential gaps in scientific research readily apparent. A significant aspect of this scientific analysis is that it may serve as a springboard for future research on ML in safety and health to discover the contemporary philosophy of science, cognitive, and cultural anthropology pertinent to this research interest.

References

Abd Aziz, F. S., Abdullah, K. H. & Samsudin, S. (2021). Bibliometric analysis of behavior-based safety (BBS): Three decades publication trends. *Webology*, *18*, 278-293. https://doi.org/10.14704/web/v18si02/web18072

Abdel-Aal, R. E. & Mangoud, A. M. (1996). Abductive machine learning for modeling and predicting the educational score in school health surveys. *Methods of Information in Medicine*, 35(03), 265-271. https://doi.org/10.1055/s-0038-1634655

Abdullah, K. H. & Abd Aziz, F. S. (2020). Safety behaviour in the Laboratory among university students. *The Journal of Behavioral Science*, *15*(3), 51-65. Retrieved from https://so06.tci-thaijo.org/index.php/IJBS/article/view/241208

Abdullah, K. H., Hashim, M. N. & Abd Aziz, F. S. (2020). A 39 years (1980-2019) bibliometric analysis of safety leadership research. *TEST Engineering and Management*, 83, 4526-4542.

Ashok, M., Madan, R., Joha, A. & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, 62, 1-17. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Badri, A., Boudreau-Trudel, B. & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern?. *Safety Science*, 109, 403-411. https://doi.org/10.1016/j.ssci.2018.06.012

Bas, G. & Koseoglu, M. A. (2019). Analysis of the warehouse work accidents in logistics sector. *Press Academia Procedia*, 9(1), 262-268. https://doi.org/10.17261/Pressacademia.2019.1102

Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?. *The Journal of Arthroplasty*, 33(8), 2358-2361. https://doi.org/10.1016/j.arth.2018.02.067

Bonifazi, G., Corradini, E., Ursino, D., Virgili, L., Anceschi, E. & De Donato, M. C. (2021). A machine learning based sentient multimedia framework to increase safety at work. *Multimedia Tools and Applications*, 81 (1), 141-169. https://doi.org/10.1007/s11042-021-10984-z

Dhawan, S. M., Gupta, B. M. & Singh, N. K. (2020). Global Machine-learning Research: a

scientometric assessment of global literature during 2009–18. *World Digital Libraries-An International Journal*, 13(2), 105-120. https://doi.org/10.18329/09757597/2020/13209

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Duan, Z. & Zhang, K. (2006). Data mining technology for structural health monitoring. *Pacific Science Review*, 8(1), 27-36.

El Naqa, I. & Murphy, M. J. (2015). What is machine learning?. In *machine learning in radiation oncology* (pp. 3-11). Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1

Gianfrancesco, M. A., Tamang, S., Yazdany, J. & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

Goh, Y. C., Cai, X. Q., Theseira, W., Ko, G. & Khor, K. A. (2020). Evaluating human versus machine learning performance in classifying research abstracts. *Scientometrics*, 125(2), 1197-1212. https://doi.org/10.1007/s11192-020-03614-2

Gordan, M., Ismail, Z. B., Razak, H. A., Ghaedi, K. & Ghayeb, H. H. (2020). Optimisation-based evolutionary data mining techniques for structural health monitoring. *Journal of Civil Engineering and Construction*, 9(1), 14-23. https://doi.org/10.32732/jcec

Gordan, M., Sabbagh-Yazdi, S. R., Ismail, Z., Ghaedi, K., Carroll, P., McCrum, D. & Samali, B. (2022). State-of-the-art review on advancements of data mining in structural health monitoring. *Measurement*, 193, 1-38. https://doi.org/10.1016/j.measurement.2022.110939

Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23(1), 40-55. https://doi.org/10.1038/s41580-021-00407-0

Gul, M. & Ak, M. F. (2018). A comparative outline for quantifying risk ratings in occupational health and safety risk assessment. *Journal of Cleaner Production*, 196, 653-664. https://doi.org/10.1016/j.jclepro.2018.06.106

Hamet, P. & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, 36-40. https://doi.org/10.1016/j.metabol.2017.01.011

Heo, Seongbong, Moonil Kim, Hangnan Yu, Woo-Kyun Lee, Jong Ryeul Sohn, Soon-Young Jung, Kyong Whan Moon & Sang Hoon Byeon (2018). Chemical accident hazard assessment by spatial analysis of chemical factories and accident records in South Korea. *International Journal of Disaster Risk Reduction*, 27, 37-47. https://doi.org/10.1016/j.ijdrr.2017.09.016

Huang VS, Morris K, Jain M, Ramesh BM, Kemp H, Blanchard J, Isac S, Sarkar B, Gothalwal V, Namasivayam V & Kumar P. (2020). Closing the gap on institutional delivery in northern India: A case study of how integrated machine learning approaches can enable precision public health. *BMJ Global Health*, 5(10), 1-14. http://dx.doi.org/10.1136/bmjgh-2020-002340

Jayatilake, S. M. D. A. C. & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of Healthcare Engineering*, 2021, 1-20. https://doi.org/10.1155/2021/6679512

Jha, I. P., Awasthi, R., Kumar, A., Kumar, V. & Sethi, T. (2021). Learning the mental health impact of covid-19 in the united states with explainable artificial intelligence:

Observational study. *JMIR Mental Health*, 8(4), 1-11. https://doi.org/10.2196/25097

Jiao, Z., Hu, P., Xu, H. & Wang, Q. (2020). Machine learning and deep learning in chemical health and safety: a systematic review of techniques and applications. *ACS Chemical Health & Safety*, 27(6), 316-334. https://dx.doi.org/10.1021/acs.chas.0c00075

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Koklonis, K., Sarafidis, M., Vastardi, M. & Koutsouris, D. (2021). Utilisation of machine learning in supporting occupational safety and health decisions in hospital workplace. *Engineering, Technology & Applied Science Research*, 11(3), 7262-7272. http://dx.doi.org/10.48084/etasr.4205

Krstić, B., Rađenović, T. & Živković, S. (2022). Occupational Health and Safety Performance Management System: Conceptual Framework, Design, and Implementation in an Enterprise. In *Handbook of Research on Key Dimensions of Occupational Safety and Health Protection Management* (pp. 1-26). IGI Global. https://doi.org/10.4018/978-1-7998-8189-6.ch001

Lalmuanawma, S., Hussain, J. & Chhakchhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139, 110059. https://doi.org/10.1016/j.chaos.2020.110059

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 1-29. https://doi.org/10.3390/s18082674

Maheronnaghsh, S., Zolfagharnasab, H., Gorgich, M. & Duarte, J. (2021). Machine learning in occupational safety and health: protocol for a systematic review. *International Journal of Occupational and Environmental Safety*, 5(1), 32-38. https://doi.org/10.24840/2184-0954_005.001_0004

Marcus, J. L., Hurley, L. B., Krakower, D. S., Alexeeff, S., Silverberg, M. J. & Volk, J. E. (2019). Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. *The Lancet HIV*, 6(10), 688-695. https://doi.org/10.1016/S2352-3018(19)30137-7

Martinez, S., del Mar Delgado, M., Marin, R. M. & Alvarez, S. (2019). Science mapping on the Environmental Footprint: A scientometric analysis-based review. *Ecological Indicators*, 106, 1-11. https://doi.org/10.1016/j.ecolind.2019.105543

Mingers, J. & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. European Journal of Operational Research, 246(1), 1-19. https://doi.org/10.1016/j.ejor.2015.04.002

Mukhamediev, R. I., Symagulov, A., Kuchin, Y., Yakunin, K. & Yelis, M. (2021). From classical machine learning to deep neural networks: A simplified scientometric review. *Applied Sciences*, 11(12), 1-26. https://doi.org/10.3390/app11125541

Naeem M, Jamal T, Diaz-Martinez J, Butt SA, Montesano N, Tariq MI, De-la-Hoz-Franco E & De-La-Hoz-Valdiris E. (2022). Trends and future perspective challenges in big data. In *Advances in Intelligent Data Analysis and Applications* (pp. 309-325). Springer, Singapore. https://doi.org/10.1007/978-981-16-5036-9_30

Niu, X. S. (2014). International scientific collaboration between Australia and China: A mixed-methodology for investigating the social processes and its implications for national innovation systems. *Technological Forecasting and Social Change*, 85, 58-68. https://doi.org/10.1016/j.techfore.2013.10.014

Paltrinieri, N., Comfort, L. & Reniers, G. (2019). Learning about risk: Machine learning for risk assessment. *Safety Science*, 118, 475-486. https://doi.org/10.1016/j.ssci.2019.06.001

Poh, C. Q., Ubeynarayana, C. U. & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, 93, 375-386. https://doi.org/10.1016/j.autcon.2018.03.022

Ruiz-Rosero, J., Ramírez-González, G. & Viveros-Delgado, J. (2019). Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications. *Scientometrics*, 121(2), 1165-1188. https://doi.org/10.1007/s11192-019-03213-w

Sachs, L. (1990). Safety and risk in a context of culture. *International Journal of Risk & Safety in Medicine*, 1(4), 255-265. https://doi.org/10.3233/jrs-1990-1402

Sarkar, S. & Maiti, J. (2020). Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Safety Science*, 131, 1-25. https://doi.org/10.1016/j.ssci.2020.104900

Shao, B., Hu, Z., Liu, Q., Chen, S. & He, W. (2019). Fatal accident patterns of building construction activities in China. *Safety Science*, 111, 253-263. https://doi.org/10.1016/j.ssci.2018.07.019

Siegel, E. (2019, February 2). *Five Ways Your Safety Depends on Machine Learning*. KDnuggets. Retrieved from https://www.kdnuggets.com/2019/02/dr-data-five-ways-safety-depends-machine-learning.html

Simsekler, M. C. E., Rodrigues, C., Qazi, A., Ellahham, S. & Ozonoff, A. (2021). A comparative study of patient and staff safety evaluation using tree-based machine learning algorithms. *Reliability Engineering & System Safety*, 208, 1-13. https://doi.org/10.1016/j.ress.2020.107416

Sirola, M. & Hulsund, J. E. (2021). Machine-Learning Methods in Prognosis of Ageing Phenomena in Nuclear Power Plant Components. *International Scientific Journal of Computing*, 20(1), 11-21. https://doi.org/10.47839/ijc.20.1.2086

Sofyan, D. & Abdullah, K. H. (2022). Scientific developments in educational innovation research in Indonesia and Malaysia: a scientometric review. *International Journal of Educational Innovation and Research*, 1(1), 42-51. https://doi.org/10.31949/ijeir.v1i1.2312

Stahl BC, Andreou A, Brey P, Hatzakis T, Kirichenko A, Macnish K, Shaelou SL, Patel A, Ryan M & Wright D. (2021). Artificial intelligence for human flourishing–Beyond principles for machine learning. *Journal of Business Research*, 124, 374-388. https://doi.org/10.1016/j.jbusres.2020.11.030

Surya, L. (2016). An exploratory study of Machine Learning and it's future in the United States. *International Journal of Creative Research Thoughts (IJCRT)*, 4(1), 862-866.

Tang, S. & Golparvar-Fard, M. (2021). Machine learning-based risk analysis for construction worker safety from ubiquitous site photos and videos. *Journal of Computing in Civil Engineering*, 35(6), 1-19. https://doi.org/10.1061/(asce)cp.1943-5487.0000979

Trung, N. D., Huy, D. T. N. & Le, T. H. (2021). IoTs, machine learning (ML), AI and digital transformation affects various industries-principles and cybersecurity risks solutions. *Webology*, 18, 501-513. https://doi.org/10.14704/web/v18si04/web18144

Tullu, M. S. (2019). Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi Journal of Anaesthesia*, 13(Suppl 1), 12-17.

https://doi.org/10.4103/sja.sja_685_18

Van Nunen, K., Li, J., Reniers, G. & Ponnet, K. (2018). Bibliometric analysis of safety culture research. *Safety Science*, 108, 248-258. https://doi.org/10.1016/j.ssci.2017.08.011

Van, T. N. & Quoc, T. N. (2021). Research Trends on Machine Learning in Construction Management: A Scientometric Analysis. *Journal of Applied Science and Technology Trends*, 2(03), 96-104. https://doi.org/10.38094/jastt203105

Vardakas, K. Z., Tsopanakis, G., Poulopoulou, A. & Falagas, M. E. (2015). An analysis of factors contributing to PubMed's growth. *Journal of Informetrics*, 9(3), 592-617. https://doi.org/10.1016/j.joi.2015.06.001

Varshney, K. R. (2016, January). Engineering safety in machine learning. In *2016 Information Theory and Applications Workshop (ITA)* (pp. 1-5). IEEE. https://doi.org/10.1109/ita.2016.7888195

Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M & Ossorio PN (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9), 1337-1340. https://doi.org/10.1038/s41591-019-0548-6

Woolley, R., Turpin, T., Marceau, J. & Hill, S. (2008). Mobility matters: Research training and network building in science. *Comparative Technology Transfer and Society*, 6(3), 159-184. https://doi.org/10.1353/ctt.0.0014

Yang, W., Fidelis, T. T. & Sun, W. H. (2019). Machine learning in catalysis, from proposal to practicing. *ACS omega*, 5(1), 83-88. https://doi.org/10.1021/acsomega.9b03673

Yong, Z., Xiaoming, Z. & Alshehri, M. D. (2021). A machine learning-enabled intelligent application for public health and safety. *Neural Computing and Applications*, 1-14. https://doi.org/10.1007/s00521-021-06301-2

Zhu, R., Hu, X., Hou, J. & Li, X. (2021). Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, 145, 293-302. https://doi.org/10.1016/j.psep.2020.08.006

Zhu, Y., Kim, D., Yan, E., Kim, M. C. & Qi, G. (2021). Analysing China's research collaboration with the United States in high-impact and high-technology research. *Quantitative Science Studies*, 2(1), 363-375. https://doi.org/10.1162/qss_a_00098