buildings

MDPI

Article

Discovering the Research Topics on Construction Safety and Health Using Semi-Supervised Topic Modeling

Kai Zhou [1], Jun Wang [1,*], Baabak Ashuri [2] and Jianli Chen [3]

[1] School of Management Engineering, Qingdao University of Technology, Qingdao 266520, China
[2] School of Civil and Environmental Engineering and School of Building Construction, Georgia Institute of Technology, Atlanta, GA 30332, USA
[3] Department of Civil Engineering, University of Utah, Salt Lake City, UT 84112, USA
* Correspondence: wangjun.gt@gmail.com

Abstract: Safety and health have been one of the major issues in the construction industry worldwide for decades, and the relevant research has correspondingly drawn much attention in the academic field. Considering the expanding size and increasing heterogeneity of this research field, this paper proposes the topic modeling approach to cluster latent topics, extract coherent keywords, and discover evolving trends over the past three decades. Focusing on a total of 1984 articles published in 27 different journal sources until February 2023, this paper applied both unsupervised topic modeling techniques—Latent Dirichlet Allocation (LDA) and Correlation Explanation (CorEx)—and their semi-supervised versions—Guided LDA and Anchored CorEx. The evolving trends and inter-relationship of 15 research topics generated by the Anchored CorEx model (the best-performing model) were analyzed. Top-listed documents of major topics were analyzed to discuss their standalone research focuses. The results of this paper provided helpful insights and implications of existing research and offered potential guides for future research on construction safety and health by helping researchers (1) select research topics of interest and clearing decaying topics; (2) extract the top words of each research topic using systematic approaches; and (3) explore the interconnection of different research topics as well as their standalone focuses.

Keywords: construction safety and health; research trends; topic modeling; natural language processing

check for updates

Citation: Zhou, K.; Wang, J.; Ashuri, B.; Chen, J. Discovering the Research Topics on Construction Safety and Health Using Semi-Supervised Topic Modeling. *Buildings* **2023**, *13*, 1169. https://doi.org/10.3390/buildings13051169

Academic Editor: Yuting Chen

Received: 23 March 2023
Revised: 21 April 2023
Accepted: 27 April 2023
Published: 28 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 and Background

Safety and health have been continuously considered as one of the top concerns in the construction industry across the globe. In the U.S., the number of construction fatalities has steadily increased over the past decade, while the fatality rate has not decreased. A total of 1061 fatalities were reported from the U.S. construction industry in 2019 [1], and 1752 fatalities were reported from the Chinese construction industry for the first half-year of 2018 [2]. The fatal injury rate in the construction industry was 1.74 per 100,000 in the UK in 2019, but it was still almost four times the all-industry rate [3]. As such, safety and health problems in the construction industry have become and continue to be a trendy research area, and the number of relevant scientific publications has been continuously increasing over the past thirty years.

Due to the long-lasting research popularity and the large number of publications, it is necessary and critical to systematically understand what research topics in the field of construction safety and health are studied and how they have evolved in the past. Some review types of studies have been carried out to summarize and analyze such topics and trends. Zhou et al. selected and evaluated 119 relevant papers published from 1986 to 2012 to examine the trends of advanced technology applications in the construction safety domain [4]. Skibniewski summarized the research trends on how information technology applications have facilitated construction safety, based on 71 articles published in *Automation in Construction* from 2000 to 2014 [5]. Alruqi et al. examined 107 articles published from 2000 to 2016 to study safety climate dimensions and their relationship to construction safety performance [6]. More recently, Sarkar and Maiti applied a science mapping approach to review 232 journal articles published from 1995 to 2019 to understand the application of machine learning in construction occupational accidents [7].

These types of qualitative reviews are usually related to some specific topics underneath construction safety and health research, and the number of articles that are considered is often limited. As the number and heterogeneity of relevant research papers continue to increase, it becomes increasingly difficult to obtain a synthetic image of the research topics that are being investigated [8]. Construction safety and health can be related to numerous factors such as safety climate, safety culture, safety training, worker behavior, situation awareness, and technology application, further exacerbating this problem. In this case, text mining can help to analyze vast amounts of research objectively [9].

Given the apparent limitations of conducting qualitative reviews, scientometric analysis has been recently adopted as a quantitative method to evaluate the importance of articles and authors and facilitate the review process in the field of construction safety and health [7,10,11]. Despite its ability to discover the inherent relationships among research works using graphic representation [12], such analysis often fails to provide topic-related information for researchers to better understand different research contexts in detail [13]. Topic modeling has great potential to discover and evaluate latent topics in a research domain, but little relevant work has been performed in the field of construction safety and health.

Topic modeling utilizes statistical and optimization algorithms to extract semantic information from a large collection of texts, and it has become an emerging method of systematically examining textual data [14]. It has been widely applied to process and analyze textual data for various purposes, such as understanding scientific research trends [15,16], exploring social networks [17,18], analyzing political attention [19,20], facilitating biomedical recommendations [21], and evaluating public health [22]. It has also been previously adopted by researchers in the architecture, engineering, construction, and facilities management industry to facilitate textual data processing and analysis for various purposes, such as identifying patterns in construction defect litigation cases [23], analyzing public concerns in mega-infrastructure projects [24], and discovering themes and trends in transportation research [13].

Much research has been conducted on applying topic modeling techniques to overview scientific papers and discover research topics, and those works can generally be summarized into two categories. The first category relates to utilizing topic modeling to overview scientific papers in general, and the scope is not limited to any specific research domain. For example, Hall et al. applied unsupervised topic modeling to the ACL Anthology to analyze historical trends in the field of Computational Linguistics from 1978 to 2006 [15]. Yau et al. utilized LDA and its extensions to separate a set of scientific documents into several clusters and evaluated the clustering results [25].

Different from the first category, the second one focuses on applying topic modeling to discover research trends in a specific research field. For instance, Jiang et al. employed a topic modeling-based bibliometric analysis to evaluate 1726 articles in the field of hydropower research, to discover research development, current trends, and intellectual structure [14]. Choi et al. applied LDA to the abstracts of 2356 documents to discover research topics and trends in the area of personal information privacy [9]. Carnot et al. also applied LDA to analyze research published in the proceedings of well-established dependability conferences and discover research trends [8]. Amado et al. leveraged topic modeling to explore research trends on big data in marketing based on the analysis of 1560 articles [26]. Most recently, Chen et al. utilized structural topic modeling to process 3963 articles published in *Computer & Education* to detect latent topics and trends in educational technologies [27]. These related works have largely automated and facilitated the process of overviewing, understanding, and clustering scientific papers in various research domains, especially when the quantity and heterogeneity of scientific papers have increased significantly in many research fields.

To this end, this paper leverages topic modeling, a natural language processing technique, to systematically cluster hotspots of construction safety and health research. Coherent keywords of each of those topics were extracted, and the evolving trend of each topic was analyzed using the Mann–Kendall test. Research insights were also discussed from both inter-topic and standalone perspectives. This study is expected to outline an overall picture of the state-of-the-art research trends, provide helpful insights and implications of existing research, and offer potential guides for future research on construction safety and health.

2. Research Methods

This section starts with technical descriptions to help readers to understand the basic techniques of two topic modeling techniques: Latent Dirichlet Allocation (LDA) and Anchored Correlation Explanation (CorEx). A research flowchart of the proposed study that contains data collection, text mining, and topic modeling is then provided in Figure 1 and discussed in detail.

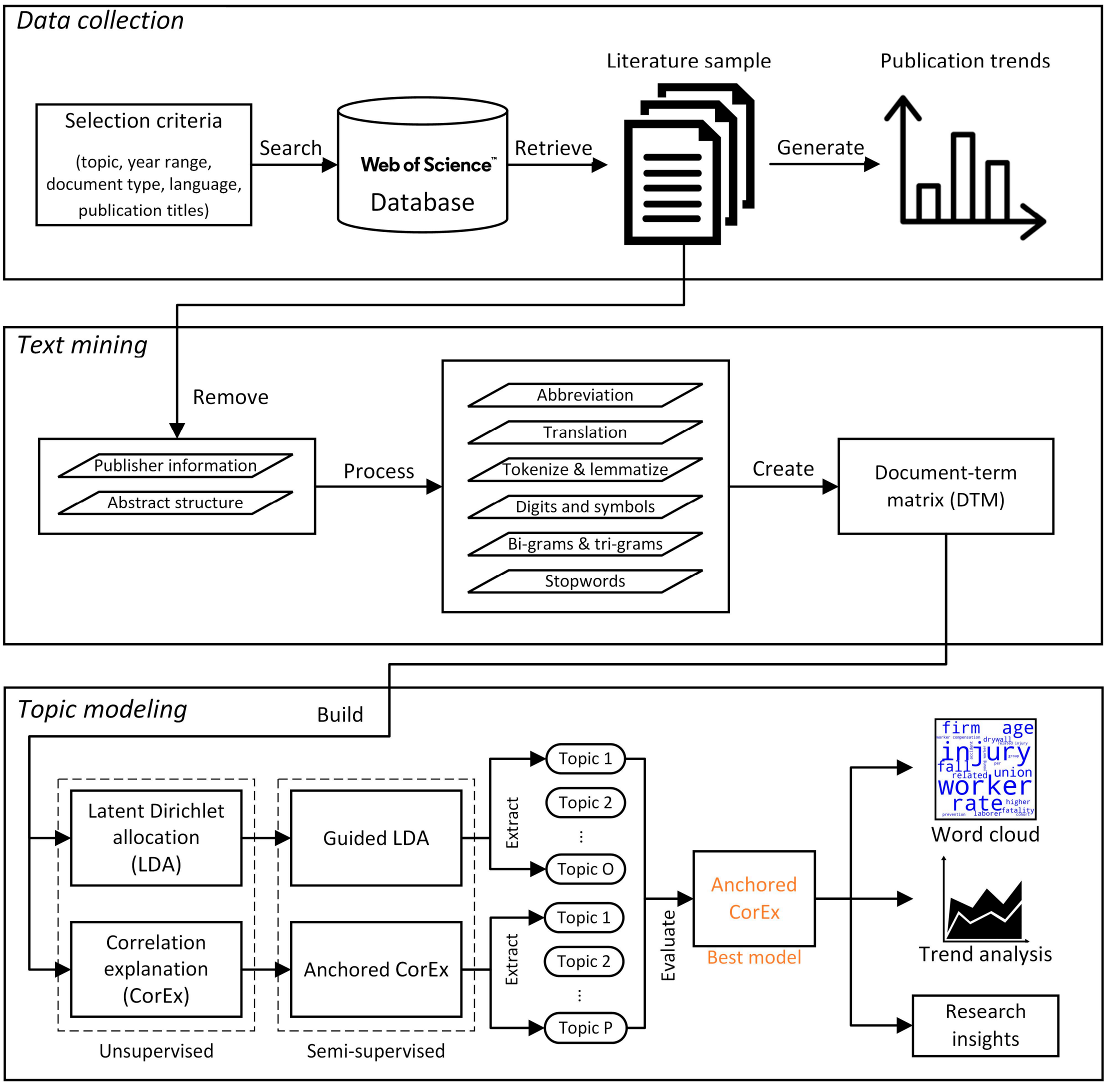


Figure 1. Flowchart for the methods of the proposed study.

2.1. Latent Dirichlet Allocation (LDA) and Correlation Explanation (CorEx)

One of the most popular and widely utilized topic models is LDA, which was initially proposed by Blei et al., and it can be graphically represented in Figure 2 [28]. LDA is a generative model, and it assumes that each document $d(d = 1 \ldots D)$ is a random mixture of topics $k(k = 1 \ldots K)$, and each topic k is a distribution of all words $n(n = 1 \ldots N)$ over the vocabulary. θ_d stands for the topic proportion of each document d, β_k stands for the word distribution of each topic k, while α and η are the Dirichlet priors of θ and β, respectively. Based on θ, the topic assignment of the nth word in the dth document $Z_{d,n}$ is determined. The shaded variable $W_{d,n}$ denotes each word n in each document d, and this is the only observed variable in the entire LDA model. Each plate represents the corresponding number of repetitions.

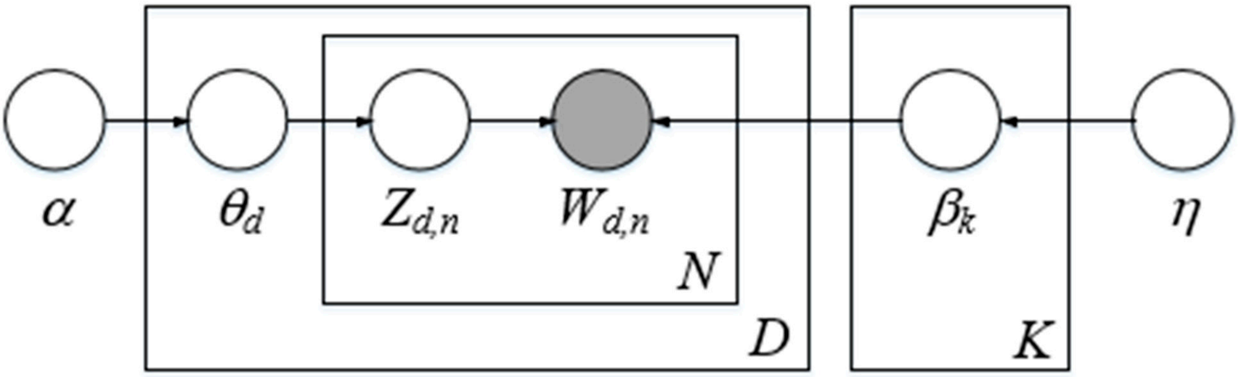


Figure 2. Graphical model representation of LDA adapted from Blei et al. [28].

Another interesting yet quite different topic model is CorEx, which was more recently proposed by Gallagher et al. [29] based on the framework that was brought up by Steeg et al. [30]. Based on information theory, CorEx model searches for topics that most explain a set of documents using a mechanism that is different from the generative assumption of the LDA model. More specifically, CorEx model aims to maximize the total correlation (TC) of a group of words and the latent topics, which can be expressed as

$$TC(X_G; Y) = \sum_{i \in G} I(X_i : Y) - I(X_G : Y) \quad (1)$$

where X_G represents a group of word types, Y denotes a topic to be learned, and I stands for mutual information. CorEx model essentially groups multiple sets of words into multiple topics and seeks to maximally explain the dependencies of words in documents through latent topics.

2.2. Semi-Supervised Topic Modeling

Both the original LDA and CorEx are unsupervised models, and they, therefore, do not require any prior knowledge of the number and themes of topics to proceed with the topic modeling process. The inferred topics are generated from the statistical structure of the document-term matrix data based on the maximum likelihood or maximum total correlation principle. However, domain experts could have a certain level of knowledge on what words are semantically closer and should be put into the same topic. Researchers have found that by injecting prior knowledge into the traditional topic modeling process, unsupervised learning problems can be converted to semi-supervised learning problems to improve the quality and interpretability of the inferred topics. Therefore, this paper tested both unsupervised learning topic models (LDA and CorEx) and their semi-supervised versions (Guided LDA [31] and Anchored CorEx [29]) to find the model with the best performance.

2.3. Methods of the Proposed Study

A flowchart for the methods of the proposed study is presented in Figure 1, and it contains three sections: data collection, text mining, and topic modeling.

2.3.1. Data Collection

In the data collection section, pre-defined data selection criteria, as shown in Table 1, were applied to the Web of Science database to retrieve the literature sample. Records including title, keywords, abstract, publication year, and journal source of 5779 articles were initially retrieved. Because this study focused on the safety and health issues of construction workers, a thorough manual screening process was conducted to further remove articles that did not fit into the research scope. Such removal was applied to articles that focus on the safety and health issues of workers in other industries [32,33], articles that deal with the safety and health issues of the structure but not workers [34,35], articles that discuss non-occupational safety and health issues [36,37], and articles that are not relevant at all. Abstracts of eight articles that were missing in the original dataset were manually filled in because the abstract of an article contains rich textual information for text mining, and the topic modeling technique generally performs better for large texts than for short texts. Finally, after removing review articles, 1984 original articles remained in the literature sample for publication trend analysis from both chronological and journal source perspectives.

Table 1. Data selection criteria.

Data Source	Web of Science
Topic	Construction Safety OR Construction Health
Year range	Until February 2023
Document type	Articles
Language	English
Journal source	(Accident Analysis and Prevention) OR (Advanced Engineering Informatics) OR (Advances in Civil Engineering) OR (American Journal of Industrial Medicine) OR (Applied Ergonomics) OR (Automation in Construction) OR (Buildings) OR (Construction Management and Economics) OR (Engineering Construction and Architectural Management) OR (International Journal of Construction Management) OR (International Journal of Environmental Research and Public Health) OR (International Journal of Industrial Ergonomics) OR (International Journal of Occupational Safety and Ergonomics) OR (Journal of Civil Engineering and Management) OR (Journal of Cleaner Production) OR (Journal of Computing in Civil Engineering) OR (Journal of Construction Engineering and Management) OR (Journal of Engineering Design and Technology) OR (Journal of Management in Engineering) OR (Journal of Occupational and Environmental Medicine) OR (Journal of Safety Research) OR (Occupational and Environmental Medicine) OR (Reliability Engineering & System Safety) OR (Safety Science) OR (Sensors) OR (Sustainability) OR (Work-A Journal of Prevention Assessment & Rehabilitation)

2.3.2. Text Mining

In the text mining section, the literature sample was cleaned to remove unnecessary and noisy information and was further processed using various text mining techniques to create the desired data format–document term matrix (DTM)—for topic modeling.

Given the features of academic publications, some articles included the publisher's copyright arguments in the abstract, such as 'Elsevier Ltd.' or 'All rights reserved'. Such arguments were removed from the dataset because they did not contain any meaningful information. Because some journals have specific formatting requirements for the abstract part by incorporating such terms as ‘Objectives’, ‘Methods’, ‘Results’, and ‘Conclusions’, these terms were also removed for the same reason.

After removing those unnecessary terms, the next step is to conduct data processing. It is common for academic publications to utilize a large number of abbreviations for simplicity reasons, but abbreviations are semantically much less meaningful than their corresponding full expressions. What is worse, the same abbreviation can refer to different full expressions under different contexts. For example, ‘CI’ can refer to ‘confidence interval’, ‘construction industry’, or ‘composite indicator’ in different scenarios, but topic models would consider them the same if they are presented in the abbreviated form. The full expressions of abbreviations were therefore restored using Neumann et al.’s algorithm [38].

Given the diverse backgrounds of researchers, both American and British English exist in the literature sample. To ensure that topic models understand that ‘building information modeling’ equates to ‘building information modelling’, a translator [39] was used to convert British spellings to American ones. Using the natural language toolkit (NLTK [40]) package in Python, texts in the literature sample were further tokenized into words that were attached with part-of-speech (POS) tags, which can be used to identify the lexical category of each word in a sentence. Based on those tags, symbols (punctuations, parentheses, equal signs, etc.) and cardinal digits were removed from the texts because they did not contain enough semantic information. Each of the remaining words was lemmatized to reduce inflectional forms to a common base form.

In order to better capture the meanings of a document, it would be useful to analyze two or three words together as a phrase instead of treating them separately [8]. For example, the phrase ‘real time’ can have very different meanings from either ‘real’ or ‘time’. Specifically, trigrams (phrases consisting of three words, e.g., ‘unmanned aerial vehicle’) and bigrams (phrases consisting of two words, e.g., ‘real time’) are often identified and treated as one word in the topic modeling analysis. Both the pointwise mutual information (PMI) [41] and the term frequency technique were adopted to generate an initial shortlist of target trigrams and bigrams, which were later manually screened to ensure their suitability with the context of this study. Finally, stop-words and global salient words were also removed because they occur so frequently that they do not help decompose the document collection.

2.3.3. Topic Modeling

As an important hyperparameter of a topic model, the number of topics is usually not easy to determine. There is no perfect answer to the question of setting the optimal number of topics in this case because topic modeling is an unsupervised learning technique. Previous studies have utilized predictive likelihood (or equivalently, perplexity) as a measure for estimating the optimal number of topics [9,14], but it was also found that models which achieve better predictive likelihood may not yield human interpretable topics [42]. Therefore, other research has avoided relying on this measure, but chooses to check the quality of topics with different numbers [43] or just use a subjective topic number based on their experience [8,13]. After experimenting with different topic numbers, this study eventually generated 15 topics based on the subjective analysis of the literature sample.

Due to the lack of human interpretability of topic models that achieve better perplexity, a different metric—topic coherence—is adopted to help evaluate the performance of different topic models. Topic coherence relies on the co-occurrence of top words in each topic, and Roder et al. found that the C_v measure of topic coherence had the best performance among other measures [44].

3. Results and Discussion

This section first presents the publication trends and the recurring trigrams and bigrams of the literature sample. Results of the four different topic models—two unsupervised ones and two semi-supervised ones—are then provided, and the anchored CorEx model has the highest coherence score among the four models. Finally, research trends, word clouds, topic co-occurrence heatmap, as well as research insights of each topic are analyzed and discussed based on the results of the best topic model.

3.1. Publication Trends

Figure 3 shows the number of articles published each year, and the earliest document that was published by Laufer and Ledbetter can be traced back to 1986 [45]. It is clear to observe that the number of published articles has significantly increased over the past decades, especially since 2008. Figure 4 presents the number of relevant articles published in each journal, and it is interesting to find that more than 50% of relevant articles were published in four journal sources: *Journal of Construction Engineering and Management*, *Safety Science*, *Automation in Construction*, and *American Journal of Industrial Medicine*.

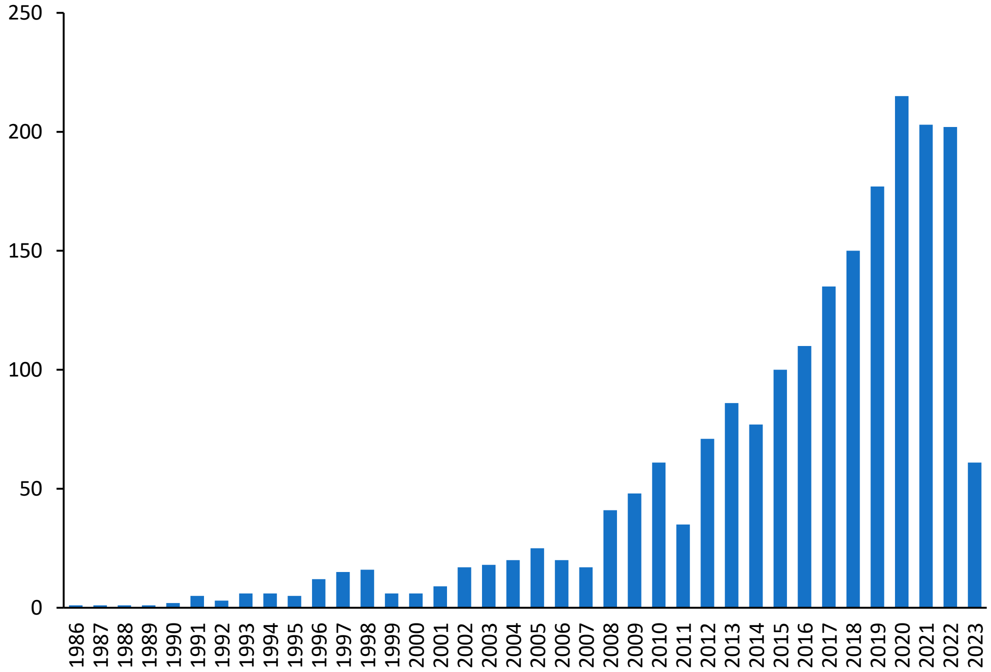


Figure 3. Number of articles published each year.

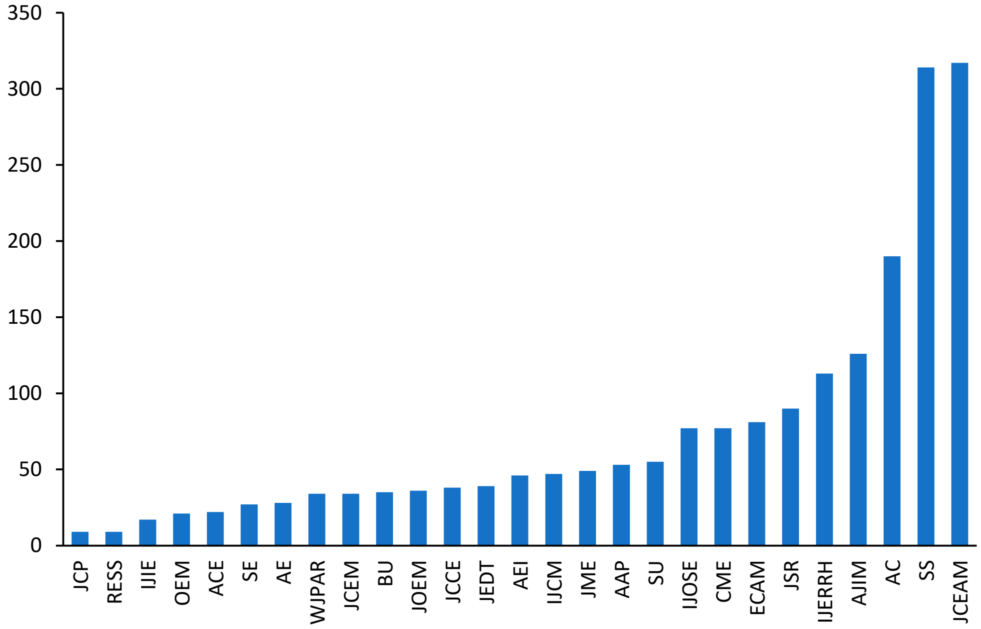


Figure 4. Number of articles published in each journal.

3.2. Trigrams and Bigrams

After a series of text mining processing discussed in Section 3.2, the 18 most popular trigrams and 33 most popular bigrams were identified and selected from the dataset, and their occurrences and percentage (based on the literature sample of 301,085 words in total) are presented in Table 2. Note that the selection of trigrams or bigrams is based on both objective measures like PMI and term occurrences and subjective domain knowledge and experience, and there is no academic consensus yet on how trigrams and bigrams should be selected from the literature sample [8].

Table 2. List of selected trigrams and bigrams and their corresponding occurrences and percentage.

Phrase	Occurrences	Percentage	Phrase	Occurrences	Percentage
Trigrams			virtual reality	196	0.065%
building information modeling	261	0.087%	prevention design	194	0.064%
structural equation modeling	182	0.060%	near miss	180	0.060%
personal protect equipment	88	0.029%	tower crane	171	0.057%
unmanned aerial vehicle	87	0.029%	machine learning	130	0.043%
real time locate	64	0.021%	deep learning	126	0.042%
convolutional neural network	54	0.018%	computer vision	125	0.042%
analytic hierarchy process	51	0.017%	safety regulation	91	0.030%
artificial neural network	42	0.014%	safety inspection	84	0.028%
support vector machine	38	0.013%	human error	83	0.028%
inertial measurement unit	37	0.012%	leading indicator	78	0.026%
natural language processing	35	0.012%	neural network	72	0.024%
case based reasoning	34	0.011%	psychological contract	71	0.024%
exploratory factor analysis	34	0.011%	internet thing	69	0.023%
social network analysis	33	0.011%	Bayesian network	66	0.022%
radio frequency identification	33	0.011%	confidence interval	58	0.019%
hand arm vibration	32	0.011%	eye tracking	57	0.019%
low back pain	31	0.010%	ethnic minority	55	0.018%
confirmatory factor analysis	30	0.010%	safety investment	52	0.017%
			root cause	50	0.017%
Bigrams			heart rate	50	0.017%
safety management	981	0.326%	leading cause	49	0.016%
safety climate	897	0.298%	heavy equipment	48	0.016%
safety performance	887	0.295%	site layout	48	0.016%
unsafe behavior	361	0.120%	lean construction	46	0.015%
safety culture	270	0.090%	self reported	44	0.015%
real time	258	0.086%	psychological capital	42	0.014%

3.3. Topic Models Results

Selected results of the first two unsupervised learning topic models—LDA and CorEx—were presented in Table 3. The coherence score of a particular topic, calculated based on the top 10 words of each topic, was presented in the last column. The average coherence score of all topics for the LDA model and the CorEx model was 0.4556 and 0.5480, respectively. Based on the top 10 words of each topic, a topic name was manually assigned based on their semantic meanings. For example, the top words of Topic 5 from the LDA model suggested that the topic was related to utilizing real-time technologies to detect and monitor workers and equipment, so the name 'object detection' can be assigned to this topic. In another example, the top words of Topic 2 from the CorEx model suggested that the topic was related to understanding safety climate and workers' safety perception and behavior using questionnaires and surveys, so the name 'safety climate' can be assigned to this topic. While it is not difficult to identify the themes of these topics, there are some other topics whose top words may be uninterpretable. For example, the top words of Topic 2 from the LDA model and Topic 12 from the CorEx model belong to multiple themes, so their names were labeled as 'unidentified'. Note that the coherence scores of these 'unidentified' topics were usually much lower than those of other topics, indicating that a high coherence score is indeed associated with better human interpretability.

Table 3. Selected results of the LDA model and the CorEx model.

Topic Number	Topic Name	Top 10 Words of Each Topic	Coherence Score
		Selected results of the LDA model	
2	Unidentified	exposure, hearing, load, lift, path, dust, protect, loss, noise, crane	0.2091
5	Object detection	worker, monitoring, equipment, fatigue, real_time, detection, sensor, technology, object, activity	0.5966
6	Safety climate	factor, safety_climate, safety_performance, safety_management, accident, management, communication, incident, safety_culture, worker	0.5308
		... ...	
		Selected results of the CorEx model	
1	Real-time detection	detection, automatically, sensor, real_time, accuracy, monitoring, automated, algorithm, wearable, machine_learning	0.8030
2	Safety climate	safety_climate, questionnaire, structural_equation_modeling, influence, survey, behavior, relationship, perception, positive, positively	0.7394
12	Unidentified	executive, talk, involved, piece, act, supplemented, violation, earthmoving, plant, workflow	0.3648
		... ...	

Due to the unsupervised nature of the LDA and CorEx models, a certain number of topics from both models (e.g., Topic 2 from the LDA model and Topic 12 from the CorEx model) are difficult to be interpreted by humans. To alleviate this difficulty, 15 topic priors (topic names and guiding words) were manually created based on those high-quality topics generated by LDA and CorEx models and our best knowledge to guide the topic modeling process.

Leveraging topic priors shown in Table 4, two semi-supervised learning topic models—the Guided LDA model and the Anchored CorEx model—were built. The average coherence score for the Guided LDA model was 0.5637, and the average coherence score for the Anchored CorEx model was 0.6779. Note that both semi-supervised models outperformed their corresponding unsupervised version from the perspective of coherence score, and the Anchored CorEx model had the best performance among all models. Results of the Anchored CorEx model were presented in Table 5 and used for further analysis and discussions.

Table 4. Topic priors and the guiding words.

Topic Number	Topic Name	Guiding Words
1	Object tracking	'tracking', 'monitoring', 'detection'
2	Safety climate	'safety_climate', 'safety_culture'
3	Ergonomics	'posture', 'musculoskeletal', 'ergonomic'
4	Design	'design', 'prevention_design'
5	Fall	'fall', 'fatality', 'fatal'
6	Safety training	'training'
7	Health and disease	'cancer', 'disease'
8	Worker behavior	'behavior', 'attitude'
9	Equipment contact	'proximity', 'near_miss', 'equipment'
10	Real-time technology	'real_time', 'technology', 'sensor'
11	Survey	'survey', 'questionnaire'
12	Machine learning	'machine_learning'
13	Computer vision	'computer_vision', 'image', 'video'
14	Interview	'delphi', 'interview'
15	Structural equation modeling	'structural_equation_modeling'

Table 5. Top 10 words of each topic from the Anchored CorEx model.

Topic Number	Topic Name	Top 10 Words of Each Topic	Coherence Score
1	Object tracking	monitoring, detection, tracking, wearable, object, detecting, detect, locate, monitor, signal	0.7438
2	Safety climate	safety_climate, safety_culture, positive, supervisor, commitment, organizational, safety_performance, perception, leadership, organization	0.6821
3	Ergonomics	musculoskeletal, ergonomic, posture, disorder, ergonomics, back, awkward, pain, physical, task	0.8674
4	Design	design, prevention_design, designer, phase, engineer, architect, designing, building, planning, tool	0.6612
5	Fall	fatality, fall, fatal, injury, height, struck, leading_cause, death, occurred, protect	0.7609
6	Safety training	training, education, program, skill, trainee, virtual_reality, apprentice, experience, trainer, curriculum	0.6373
7	Health and disease	disease, cancer, age, male, exposure, occupational, worker, older, safety_management, population	0.6023
8	Worker behavior	behavior, attitude, unsafe_behavior, norm, cognitive, supportive, safe, unsafe, planned, cost	0.4768
9	Equipment contact	equipment, near_miss, proximity, operator, collision, blind, crane, operation, personal_protect_equipment, vehicle	0.5681
10	Real-time technology	technology, real_time, sensor, sensing, wireless, application, device, internet_thing, positioning, smart	0.7425
11	Survey	survey, questionnaire, administered, response, respondent, completed, case, country, demographic, time	0.4655
12	Machine learning	machine_learning, algorithm, automatically, accuracy, automated, proposes, predict, dataset, neural_network, feasibility	0.7608
13	Computer vision	image, video, computer_vision, deep_learning, camera, visual, recognition, automatic, convolutional_neural_network, vision	0.8118
14	Interview	interview, delphi, structured, qualitative, semi, depth, theme, expert, thematic, practice	0.6505
15	Structural equation modeling	structural_equation_modeling, relationship, influence, mediating, effect, psychological, positively, moderating, role, stress	0.7377

According to the results in Table 5, some topics, such as *ergonomics* and *computer vision*, reached very high coherence scores (greater than 0.8), indicating that the top 10 words of these topics are highly coherent. Screening the top 10 words of these high-quality topics, it was found that many of these words were indeed semantically close to their topic names. For example, words such as 'musculoskeletal', 'ergonomic', 'posture', 'disorder', 'back', and 'pain' are all related to the *ergonomics* theme, and words such as 'image', 'video', 'deep_learning', 'camera', 'convolutional_neural_network', 'visual', 'vision', and 'recognition' are all related to the *computer vision* theme. Some other topics, such as *worker behavior* and *equipment contact*, reached lower coherence scores between 0.4 and 0.6. Despite the existence of some relatively low-quality topics, it is worth mentioning that the overall quality of topics generated by the Anchored CorEx model is still high.

3.4. Research Trends and Insights

Besides extracting top words and calculating the coherence score of each topic, the Anchored CorEx model is also able to categorize documents into different topics based on the calculated value of total correlation for each topic. Utilizing such categorization, the proportion of topics was analyzed to understand the topic trend evolution over time. Results that display the proportional evolution of different topics are plotted in Figure 5. Mann–Kendall statistical test [46,47] was used to quantitatively test the significance level of each topic's trend, and results are shown in Table 6.

Table 6. Results of Mann–Kendall statistical test and chronological topic proportion (in percentage).

Topic Name	Mann-Kendall Results	≤2000	2001–2005	2006–2008	2009–2011	2012–2013	2014–2015	2016–2017	2018	2019	2020	2021	2022
Object tracking	increasing	2.3	2.4	2.4	4.0	7.0	6.5	6.6	6.8	9.0	7.5	6.7	7.3
Safety climate	no trend	1.7	2.9	5.9	5.9	4.8	4.1	5.5	2.9	7.6	5.5	4.5	4.1
Ergonomics	decreasing	11.3	4.8	6.5	5.9	3.5	4.1	3.8	5.5	2.5	4.4	2.2	3.0
Design	no trend	4.5	8.1	7.7	4.5	7.5	8.7	5.4	4.4	5.3	4.8	4.7	5.5
Fall	decreasing	15.3	16.2	10.7	8.8	10.6	9.6	10.0	7.8	8.4	7.5	8.5	6.7
Safety training	no trend	6.8	9.5	10.7	9.9	9.8	5.7	6.4	5.2	7.2	6.6	6.8	7.6
Health and disease	decreasing	29.4	17.6	8.3	12.2	6.5	6.3	4.9	3.4	2.1	3.3	2.7	4.8
Worker behavior	increasing	4.5	7.6	8.3	7.9	7.0	7.0	8.1	9.9	8.0	8.1	9.4	9.4
Equipment contact	no trend	5.1	6.2	4.7	7.4	7.3	8.7	6.0	5.5	4.9	5.7	5.2	3.5
Real-time technology	increasing	0.6	1.9	2.4	3.7	8.0	7.2	7.4	7.0	8.8	7.9	8.1	7.8
Survey	no trend	13.6	12.4	15.4	13.0	7.8	8.9	10.4	12.0	8.6	10.6	10.6	10.1
Machine learning	increasing	0.0	3.3	3.6	4.5	7.8	9.8	10.1	12.2	12.7	11.9	13.2	10.8
Computer vision	increasing	1.7	0.0	3.0	2.5	2.3	3.9	4.1	6.0	5.7	6.0	6.5	7.3
Interview	no trend	3.4	5.7	7.7	7.4	7.3	6.1	5.8	5.5	3.5	3.7	4.0	5.5
Structural equation modeling	increasing	0.0	1.4	3.0	2.3	2.8	3.5	5.4	6.0	5.7	6.6	7.0	6.6

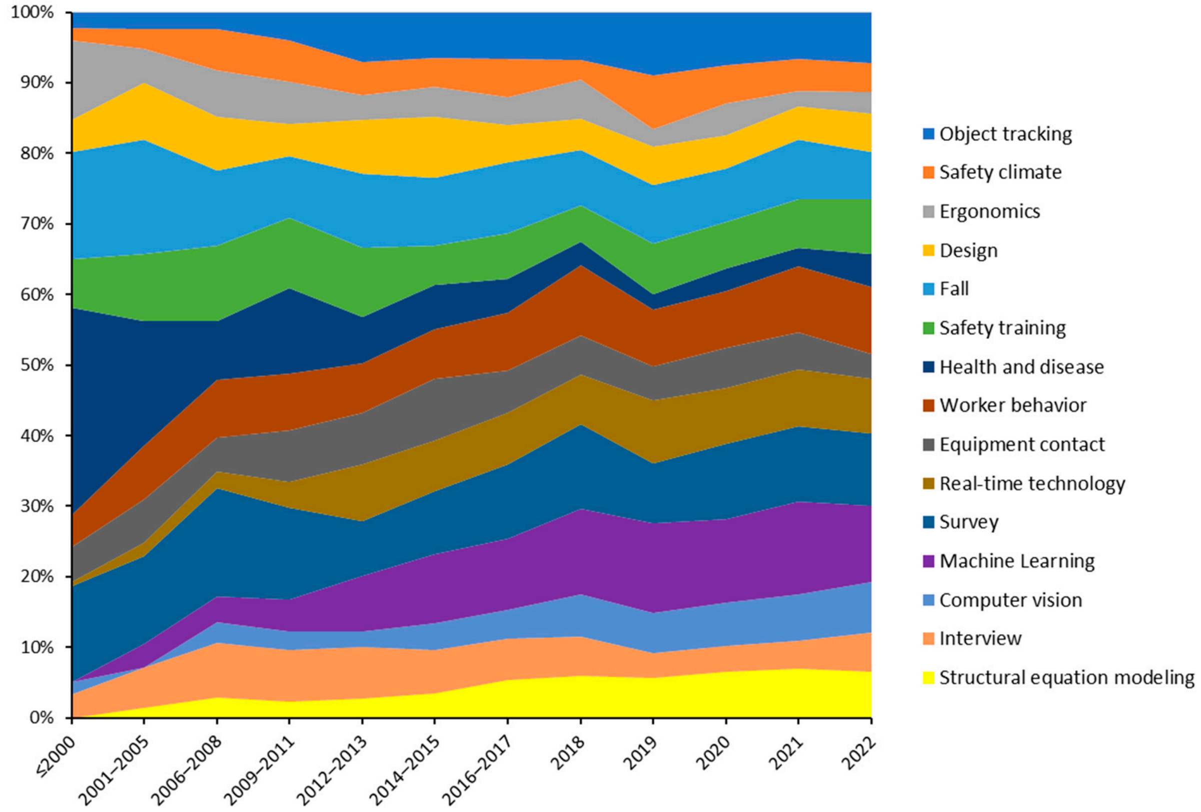
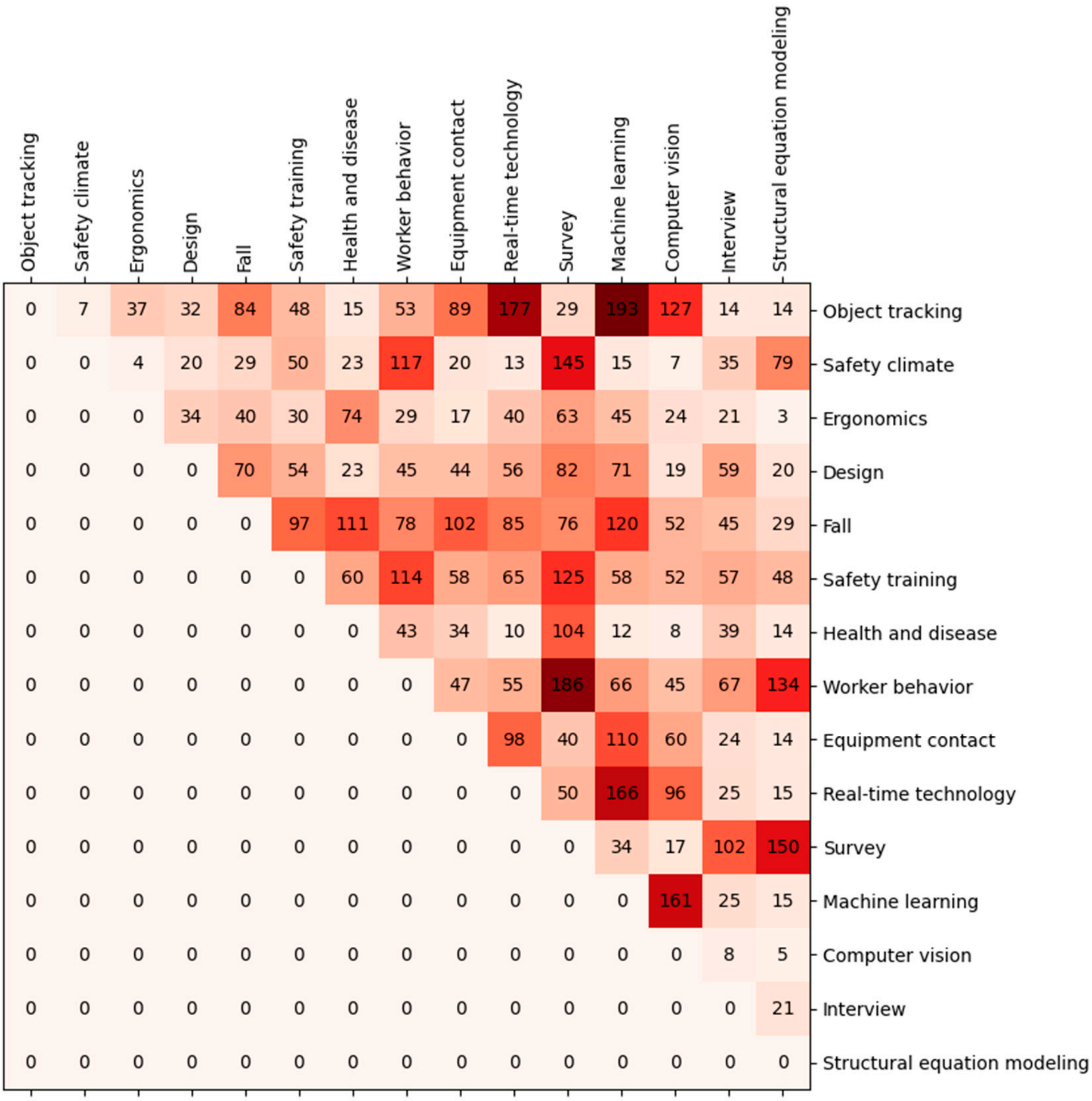


Figure 5. Topic proportion over time based on the Anchored CorEx model.

It is possible that a document can be related to multiple topics; the topic modeling technique is capable of capturing those mixed-topic documents by labeling them with more than one topic. Leveraging this feature, the inter-relationship (co-occurrence) between different topics can be analyzed to explore how different topics have interacted with each other and understand the hotspots of applying certain techniques to solving construction safety and health research problems. A co-occurrence matrix (M) was created and displayed in Figure 6, and $M_{i,j}$ denotes the number of documents that were both classified as the ith and jth topic. Note that some documents were labeled with only one topic, and some others were labeled with multiple topics. By looking into some of the global hotspots (matrix element value greater than 150), $M_{1,10}$ and $M_{1,12}$ denote that *real-time technologies* and *machine learning* techniques have been frequently utilized for *object-tracking* purposes; $M_{8,11}$ denotes that *survey* and questionnaire has often been utilized to understand *worker behavior*; $M_{10,12}$ denotes that *real-time technologies* have been regularly used in combination with *machine learning* techniques; and $M_{12,13}$ denotes that *machine learning* techniques have been commonly adopted in the *computer vision* domain.

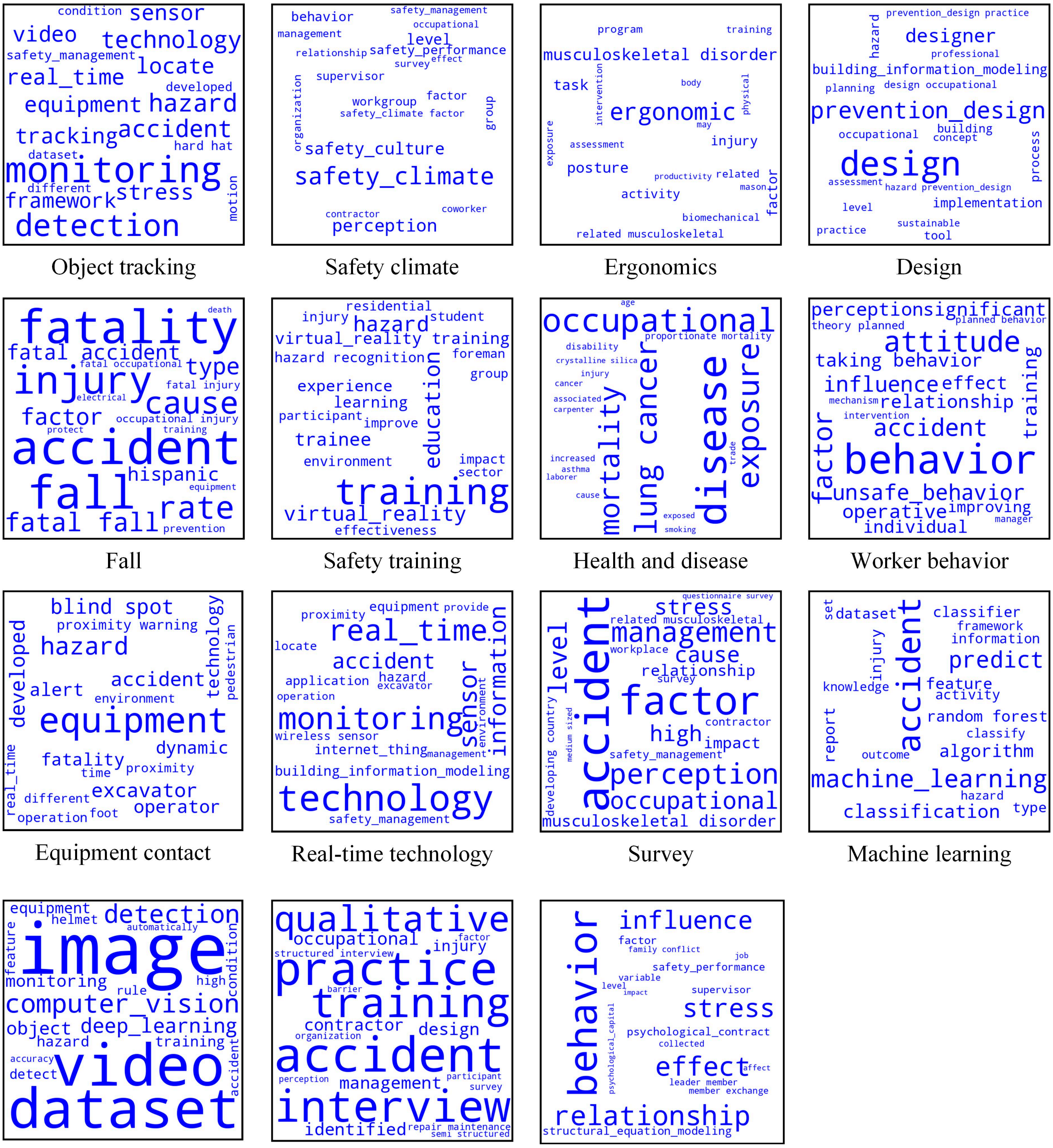
Results of this study suggest that there is increasing popularity in applying cutting-edge technologies, e.g., *object tracking*, *real-time technology*, *machine learning*, *computer vision*, and *structural equation modeling*, to construction safety and health research. To further explore the hotspots and the inter-relationship of these topics, the top ten documents of each of these topics were shortlisted and investigated for more details. *Object tracking* was primarily conducted with the help of sensors [48,49], computer vision techniques [50–52], and other real-time technologies [53]; *real-time technologies* have been widely adopted for various safety-related monitoring tasks on the job site [48,54–59]; *machine learning* (or deep learning) algorithms have been commonly utilized to understand and classify accident narratives or safety reports [60–63] and predict injury severity and outcomes [64–67]; *computer vision* techniques have been frequently leveraged for safety equipment monitoring [68–70] and unsafe behavior detection [71–73]; and *structural equation modeling* has been generally adopted to understand the relationship between safety and health performance and various other factors like stress [74,75], safety culture and climate [76,77], psychological capital [77,78], and psychological contract [79,80].



	Object tracking	Safety climate	Ergonomics	Design	Fall	Safety training	Health and disease	Worker behavior	Equipment contact	Real-time technology	Survey	Machine learning	Computer vision	Interview	Structural equation modeling
Object tracking	0	7	37	32	84	48	15	53	89	177	29	193	127	14	14
Safety climate	0	0	4	20	29	50	23	117	20	13	145	15	7	35	79
Ergonomics	0	0	0	34	40	30	74	29	17	40	63	45	24	21	3
Design	0	0	0	0	70	54	23	45	44	56	82	71	19	59	20
Fall	0	0	0	0	0	97	111	78	102	85	76	120	52	45	29
Safety training	0	0	0	0	0	0	60	114	58	65	125	58	52	57	48
Health and disease	0	0	0	0	0	0	0	43	34	10	104	12	8	39	14
Worker behavior	0	0	0	0	0	0	0	0	47	55	186	66	45	67	134
Equipment contact	0	0	0	0	0	0	0	0	0	98	40	110	60	24	14
Real-time technology	0	0	0	0	0	0	0	0	0	0	50	166	96	25	15
Survey	0	0	0	0	0	0	0	0	0	0	0	34	17	102	150
Machine learning	0	0	0	0	0	0	0	0	0	0	0	0	161	25	15
Computer vision	0	0	0	0	0	0	0	0	0	0	0	0	0	8	5
Interview	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
Structural equation modeling	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 6. Inter-relationship (co-occurrence) heat map between different topics.

Despite the rapid growth and increasing popularity of the above-mentioned emerging topics, many of the traditional topics have not lost their momentum yet. For example, *safety climate* has been continuously recognized as a critical factor in safety performance [81,82] and has influenced workers' behaviors through organizational structure [83,84] and supervisors' management [85,86]; *worker behavior* research has been primarily focused on identifying different factors like cognitive factors [87], psychological drivers [78,88], organizational impact [89] and occupational stress [90,91] that have impacted workers' safety behaviors. Finally, to discover each topic in more detail, the word cloud plot, a visual representation of the 20 most important words, was generated based on the top 30 documents of each topic and displayed in Figure 7.


condition sensor video technology safety_management locate real_time developed equipment hazard tracking accident hard hat dataset monitoring different stress framework motion detection
behavior safety_management occupational management level relationship safety_performance survey effect supervisor workgroup factor organization safety_climate factor group safety_culture safety_climate contractor coworker perception
program training musculoskeletal disorder task body intervention physical ergonomic may exposure assessment injury posture productivity related mason activity factor biomechanical related musculoskeletal
prevention_design practice hazard designer professional building_information_modeling planning design occupational prevention_design occupational building concept design process assessment hazard prevention_design level implementation practice sustainable tool
death fatality fatal accident fatal occupational type injury fatal injury electrical cause factor occupational injury protect training accident hispanic fall equipment rate fatal fall prevention
residential injury student hazard virtual_reality training hazard recognition foreman group experience learning participant improve education trainee environment impact sector training virtual_reality effectiveness
age occupational proportionate mortality disability crystalline silica injury cancer associated carpenter increased asthma laborer cause mortality lung cancer trade disease exposure exposed smoking
perception significant theory planned planned behavior attitude taking behavior training influence effect mechanism relationship intervention accident factor behavior unsafe_behavior operative improving individual manager
blind spot proximity warning technology hazard pedestrian developed alert accident environment equipment dynamic fatality time proximity real_time different excavator operation foot operator
proximity equipment provide real_time locate accident application hazard excavator operation sensor environment information monitoring wireless sensor internet_thing management building_information_modeling technology safety_management
questionnaire survey stress related musculoskeletal management workplace cause relationship survey level accident factor contractor high impact medium sized safety_management developing country perception occupational musculoskeletal disorder
set dataset classifier framework information injury accident predict knowledge feature activity random forest report classify outcome algorithm machine_learning hazard classification type
equipment helmet detection automatically feature image condition monitoring rule high computer_vision object deep_learning hazard training accuracy detect video accident dataset
qualitative occupational factor injury structured interview practice barrier training contractor design organization accident perception management participant survey interview identified repair maintenance semi structured
behavior influence factor family conflict job safety_performance variable level impact supervisor stress psychological_capital psychological_contract collected effect affect leader member member exchange relationship structural_equation_modeling

Object tracking

Safety climate

Ergonomics

Design

Fall

Safety training

Health and disease

Worker behavior

Equipment contact

Real-time technology

Survey

Machine learning

Computer vision

Interview

Structural equation modeling

Figure 7. Word cloud of each topic.

4. Conclusions and Future Work

This paper applied topic modeling techniques to discover the research topics and trends in the field of construction safety and health. It focused on 1984 articles that were published until February 2023 from 27 journal sources included in the Web of Science database. After text mining the original data, two unsupervised learning models (LDA and CorEx), as well as their corresponding semi-supervised versions (Guided LDA and

Anchored CorEx), were built. Semi-supervised learning models have outperformed unsupervised ones by achieving a higher coherence score and demonstrating better human interpretability, and the Anchored CorEx model has achieved the best performance among all four models.

Based on the results of the Anchored CorEx model, trend evolution analysis was conducted to explore how those topics have proportionally changed over the past three decades, and an interconnection heat map was plotted to discover how those topics are interrelated. Top-listed articles of major topics were reviewed, and the word cloud of each topic was plotted to better understand the research focus of each topic. Clear interrelationship patterns were found between object tracking and real-time technology, object tracking and machine learning, survey and worker behavior, real-time technology and machine learning, and machine learning and computer vision.

Such cutting-edge technologies as object tracking, real-time technology, machine learning, computer vision, and structural equation modeling have been increasingly applied to construction safety and health research to form emerging research topics. At the same time, several traditional topics have not lost their momentum, among which safety climate and worker behavior have been continuously studied by researchers at both organizational and individual levels from various perspectives.

The contributions of this paper are two-fold. First, it applied topic modeling techniques to discover the research topics and trends in the construction safety and health field and provided guidance for future research directions in this domain. It helped answer research questions such as ‘what themes and topics are researchers in this field interested in?’, ‘how did these research topics evolve over time?’, ‘how are these research topics interrelated?’, and ‘what is the main research focus of each topic?’. Second, different from existing related works in other domains that only applied unsupervised learning models such as LDA, the proposed research introduced two semi-supervised learning models to guide the learning process using pre-determined topic priors, thus offering more topic generation flexibility and providing higher quality analytical results.

In short, the main contributions of this work are the following:

1. The application of NLP techniques on a large dataset containing the titles, keywords, and abstracts of research papers from the construction safety and health domain;
2. The comparison between the performances of four different topic models, two unsupervised ones and two semi-supervised ones, and the identification of the best-performing model;
3. The identification of research topics in the construction safety and health domain and the analysis of their chronological trends over the past three decades;
4. The discovery of the interconnection between different research topics as well as the discussion on their standalone most representative publications.

One possible future direction of this study is expanding the current dataset that contains title, keywords, and abstract to include the full text of each article because topic modeling techniques generally perform better for large texts than for short texts, providing a potential tradeoff between even higher topic quality and shorter computational time.

Author Contributions: Conceptualization, K.Z. and J.W.; methodology, K.Z. and J.W.; software, K.Z.; validation, J.W., B.A. and J.C.; formal analysis, K.Z., J.W. and J.C.; data curation, K.Z. and J.W.; writing—original draft preparation, K.Z. and J.W.; writing—review and editing, K.Z., J.W., B.A. and J.C.; visualization, K.Z. and J.W.; supervision, J.W., B.A. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. ENR. Construction Jobsite Deaths, Fatality Rate Climb, Engineering News-Record. Available online: https://www.enr.com/articles/50908-construction-jobsite-deaths-fatality-rate-climb (accessed on 22 June 2022).
2. MEM PRC. Announcement on the Safety Situation of the Domestic Construction Industry in H1 2018, Ministry of Emergency Management of the People's Republic of China. Available online: https://www.mem.gov.cn/gk/tzgg/tb/201807/t20180725_230568.shtml (accessed on 22 June 2022).
3. HSE. Construction Statistics in Great Britain, 2020, Health and Safety Executive. Available online: https://www.hse.gov.uk/statistics/industry/construction.pdf (accessed on 22 June 2022).
4. Zhou, Z.; Irizarry, J.; Li, Q. Applying advanced technology to improve safety management in the construction industry: A literature review. *Constr. Manag. Econ.* **2013**, *31*, 606–622. [CrossRef]
5. Skibniewski, M. Research Trends in Information Technology Applications in Construction Safety Engineering and Management. *Front. Eng. Manag.* **2014**, *1*, 246–259. [CrossRef]
6. Alruqi, W.M.; Hallowell, M.R.; Techera, U. Safety climate dimensions and their relationship to construction safety performance: A meta-analytic review. *Saf. Sci.* **2018**, *109*, 165–173. [CrossRef]
7. 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]
8. Carnot, M.L.; Bernardino, J.; Laranjeiro, N.; Oliveira, H.G. Applying Text Analytics for Studying Research Trends in Dependability. *Entropy* **2020**, *22*, 1303. [CrossRef] [PubMed]
9. Choi, H.S.; Lee, W.S.; Sohn, S.Y. Analyzing research trends in personal information privacy using topic modeling. *Comput. Secur.* **2017**, *67*, 244–253. [CrossRef]
10. Jin, R.; Zou, P.X.; Piroozfar, P.; Wood, H.; Yang, Y.; Yan, L.; Han, Y. A science mapping approach based review of construction safety research. *Saf. Sci.* **2018**, *113*, 285–297. [CrossRef]
11. Vigneshkumar, C.; Salve, U.R. A scientometric analysis and review of fall from height research in construction. *Constr. Econ. Build.* **2020**, *20*, 17–35.
12. Wang, J.; Chen, J.; Hu, Y. A science mapping approach based review of model predictive control for smart building operation management. *J. Civ. Eng. Manag.* **2022**, *28*, 661–679. [CrossRef]
13. Sun, L.; Yin, Y. Discovering themes and trends in transportation research using topic modeling. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 49–66. [CrossRef]
14. Jiang, H.; Qiang, M.; Lin, P. A topic modeling based bibliometric exploration of hydropower research. *Renew. Sustain. Energy Rev.* **2016**, *57*, 226–237. [CrossRef]
15. Hall, D.; Jurafsky, D.; Manning, C.D. Studying the history of ideas using topic models. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu HI, USA, 25–27 October 2008; pp. 363–371.
16. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (Suppl. 1), 5228–5235. [CrossRef] [PubMed]
17. Weng, J.; Lim, E.P.; Jiang, J.; He, Q. Twitterrank: Finding topic-sensitive influential twitterers. In Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, NY, USA, 3–6 February 2010; pp. 261–270.
18. Roberts, K.; Roach, M.A.; Johnson, J.; Guthrie, J.; Harabagiu, S.M. EmpaTweet: Annotating and Detecting Emotions on Twitter. *Lrec* **2012**, *12*, 3806–3813.
19. Fang, Y.; Si, L.; Somasundaram, N.; Yu, Z. Mining contrastive opinions on political texts using cross-perspective topic model. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, Seattle, WA, USA, 8–12 February 2012; pp. 63–72.
20. Levy, K.E.; Franklin, M. Driving regulation: Using topic models to examine political contention in the US trucking industry. *Soc. Sci. Comput. Rev.* **2014**, *32*, 182–194. [CrossRef]
21. Zhang, Y.; Chen, M.; Huang, D.; Wu, D.; Li, Y. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Gener. Comput. Syst.* **2017**, *66*, 30–35. [CrossRef]
22. Paul, M.; Dredze, M. You are what you tweet: Analyzing twitter for public health. In Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5, No. 1.
23. Jallan, Y.; Brogan, E.; Ashuri, B.; Clevenger, C.M. Application of Natural Language Processing and Text Mining to Identify Patterns in Construction-Defect Litigation Cases. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2019**, *11*, 04519024. [CrossRef]
24. Xue, J.; Shen, G.Q.; Li, Y.; Wang, J.; Zafar, I. Dynamic Stakeholder-Associated Topic Modeling on Public Concerns in Megainfrastructure Projects: Case of Hong Kong–Zhuhai–Macao Bridge. *J. Manag. Eng.* **2020**, *36*, 04020078. [CrossRef]
25. Yau, C.-K.; Porter, A.; Newman, N.; Suominen, A. Clustering scientific documents with topic modeling. *Scientometrics* **2014**, *100*, 767–786. [CrossRef]
26. Amado, A.; Cortez, P.; Rita, P.; Moro, S. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manag. Bus. Econ.* **2017**, *24*, 1–7. [CrossRef]
27. Chen, X.; Zou, D.; Cheng, G.; Xie, H. Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Comput. Educ.* **2020**, *151*, 103855. [CrossRef]
28. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

29. Gallagher, R.J.; Reing, K.; Kale, D.; Steeg, G.V. Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge. *Trans. Assoc. Comput. Linguistics* **2017**, *5*, 529–542. [CrossRef]
30. Ver Steeg, G.; Galstyan, A. Discovering structure in high-dimensional data through correlation explanation. *Adv. Neural Inf. Process. Syst.* **2014**, 27. [CrossRef]
31. Jagarlamudi, J.; Daumé, H., III; Udupa, R. Incorporating lexical priors into topic models. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, 23–27 April 2012; pp. 204–213.
32. Bevilacqua, M.; Ciarapica, F.E.; Giacchetta, G. Industrial and occupational ergonomics in the petrochemical process industry: A regression trees approach. *Accid. Anal. Prev.* **2008**, *40*, 1468–1479. [CrossRef] [PubMed]
33. Witter, R.Z.; Tenney, L.; Clark, S.; Newman, L.S. Occupational exposures in the oil and gas extraction industry: State of the science and research recommendations. *Am. J. Ind. Med.* **2014**, *57*, 847–856. [CrossRef]
34. Luo, X.; Gan, W.; Wang, L.; Chen, Y.; Meng, X. A Prediction Model of Structural Settlement Based on EMD-SVR-WNN. *Adv. Civ. Eng.* **2020**, *2020*, 8831965. [CrossRef]
35. Han, F.; Zhang, Q.; Wang, C.; Lu, G.; Jiang, J. Structural Health Monitoring of Timber Using Electromechanical Impedance (EMI) Technique. *Adv. Civ. Eng.* **2020**, *2020*, 1906289. [CrossRef]
36. Syamlal, G.; King, B.A.; Mazurek, J.M. Tobacco product use among workers in the construction industry, United States, 2014-2016. *Am. J. Ind. Med.* **2018**, *61*, 939–951. [CrossRef]
37. Lipscomb, H.J.; Dement, J.M.; Li, L. Health care utilization of families of carpenters with alcohol or substance abuse-related diagnoses. *Am. J. Ind. Med.* **2003**, *43*, 361–368. [CrossRef]
38. Neumann, M.; King, D.; Beltagy, I.; Ammar, W. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. *arXiv* **2019**, arXiv:1902.07669. [CrossRef]
39. Available online: https://github.com/hyperreality/American-British-English-Translator (accessed on 22 February 2023).
40. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2009.
41. Church, K.; Hanks, P. Word association norms, mutual information, and lexicography. *Comput. Linguist.* **1990**, *16*, 22–29.
42. Chang, J.; Boyd-Graber, J.; Wang, C.; Gerrish, S.; Blei, D.M. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*; 2009; Volume 22, pp. 288–296. Available online: https://papers.nips.cc/paper_files/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html (accessed on 22 February 2023).
43. Zou, C. Analyzing research trends on drug safety using topic modeling. *Expert Opin. Drug Saf.* **2018**, *17*, 629–636. [CrossRef] [PubMed]
44. Röder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015; pp. 399–408.
45. Laufer, A.; Ledbetter, W.B. Assessment of Safety Performance Measures at Construction Sites. *J. Constr. Eng. Manag.* **1986**, *112*, 530–542. [CrossRef]
46. Mann, H.B. Nonparametric tests against trend. *Econom. J. Econom. Soc.* **1945**, 245–259. [CrossRef]
47. Hussain, M.M.; Mahmud, I. pyMannKendall: A python package for non parametric Mann Kendall family of trend tests. *J. Open Source Softw.* **2019**, *4*, 1556. [CrossRef]
48. Park, J.W.; Kim, K.; Cho, Y.K. Framework of Automated Construction-Safety Monitoring Using Cloud-Enabled BIM and BLE Mobile Tracking Sensors. *J. Constr. Eng. Manag.* **2017**, *143*, 05016019. [CrossRef]
49. Park, J.; Cho, Y.K.; Khodabandelu, A. Sensor-Based Safety Performance Assessment of Individual Construction Workers. *Sensors* **2018**, *18*, 3897. [CrossRef]
50. Han, S.; Lee, S. A vision-based motion capture and recognition framework for behavior-based safety management. *Autom. Constr.* **2013**, *35*, 131–141. [CrossRef]
51. Tang, S.; Golparvar-Fard, M.; Naphade, M.; Gopalakrishna, M.M. Video-Based Motion Trajectory Forecasting Method for Proactive Construction Safety Monitoring Systems. *J. Comput. Civ. Eng.* **2020**, *34*, 04020041. [CrossRef]
52. Seo, J.; Han, S.; Lee, S.; Kim, H. Computer vision techniques for construction safety and health monitoring. *Adv. Eng. Inform.* **2015**, *29*, 239–251. [CrossRef]
53. Ding, L.; Zhou, C.; Deng, Q.; Luo, H.; Ye, X.; Ni, Y.; Guo, P. Real-time safety early warning system for cross passage construction in Yangtze Riverbed Metro Tunnel based on the internet of things. *Autom. Constr.* **2013**, *36*, 25–37. [CrossRef]
54. Cheng, T.; Teizer, J. Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. *Autom. Constr.* **2012**, *34*, 3–15. [CrossRef]
55. Cheung, W.-F.; Lin, T.-H.; Lin, Y.-C. A Real-Time Construction Safety Monitoring System for Hazardous Gas Integrating Wireless Sensor Network and Building Information Modeling Technologies. *Sensors* **2018**, *18*, 436. [CrossRef] [PubMed]
56. Riaz, Z.; Parn, E.A.; Edwards, D.J.; Arslan, M.; Shen, C.; Pena-Mora, F. BIM and sensor-based data management system for construction safety monitoring. *J. Eng. Des. Technol.* **2017**, *15*, 738–753. [CrossRef]
57. Khan, M.; Khalid, R.; Anjum, S.; Tran SV, T.; Park, C. Fall prevention from scaffolding using computer vision and IoT-based monitoring. *J. Constr. Eng. Manag.* **2022**, *148*, 04022051. [CrossRef]
58. Zhou, C.; Luo, H.; Fang, W.; Wei, R.; Ding, L. Cyber-physical-system-based safety monitoring for blind hoisting with the internet of things: A case study. *Autom. Constr.* **2018**, *97*, 138–150. [CrossRef]

59. Wang, J.; Zhang, S.; Teizer, J. Geotechnical and safety protective equipment planning using range point cloud data and rule checking in building information modeling. *Autom. Constr.* **2015**, *49*, 250–261. [CrossRef]
60. Qiao, J.; Wang, C.; Guan, S.; Shuran, L. Construction-Accident Narrative Classification Using Shallow and Deep Learning. *J. Constr. Eng. Manag.* **2022**, *148*, 04022088. [CrossRef]
61. Goh, Y.M.; Ubeynarayana, C. Construction accident narrative classification: An evaluation of text mining techniques. *Accid. Anal. Prev.* **2017**, *108*, 122–130. [CrossRef]
62. Zhong, B.; Pan, X.; Love, P.E.; Ding, L.; Fang, W. Deep learning and network analysis: Classifying and visualizing accident narratives in construction. *Autom. Constr.* **2020**, *113*, 103089. [CrossRef]
63. Fang, W.; Luo, H.; Xu, S.; Love, P.E.; Lu, Z.; Ye, C. Automated text classification of near-misses from safety reports: An improved deep learning approach. *Adv. Eng. Inform.* **2020**, *44*, 101060. [CrossRef]
64. Sarkar, S.; Pramanik, A.; Maiti, J.; Reniers, G. Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Saf. Sci.* **2020**, *125*, 104616. [CrossRef]
65. Baker, H.; Hallowell, M.R.; Tixier, A.J.-P. AI-based prediction of independent construction safety outcomes from universal attributes. *Autom. Constr.* **2020**, *118*, 103146. [CrossRef]
66. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Application of machine learning to construction injury prediction. *Autom. Constr.* **2016**, *69*, 102–114. [CrossRef]
67. Koc, K.; Ekmekcioğlu, Ö.; Gurgun, A.P. Prediction of construction accident outcomes based on an imbalanced dataset through integrated resampling techniques and machine learning methods. *Eng. Constr. Arch. Manag.* 2022, *ahead-of-print*. [CrossRef]
68. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* **2020**, *112*, 103085. [CrossRef]
69. Mneymneh, B.E.; Abbas, M.; Khoury, H. Vision-Based Framework for Intelligent Monitoring of Hardhat Wearing on Construction Sites. *J. Comput. Civ. Eng.* **2019**, *33*, 04018066. [CrossRef]
70. Li, Y.; Wei, H.; Han, Z.; Huang, J.; Wang, W. Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks. *Adv. Civ. Eng.* **2020**, *2020*, 9703560. [CrossRef]
71. Luo, H.; Wang, M.; Wong, P.K.-Y.; Cheng, J.C. Full body pose estimation of construction equipment using computer vision and deep learning techniques. *Autom. Constr.* **2019**, *110*, 103016. [CrossRef]
72. Wang, Y.; Liao, P.-C.; Zhang, C.; Ren, Y.; Sun, X.; Tang, P. Crowdsourced reliable labeling of safety-rule violations on images of complex construction scenes for advanced vision-based workplace safety. *Adv. Eng. Inform.* **2019**, *42*, 101001. [CrossRef]
73. Ding, L.; Fang, W.; Luo, H.; Love, P.E.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [CrossRef]
74. Liang, Q.; Zhou, Z.; Li, X.; Hu, Q.; Ye, G. Revealing the mechanism of stress generation for construction frontline professionals through development of structural stressors–coping–stress models. *Saf. Sci.* **2022**, *150*, 105708. [CrossRef]
75. Liu, Q.; Feng, Y.; London, K.; Zhang, P. Influence of personal characteristics and environmental stressors on mental health for multicultural construction workplaces in Australia. *Constr. Manag. Econ.* **2022**, *41*, 116–137. [CrossRef]
76. Danso, F.O.; Adinyira, E.; Manu, P.; Agyekum, K.; Ahadzie, D.K.; Badu, E. The mediating influence of local cultures on the relationship between factors of safety risk perception and Risk-Taking behavioural intention of construction site workers. *Saf. Sci.* **2021**, *145*, 105490. [CrossRef]
77. He, C.; McCabe, B.; Jia, G. Effect of leader-member exchange on construction worker safety behavior: Safety climate and psychological capital as the mediators. *Saf. Sci.* **2021**, *142*, 105401. [CrossRef]
78. He, C.; Jia, G.; McCabe, B.; Chen, Y.; Sun, J. Impact of psychological capital on construction worker safety behavior: Communication competence as a mediator. *J. Saf. Res.* **2019**, *71*, 231–241. [CrossRef]
79. Liang, H.; Shi, X.; Yang, D.; Liu, K. Impact of mindfulness on construction workers' safety performance: The mediating roles of psychological contract and coping behaviors. *Saf. Sci.* **2021**, *146*, 105534. [CrossRef]
80. Newaz, M.T.; Davis, P.; Jefferies, M.; Pillay, M. Examining the Psychological Contract as Mediator between the Safety Behavior of Supervisors and Workers on Construction Sites. *J. Constr. Eng. Manag.* **2020**, *146*, 04019094. [CrossRef]
81. Chen, Y.; McCabe, B.; Hyatt, D. Impact of individual resilience and safety climate on safety performance and psychological stress of construction workers: A case study of the Ontario construction industry. *J. Saf. Res.* **2017**, *61*, 167–176. [CrossRef] [PubMed]
82. Chen, Y.; McCabe, B.; Hyatt, D. A resilience safety climate model predicting construction safety performance. *Saf. Sci.* **2018**, *109*, 434–445. [CrossRef]
83. Gao, R.; Chan, A.P.; Utama, W.P.; Zahoor, H. Multilevel Safety Climate and Safety Performance in the Construction Industry: Development and Validation of a Top-Down Mechanism. *Int. J. Environ. Res. Public Health* **2016**, *13*, 1100. [CrossRef]
84. Chen, Q.; Jin, R. Multilevel Safety Culture and Climate Survey for Assessing New Safety Program. *J. Constr. Eng. Manag.* **2013**, *139*, 805–817. [CrossRef]
85. Chan, D.W.M.; Cristofaro, M.; Nassereddine, H.; Yiu, N.S.N.; Sarvari, H. Perceptions of Safety Climate in Construction Projects between Workers and Managers/Supervisors in the Developing Country of Iran. *Sustainability* **2021**, *13*, 10398. [CrossRef]
86. Törner, M.; Pousette, A. Safety in construction–a comprehensive description of the characteristics of high safety standards in construction work, from the combined perspective of supervisors and experienced workers. *J. Saf. Res.* **2009**, *40*, 399–409. [CrossRef] [PubMed]

87. Goh, Y.M.; Binte Sa'adon, N.F. Cognitive factors influencing safety behavior at height: A multimethod exploratory study. *J. Constr. Eng. Manag.* **2015**, *141*, 04015003. [CrossRef]
88. Liu, Q.; Xu, N.; Jiang, H.; Wang, S.; Wang, W.; Wang, J. Psychological Driving Mechanism of Safety Citizenship Behaviors of Construction Workers: Application of the Theory of Planned Behavior and Norm Activation Model. *J. Constr. Eng. Manag.* **2020**, *146*, 04020027. [CrossRef]
89. Man, S.S.; Chan, A.H.S.; Alabdulkarim, S.; Zhang, T. The effect of personal and organizational factors on the risk-taking behavior of Hong Kong construction workers. *Saf. Sci.* **2021**, *136*, 105155. [CrossRef]
90. Liang, Q.; Zhou, Z.; Ye, G.; Shen, L. Unveiling the mechanism of construction workers' unsafe behaviors from an occupational stress perspective: A qualitative and quantitative examination of a stress–cognition–safety model. *Saf. Sci.* **2021**, *145*, 105486. [CrossRef]
91. Chen, Y.; McCabe, B.; Hyatt, D. Relationship between individual resilience, interpersonal conflicts at work, and safety outcomes of construction workers. *J. Constr. Eng. Manag.* **2017**, *143*, 04017042. [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.