Artificial Intelligence in Geosciences 5 (2024) 100084

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

Water resource forecasting with machine learning and deep learning: A scientometric analysis

Chanjuan Liu [a], Jing Xu [b], Xi'an Li [c], Zhongyao Yu [a], Jinran Wu [d,*]

[a] School of Business Administration and Customs, Shanghai Customs College, Shanghai 201204, PR China
[b] College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou 225009, PR China
[c] Ceyear Technologies Co., Ltd., Qingdao 266555, PR China
[d] The Institute for Positive Psychology and Education, Australian Catholic University, Banyo 4014, Australia

ARTICLE INFO

Keywords:
Water forecasting
Machine learning/deep learning
Web of Science
Visualization

ABSTRACT

Water prediction plays a crucial role in modern-day water resource management, encompassing both hydrological patterns and demand forecasts. To gain insights into its current focus, status, and emerging themes, this study analyzed 876 articles published between 2015 and 2022, retrieved from the Web of Science database. Leveraging CiteSpace visualization software, bibliometric techniques, and literature review methodologies, the investigation identified essential literature related to water prediction using machine learning and deep learning approaches. Through a comprehensive analysis, the study identified significant countries, institutions, authors, journals, and keywords in this field. By exploring this data, the research mapped out prevailing trends and cutting-edge areas, providing valuable insights for researchers and practitioners involved in water prediction through machine learning and deep learning. The study aims to guide future inquiries by highlighting key research domains and emerging areas of interest.

Contents

* Corresponding author.
E-mail addresses: liuchanjuan@shcc.edu.cn (C. Liu), xujing7503@yzu.edu.cn (J. Xu), lixian@ceyear.com (X. Li), yuzhongyao@shcc.edu.cn (Z. Yu), ryan.wu@acu.edu.au (J. Wu).

https://doi.org/10.1016/j.aiig.2024.100084
Received 7 March 2024; Received in revised form 6 June 2024; Accepted 9 August 2024
Available online 10 August 2024

1. Introduction

Water resources play a vital role in the survival and progress of humanity. Therefore, forecasting water availability is essential for ensuring the sustainable management of this invaluable resource (Yaseen et al., 2019; Wei et al., 2018). Accurate predictions regarding water supply can significantly aid in disaster preparedness (Paudyal, 2002; Chang et al., 2014; Fang et al., 2013), agricultural scheduling (Brédy et al., 2020), efficient water resource allocation (Fullerton and Molina, 2010; Chang and Guo, 2020), and the protection of water ecosystems (Liao and Sun, 2010; Choi et al., 2023). Machine learning and deep learning techniques have emerged as potent tools for prediction, attracting significant attention in water forecasting endeavors, encompassing hydrological predictions and forecasting urban residential water consumption patterns (Latif and Ahmed, 2023).

Over the past decades, there has been a notable surge in documents and researchers exploring water forecasting utilizing machine learning and deep learning techniques, which are advancing rapidly (Wee et al., 2021). Presently, there exists a substantial body of literature and comprehensive overview materials within this domain. However, a comprehensive literature review analyzing the evolution of water forecasting using machine learning and deep learning is lacking. This review should encompass all available literature on the subject. In addition, a limited number of studies have successfully incorporated visualization software with literature reviews to explore the advancement of water forecasting within the context of machine learning and deep learning methodologies. To bridge this research gap, our article aims to comprehensively analyze the current focal areas and prevailing state of water forecasting by leveraging machine learning and deep learning techniques. By doing so, we explore the emerging trends and potential prospects in this field. To achieve this, we utilize bibliometric visualization tools to provide an immediate overview of research activities in this burgeoning field.

Thanks to advancements in information technology, a range of visualization tools, including Gephi, CiteSpace, and SCI2, are available for analysis. A complex system typically comprises numerous interconnected components, exhibiting emergent properties not readily apparent from the individual parts. Among bibliometric visualization tools, CiteSpace software stands out for its focus on layout algorithms and its ability to represent co-occurring networks that capture time, frequency, and intermediary centrality simultaneously. Notably, CiteSpace boasts user-friendly operation, robust functionality, and high objectivity (Chen, 2006, 2017, 2018). Hence, for this study, we employed CiteSpace software, developed by Professor Chen Chaomei, as the scientific mapping tool. Through CiteSpace, we aim to assess the current state, trends, and frontiers of research in water forecasting using machine learning and deep learning techniques. Additionally, we seek to identify core technologies and their co-citations, as well as pivotal literature within the domain from 2015 to 2022.

Our research endeavors to analyze documents related to water forecasting by employing machine learning and deep learning approaches utilizing online databases. To ensure thorough insights, we advocate for the incorporation of visualization technologies. Consequently, we utilize scientific knowledge mapping software, established literature measurement techniques, and information visualization tools to quantitatively analyze articles concerning water forecasting with machine learning and deep learning from 2015 to 2022. The specific issues addressed in this paper were to: (1) understand the quantitative changes in total publications and times cites on water forecasting with machine learning and deep learning, as well as distribution in country, institution, research area, and journal; (2) identify the clusters of network in the documents co-citation analysis, and detect the burst of citation; (3) investigate emerging topics of water forecasting with machine learning and deep learning and clusters of network with keywords co-occurrence analysis; (4) explore collaboration network of author and institution based on documents related to water forecasting with machine learning and deep learning.

The structure of this paper is as follows: In Section 2, we outline the methodologies and materials utilized throughout our research. Moving forward, Section 3 details the findings of our analysis. Section 4 then delves into a comprehensive review of water forecasting techniques employing machine learning and deep learning. Finally, Section 5 wraps up the paper with conclusive remarks and a summary of our conclusions.

2. Methods and materials

2.1. Research method

In this study, CiteSpace software was chosen for visualizing knowledge graphs in the analysis of time series forecasting employing machine learning and deep learning techniques in hydrology.

CiteSpace, developed by Dr. Chen Chaomei at Drexel University, is a Java application that is freely accessible (Chen, 2006). It is tailored for visualizing and analyzing trends in scientific literature, providing features to comprehend network and historical patterns. These include pinpointing emerging topics, identifying citation hotspots, and automatically clustering networks based on terms from citing articles. Additionally, CiteSpace allows for the examination of structural and temporal aspects across various networks derived from scientific publications, such as collaboration, author co-citation, and document co-citation networks. In our research, we employed CiteSpace (version R6.1.6) to visualize raw data and create a visual knowledge map, as illustrated in Fig. 1.

We initiated a project named “WF WOS” and imported 876 studies, including full records and cited references, into CiteSpace visualization software in plain text format. Duplicate entries were removed as part of this process.

Following this, parameters were configured, specifying the time interval from 2015 to 2022 with yearly time slices. The analysis focused on information from the title, abstract, author names, and keywords of the articles. Node types selected included country, institution, author, cited journal, reference, cited author, and keyword, with the top 10 most cited items chosen for each time slice. Additionally, the top 20% selection criteria were applied, and pathfinder and pruning section network pruning modes were utilized, along with cluster view-static mode for visualizing the merged network.

CiteSpace was then executed to extract data networks covering authors, institutions, and countries with significant publication volumes, widely cited references, influential authors and journals in the research field, and keywords appearing two or more times.

Finally, the obtained networks and data were analyzed to identify frontiers, key topics, development trends, and collaborative patterns within the domain of big data.

2.2. Materials

The ISI Web of Science database serves as a standardized indexing tool for assessing the reliability of machine learning and deep learning research in water forecasting. To ensure the accuracy and comprehensiveness of our analysis, we conducted an advanced search within the Web of Science platform. Our search criteria included terms such as “Machine Learning”, “Deep Learning”, “statistical”, and “neural networks” in conjunction with keywords related to water, floods, hydrology, and forecasting. We specified English as the language and selected “article” as the document type. The search was constrained to the period from 2015 to 2022.

Following this retrieval method, we meticulously reviewed article contents and identified 876 relevant articles. The output files were saved in the format “download_*.txt” after renaming.

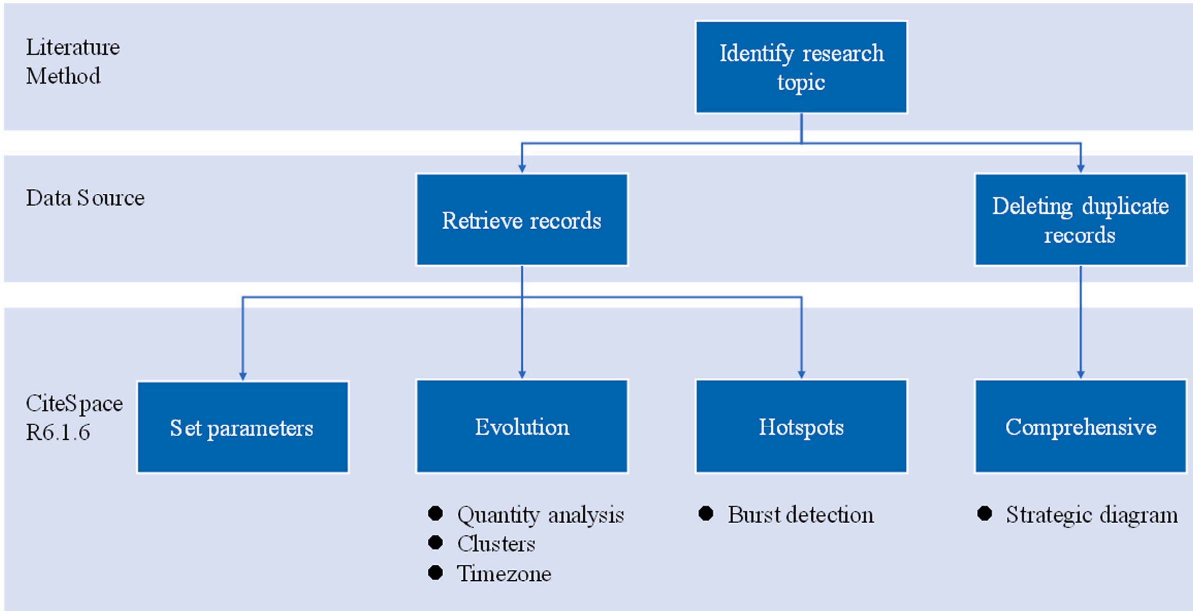

Fig. 1. The research processes.

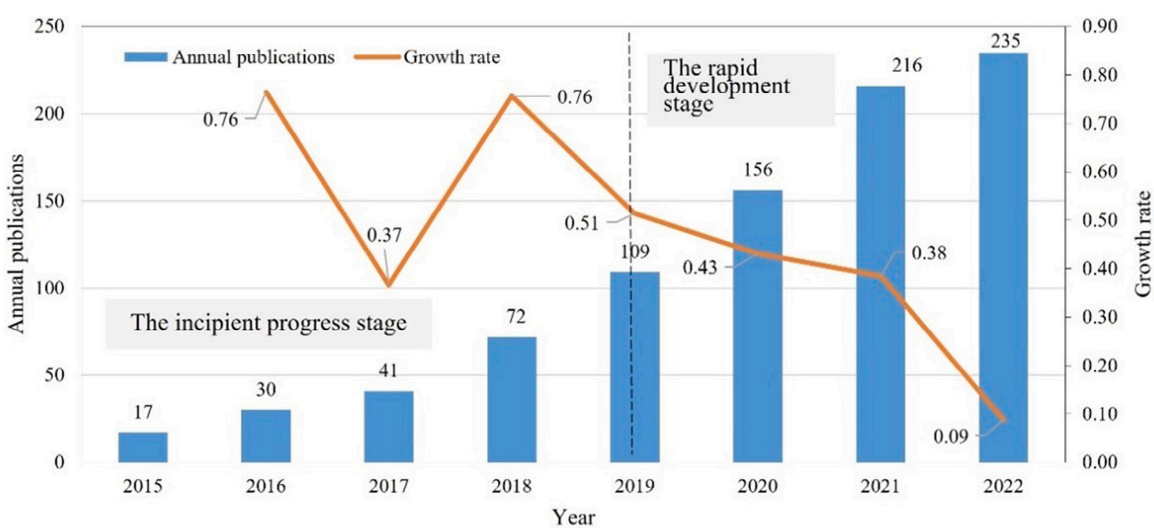

Fig. 2. Annual publication statistics of water forecasting with machine learning/deep learning research (2015–2022).

3. Results

3.1. Publishing outputs overall characteristics

One approach to assess the research progress within a discipline is by quantifying the volume of academic literature in the respective field. Plotting the time distribution graph of literature quantity serves as an effective means to evaluate the research status within the discipline, enabling further prediction of its developmental dynamics and trends (Bicheng et al., 2023; Sun et al., 2020).

As depicted in Fig. 2, the period from 2015 to 2022 witnessed two distinct stages in the water forecasting research domain employing machine learning and deep learning techniques: an initial development phase spanning from 2015 to 2018, followed by a period of rapid growth from 2019 to 2022. Throughout this timeframe, the number of articles published in this field exhibited a consistent upward trajectory, culminating in a peak of 235 publications in 2022. However, it is notable that since 2018, the growth rate of publication numbers has gradually declined on an annual basis.

3.2. Collaboration network analysis

Analyzing collaboration can enhance our understanding of the collaborative network, revealing the key countries, organizations, and authors making significant contributions in the field of water forecasting utilizing machine learning and deep learning techniques (Wu et al., 2020; Zhang et al., 2020).

3.2.1. Country collaboration network analysis

In the CiteSpace analysis software, we utilized the “Country” parameter to generate collaboration network maps depicting countries engaged in water forecasting research utilizing machine learning and deep learning techniques, as shown in Fig. 3. In the graph, nodes represent different countries cited in the documents, differentiated by colors and sizes. Larger nodes indicate a greater volume of associated papers, while smaller nodes signify fewer papers.

As depicted in Fig. 3, the top ten nations with the highest number of publications include the United States, China, Iran, India, England, South Korea, Vietnam, Australia, Canada, and Spain. The size of each

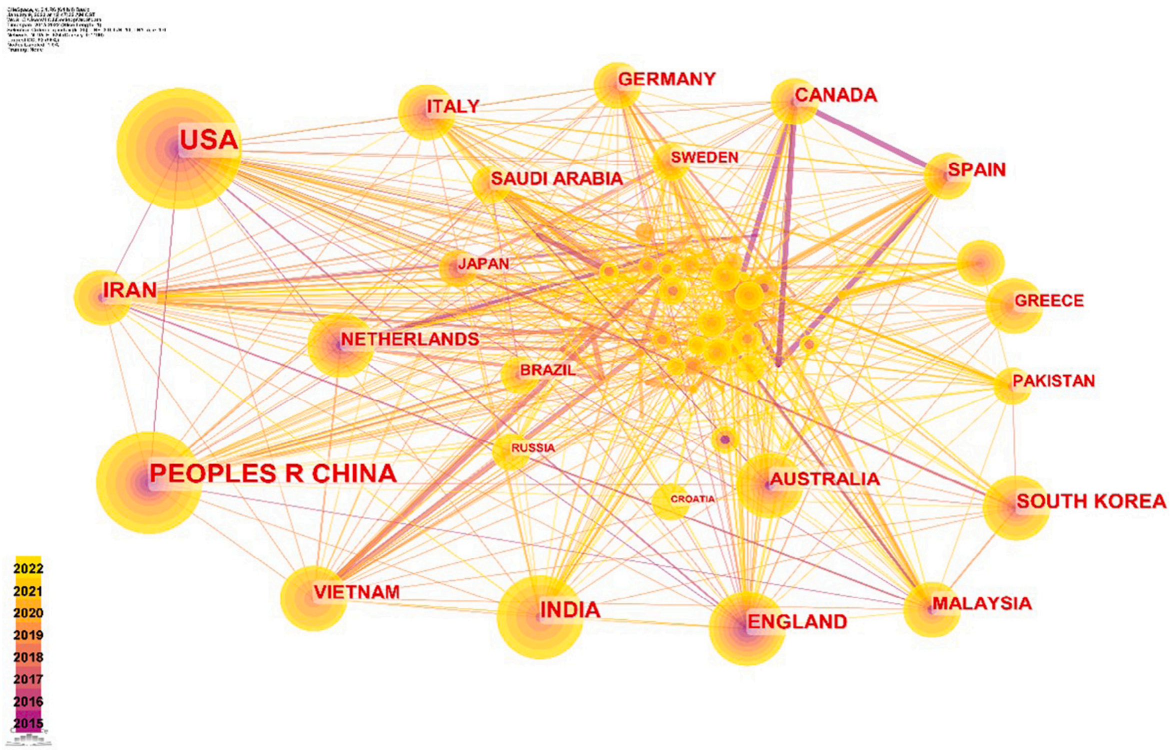

Fig. 3. Country collaboration network map.

Table 1
Top 10 most productive countries and top ten countries with the highest degree regarding water forecasting with machine learning/deep learning.

Ranking	Frequency	Countries	Ranking	Degree	Countries
1	215	USA	1	49	USA
2	191	China	2	45	Iran
3	67	Iran	3	42	China
4	62	India	4	36	Vietnam
5	48	England	5	33	India
6	42	South Korea	6	33	England
7	39	Vietnam	7	30	Malaysia
8	39	Australia	8	28	Spain
9	39	Canada	9	28	Germany
10	36	Spain	10	28	Netherlands

node in the knowledge map corresponds to the number of publications attributed to that country. Lines between nodes represent collaboration relationships, with a greater number of lines indicating more extensive cooperation.

Table 1 presents the top ten prolific countries categorized by both publication frequency and degree centrality. Degree centrality indicates the extent to which a country collaborates with others; hence, a higher degree centrality suggests more cooperative relationships. While the United States ranks first in both publication count and degree centrality within its collaboration network, it is important to note that countries with the highest publication numbers do not necessarily exhibit the most extensive cooperation with others. The United States leads with 215 papers, followed by China (191), Iran (67), India (62), England (48), South Korea (42), Vietnam (39), Australia (39), Canada (39), and Spain (36) in terms of publication frequency. In terms of degree centrality, the United States also holds the top position with a degree centrality of 49, followed by Iran (45), China (42), Vietnam (36), India (33), England (33), Malaysia (30), Spain (28), Germany (28), and the Netherlands (28).

3.2.2. Institution collaboration network analysis

Fig. 4 illustrates the academic collaborations among various institutions engaged in water forecasting research using machine learning and deep learning methodologies, employing similar operational strategies as depicted in the national cooperation network shown in Fig. 3.

The institution collaboration network comprises 284 nodes, 470 edges, and a network density of 0.0117. While the University of the Chinese Academy of Sciences ranks highest in terms of publication volume, Duy Tan University, situated in the coastal city of Da Nang, Vietnam, holds the topmost position in the entire collaboration network, boasting a degree centrality of 34. Notably, some organizations demonstrate minimal collaboration with other institutions, indicating a lack of cooperation among different organizations.

Table 2 presents the top 10 prolific institutions ranked by both publication frequency and degree centrality of publications. In terms of publication frequency, the University of Chinese Academy of Sciences leads the ranking with 20 documents, followed by Duy Tan University (19), University of Tehran (18), Texas A&M University (17), Ton Duc Thang University (16), Hohai University (15), Obuda University (11), National Technical University of Athens (11), Zhejiang University (10), and University of Gour Banga (9).

However, in the institution collaboration network, Duy Tan University holds the top position in terms of degree centrality, with a score of 34, followed by Ton Duc Thang University (26), University of Tehran (22), Lulea University of Technology (17), Obuda University (16), Hohai University (14), University of Twente (14), Transilvania University of Brasov (14), Tarbiat Modares University (14), and University of Transport Technology in Vietnam (13). Consequently, Duy Tan University, located in the coastal city of Da Nang, Vietnam, exhibits the broadest international cooperation, surpassing other universities in terms of international collaboration efforts.

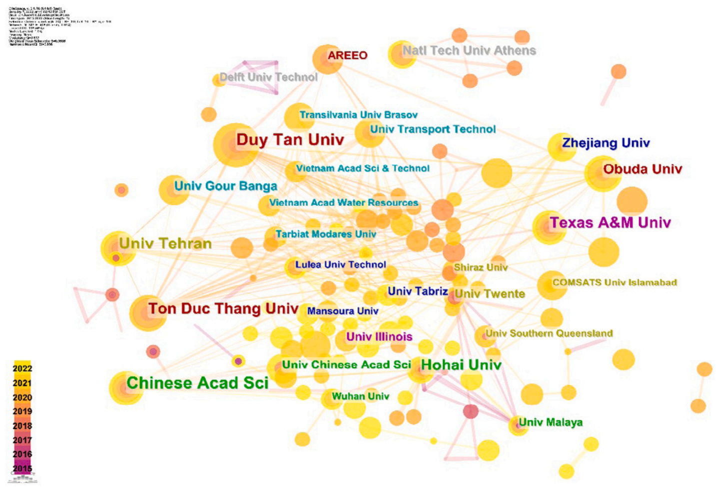

Fig. 4. Institution collaboration network map.

Table 2
Top 10 most productive institutions and top ten institutions with the highest degree regarding water forecasting with machine learning/deep learning.

Ranking	Frequency	Institutions	Ranking	Degree	Institutions
1	20	Chinese Acad Sci	1	34	Duy Tan Univ
2	19	Duy Tan Univ	2	26	Ton Duc Thang Univ
3	18	Univ Tehran	3	22	Univ Tehran
4	17	Texas A&M Univ	4	17	Lulea Univ Technol
5	16	Ton Duc Thang Univ	5	16	Obuda Univ
6	15	Hohai Univ	6	14	Hohai Univ
7	11	Obuda Univ	7	14	Univ Twente
8	11	Natl Tech Univ Athens	8	14	Transilvania Univ Brasov
9	10	Zhejiang Univ	9	14	Tarbiat Modares Univ
10	9	Univ Gour Banga	10	13	Univ Transport Technol

3.2.3. Author collaboration network analysis

Fig. 5 illustrates the academic collaboration among authors with similar professional backgrounds. Generally, authors' research endeavors are affiliated with their respective organizations, hence collaboration among authors aligns closely with the institution's collaboration network.

Within Fig. 5, four prominent research teams and several smaller teams are discernible. However, the author's collaboration network comprises 182 nodes, representing 182 authors, with 236 edges, resulting in a network density of 0.0143, indicating relatively low cohesion. Collaboration among authors in the water forecasting research domain utilizing machine learning and deep learning techniques appears to be less extensive. While some noticeable research teams are evident, numerous teams or authors lack substantial cooperation with their peers. The largest research team includes Hiep Van Le, Tran Van Phong, AI-ansari Nadhir, Huu Duy Nguyen, Binh Thai Pham, and Prakash Indra, among others. The second-largest research team is led by Rahmati Omid.

Table 3 presents the top 10 most prolific authors categorized by both publication frequency and degree centrality of publications. While Mosavi Amir ranks as the most prolific author with 9 publications, he does not exhibit the most extensive collaborative relationships. Conversely, despite Costache Romulus ranking third in terms of publication count, he emerges as the author with the most extensive cooperation with other researchers.

3.3. Co-citation analysis

In 1973, American intelligence scientist Henry Small introduced the concept of co-citation analysis in his article titled "Co-citation in the scientific literature: A new measure of the relationship between publications" (Small, 1973). This concept suggests that when two or more documents share common authors, they are considered to have an association relationship.

3.3.1. Journal co-cited analysis

We utilized the "cited journals" parameter in the CiteSpace analysis software to generate the co-citation network of journals related to water forecasting research employing machine learning and deep learning techniques, as depicted in Fig. 6. Nodes in the graph represent different journals co-cited in the documents, distinguished by colors and sizes. Larger nodes indicate greater attention received.

Table 4 presents the top 15 prolific co-cited journals, categorized by publication volume. The *Journal of Hydrology* received the highest number of citations with 355 papers, followed by the *Science of the Environment* (322), and *Water* (248). Additionally, two cited journals had a Journal Citation Indicator greater than 10, indicating the presence of core journals in the field of water forecasting research employing machine learning and deep learning techniques.

Drawing from the aforementioned, journal co-citation analysis was employed to explore the core journals in water forecasting research utilizing machine learning and deep learning techniques. This analysis provided insights into the academic caliber and quality of each journal, assessed through citation perspectives. Notably, publications in renowned peer-reviewed journals such as *Science* and *Nature* are esteemed for their original research contributions, alongside reviews and analyses of current research and science policy. The top 15 comprehensive scientific journals identified through this analysis serve as valuable reference materials for establishing the research authority within the domain of water forecasting employing machine learning and deep learning techniques.

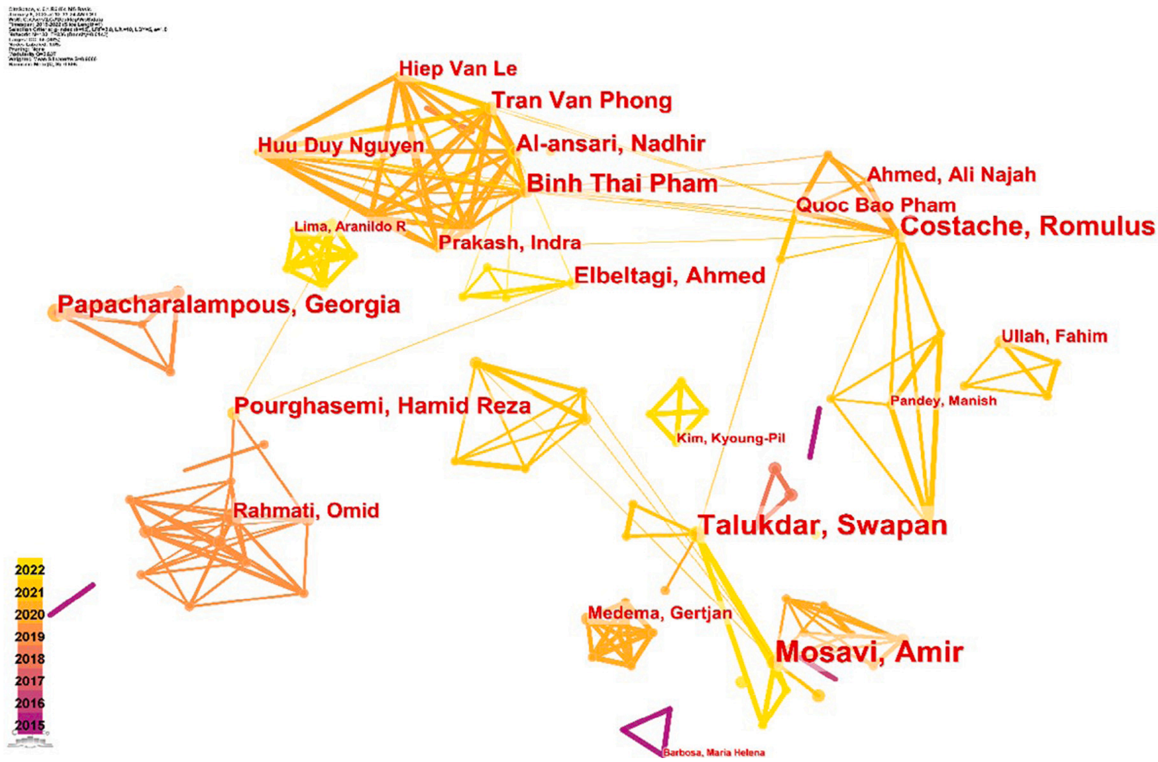

Fig. 5. Author collaboration network map.

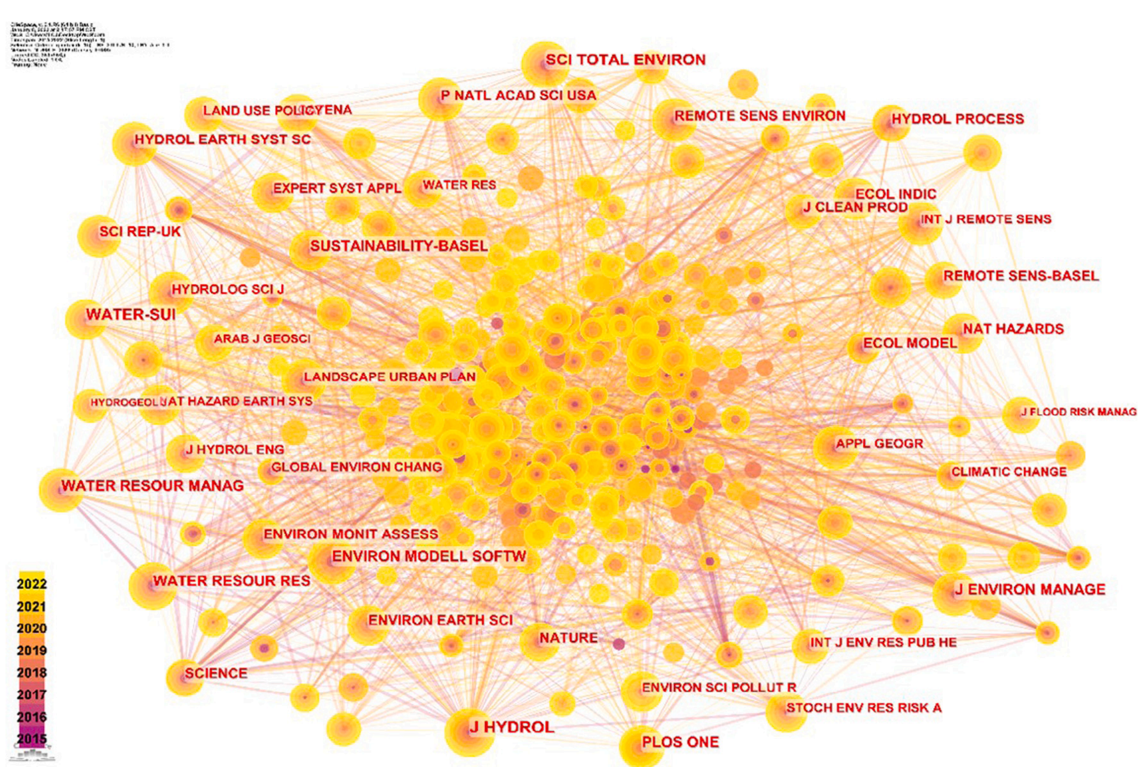

Fig. 6. Journal co-citation network map.

Table 3
Top 10 most productive authors and top ten authors with the highest degree regarding water forecasting with machine learning/deep learning.

Ranking	Frequency	Authors	Ranking	Degree	Authors
1	9	Mosavi, Amir	1	14	Costache, Romulus
2	8	Talukdar, Swapan	2	13	Binh Thai Pham
3	7	Costache, Romulus	3	11	Mosavi, Amir
4	6	Binh Thai Pham	4	11	AI-ANSARI, Nadhir
5	6	Papacharalampous, Georgia	5	10	Tran Van Phong
6	5	Binh Thai Pham	6	10	Prakash, Indra
7	5	Tran Van Phong	7	10	Huu Duy Nguyen
8	5	Pourghasemi, Hamid Reza	8	9	Rahmati, Omid
9	5	Elbeltagi, Ahmed	9	9	Hiep Van Le
10	4	Pourghasemi, Hamid Reza	10	9	Dieu Tien Bui

Table 4
Top 15 most cited journals of water forecasting with machine learning/deep learning documents.

Ranking	Counts	Journals	Journal citation indicator
1	355	*Journal of Hydrology*	1.59 (2021)
2	322	*Science of the Total Environment*	1.77 (2021)
3	248	*Water*	0.68 (2021)
4	225	*Water Resources Research*	1.36 (2021)
5	224	*Sustainability*	0.65 (2021)
6	207	*Water Resources Management*	1.03 (2021)
7	205	*Journal of Environmental Management*	1.38 (2021)
8	204	*Environmental Modelling & Software*	1.03 (2021)
9	194	*PloS One*	0.88 (2021)
10	172	*Science*	10.15 (2021)
11	150	*Nature*	10.86 (2021)
12	150	*Remote Sensing*	1.09 (2021)
13	143	*Environmental Earth Sciences*	0.67 (2021)
14	143	*Hydrology & Earth System Sciences*	1.48 (2021)
15	140	*Environmental Monitoring & Assessment*	0.53 (2021)

Table 5
Top 10 most cited authors of water forecasting with machine learning/deep learning documents.

Rankings	Citation counts	Authors
1	47	Bui, DT
2	44	Rahmati, O
3	40	Pourghasemi, HR
4	38	Arabameri, A
5	37	Lee, S
6	36	Pham, BT
7	36	Nash, J.E.
8	36	Youssef, AM
9	35	Tehrany, MS
10	34	Chen, W

3.3.2. Authors co-cited analysis

The co-cited author network serves as a valuable resource for identifying academic experts within specific domains. As depicted in Fig. 7, by selecting suitable pathfinder options and removing irrelevant nodes, we observe a network comprising 244 co-cited authors and 562 connections among them. In this visualization, the size of each node corresponds to the frequency of citations for the respective author, while closer distances between nodes indicate higher citation frequencies between the corresponding authors.

Table 5 showcases the top ten prolific co-citation authors, categorized by citation counts. These authors contribute to water forecasting research utilizing machine learning and deep learning techniques from various scientific perspectives. Bui, DT, Rahmati, O, Pourghasemi, HR, Arabameri, A, Youssef, AM, and Tehrany MS focus on flood hazard prediction, particularly in flood susceptibility prediction, employing diverse machine learning methods. Lee, S specializes in employing machine learning algorithms for landslide susceptibility mapping prediction. Meanwhile, Pham, BT, Nash, J.E., and Chen, W explore the application of hybrid machine learning models for river water salinity prediction and river flow forecasting.

3.4. Keywords network and clustering

We utilized the “keywords” parameter in the CiteSpace analysis software to generate the co-citation network of documents related to water forecasting employing machine learning and deep learning, as depicted in Fig. 8. Keywords offer a broad representation of the article’s subject matter, and their frequency, correlation degree, and emergence can unveil the research hotspots within the field.

Table 6 lists the top most frequently used keywords in the download documents and the keywords with the highest degree in the co-keywords network. According to Table 6, the keywords with the highest frequency of occurrence are prediction and machine learning, both appearing 128 times. The following are artificial neural network (107), climate change (101), impact (92), neural network (89), management (63), water (60), random forest (51), region (45), risk (44), system (42), area (41), classification (40), and support vector machine (38).

The important keywords obtained by ranking based on degree centrality are slightly different from the keywords obtained by frequency. It can be seen that the keywords “prediction” and “climate change” not only have the largest number of publications but also have high importance of research according to degree. However, according to the degree of centrality, keywords such as “land use change, gi, simulation, hazard, logistic regression, spatial prediction, cellular automata, and dynamics” are shown to be more important, but the frequency is relatively small, indicating that research involving these important keywords are not yet sufficient, and further attention and research are needed in the future.

Based on the keywords co-occurrence map, the LLR algorithm is used to obtain the keyword clustering network map as shown in Fig. 9. The modular value Q obtained from this clustering is 0.5677 ($Q > 0.3$ indicates significant network structure), and the average contour value S is 0.6707 ($S > 0.5$ indicates high cluster member consistency). Both Q and S values are within the normal range, so the clustering effect is significant and the results are reasonable. As can be seen from Fig. 9, under the LLR algorithm, a total of 10 clusters have been formed, and the label words of these clusters reflect the main research topics water forecasting with machine learning/deep learning documents since 2015: (0) landslide susceptibility, (1) drinking water, (2) time series, (3) support vector regression, (4) machine learning, (5) cellular automata, (6) water quality, (7) sustainable development, (8) risk factors, and (9) compressive strength.

3.5. Keywords bursts analysis

Keyword burst detection serves as an indicator of research focus shifts (Zhang et al., 2020; Liu et al., 2022b,a). Hence, we generated

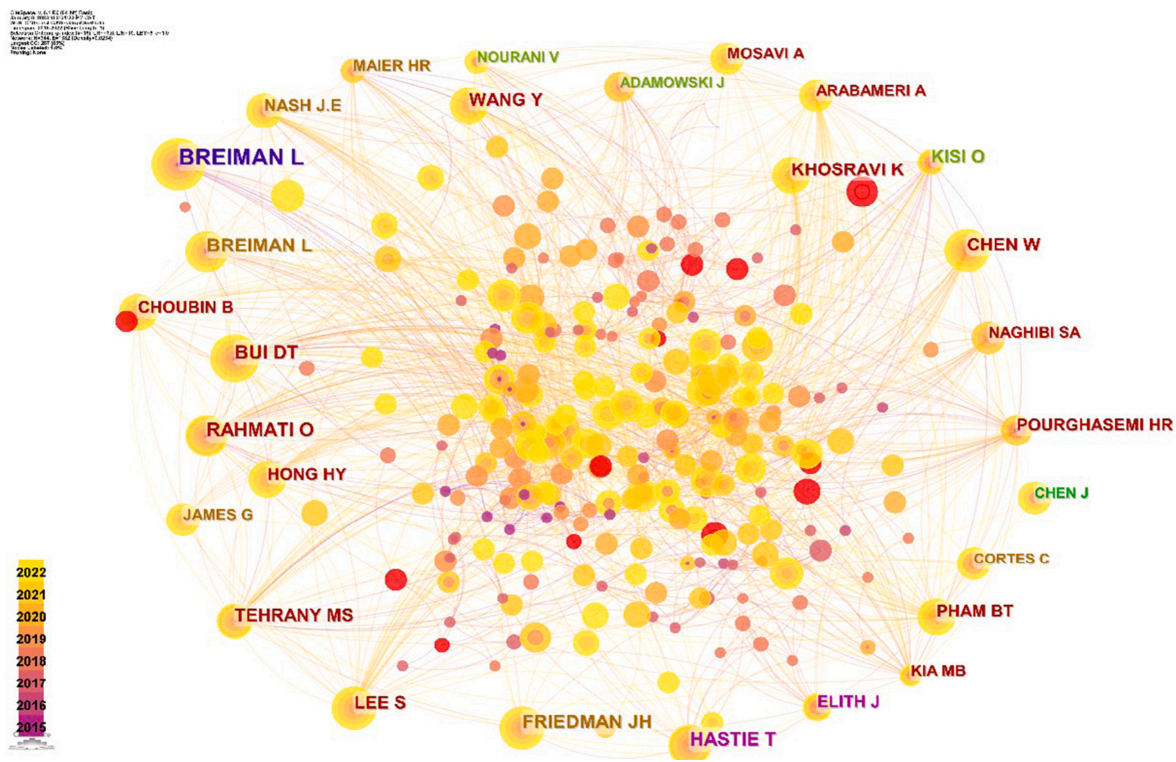

Fig. 7. Author co-citation network map.

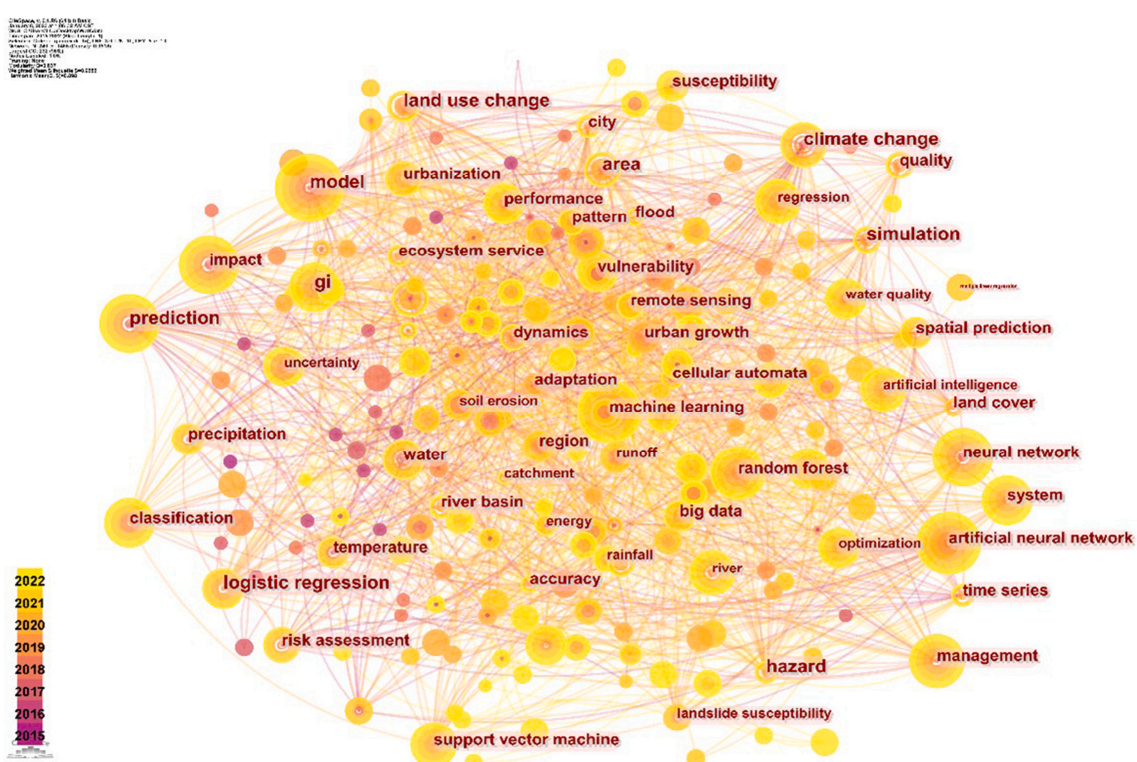

Fig. 8. Keywords co-citation network map.

Table 6
Top 15 keywords of water forecasting with machine learning/deep learning documents.

Rankings	Frequency	Keywords	Rankings	Degree	Keywords
1	128	prediction	1	43	area
2	128	machine learning	2	42	prediction
3	107	artificial neural network	3	41	climate change
4	101	climate change	4	37	land use change
5	92	impact	5	35	gi
6	89	neural network	6	35	simulation
7	63	management	7	35	hazard
8	60	water	8	34	logistic regression
9	51	random forest	9	33	machine learning
10	45	regression	10	33	impact
11	44	risk	11	32	artificial neural network
12	42	system	12	31	spatial prediction
13	41	area	13	31	cellular automata
14	40	classification	14	30	neural network
15	38	support vector machine	15	29	dynamics

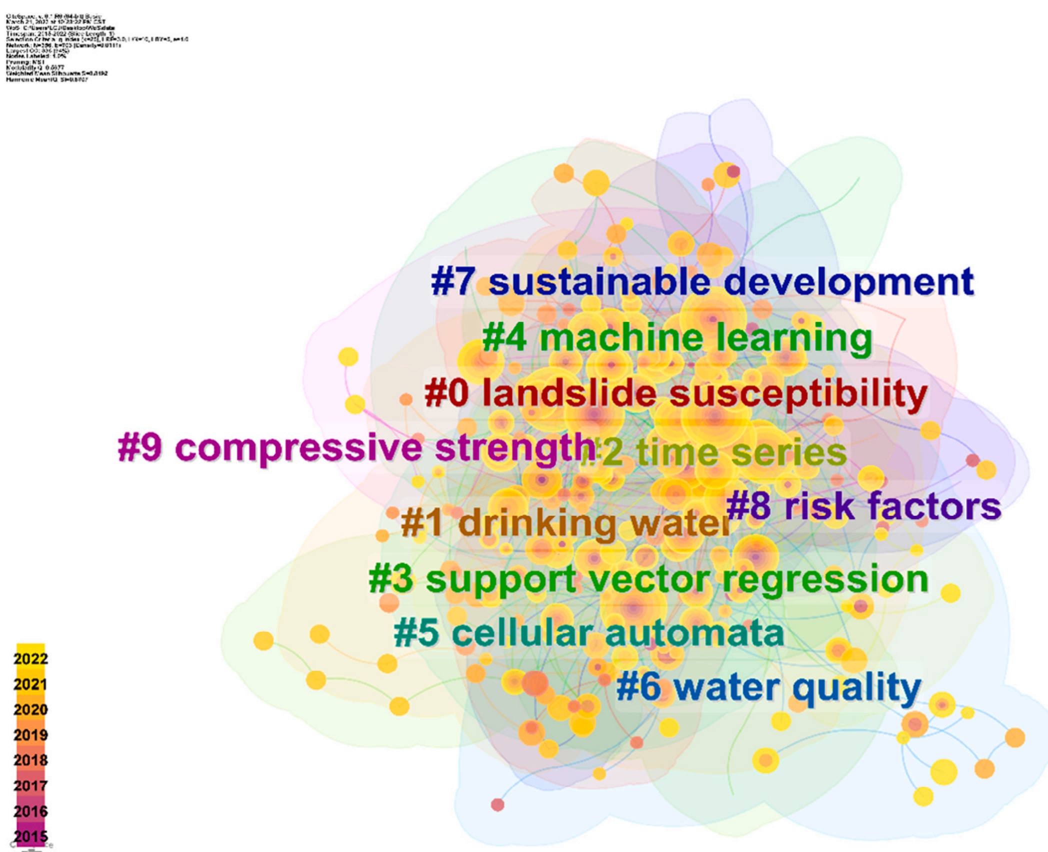

Fig. 9. Keywords clustering map.

a diagram illustrating the burst detection of keywords, identifying the top 16 keywords exhibiting the strongest citation bursts. In Fig. 10, these keywords are arranged chronologically, with a red elongated bar denoting the duration of each burst.

Fig. 10 displays the top 16 keywords with the most significant bursts observed from 2015 to 2022. The earliest burst is associated with the keyword "climate variability" (initiated in 2015), while the most recent is linked to "flash flood" (commencing in 2019). Notably, "temperature" exhibits the longest burst duration, spanning from 2015 to 2017. Moreover, keywords such as "climate variability", "activation", "health", "exposure", "difference water index", "abundance", "flood risk", "hybrid model", "soil erosion", "geographically weighted regression", "soil", "cover", "urban growth", "association", and "flash flood" each experience bursts lasting only one year. These terms signify sudden research hotspots.

4. Summary of water forecasting with machine learning/deep learning

By the temporal segmentation outlined in Fig. 2, we have synthesized foundational field insights to offer a clear overview of water forecasting research utilizing machine learning and deep learning methodologies. The summary of our findings is presented in Table 7.

As demonstrated in Table 7, we have divided the entire domain into two distinct phases based on the delineation provided in Fig. 2. We proceed to elaborate on the specific details of each phase, including the total article count, the most active countries and institutions, and the articles with the highest influence, as indicated by citation counts. Furthermore, we spotlight the top five authors who have made noteworthy contributions, assessed by the impact of their publications, i.e., articles with the highest citation rates. By integrating our research outcomes

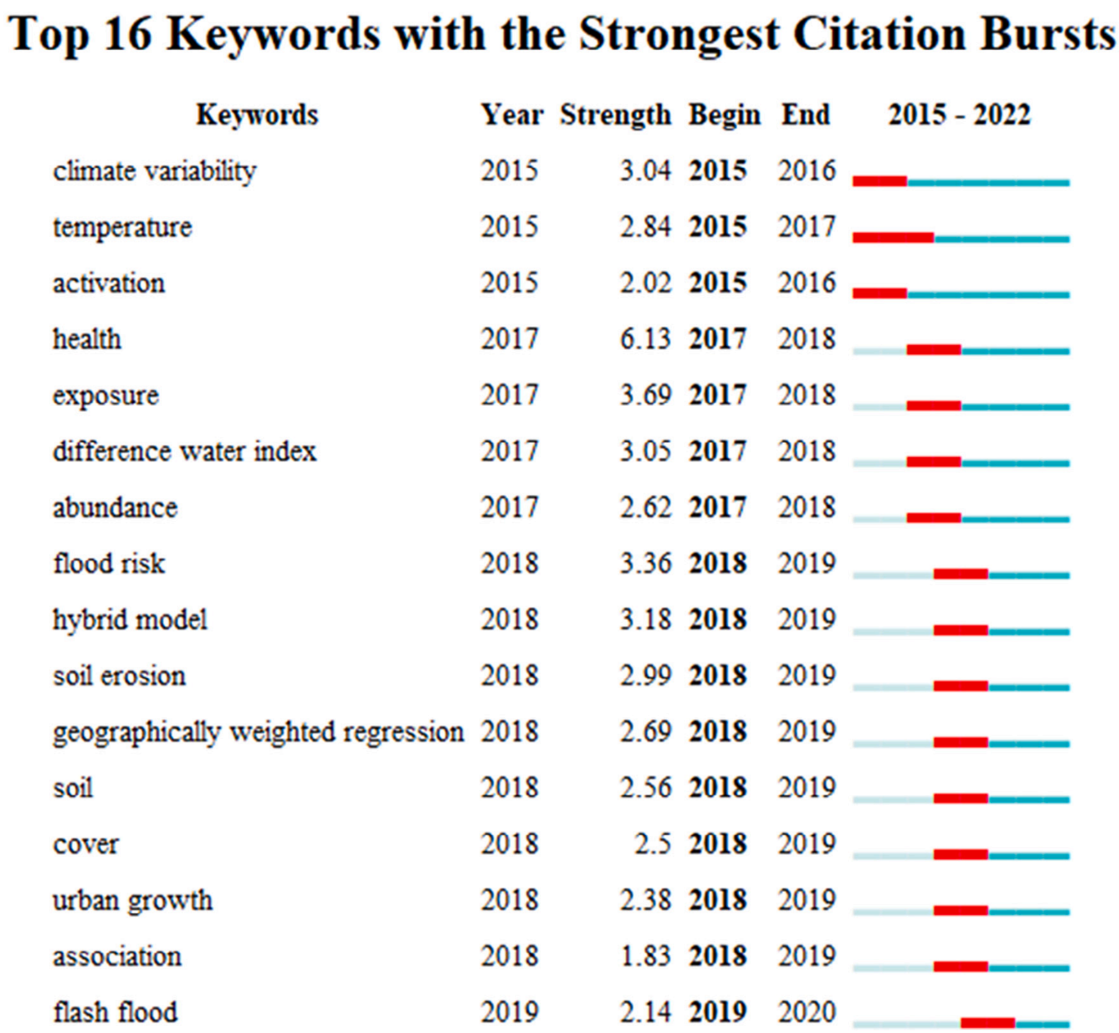

Fig. 10. Time sequence of keywords.

Table 7
Review stage in water forecasting with machine learning/deep learning research.

	The incipient progress stage (2015–2018)	The rapid development stage (2019–2022)
Number of articles	160	655
Top 3 of the most active countries	Iran China Canada	China USA Iran
Top 5 of the most active institutions	Islamic Azad University McGill University University of Malaya Chinese Academy of Sciences Harbin Engineering University	Egyptian Knowledge Bank Islamic Azad University Duy Tan University Ton Duc Thang University Chinese Academy of Sciences
Most influential literature	Shafizadeh-Moghadam et al. (2018)	Yaseen et al. (2019)
Top 5 of the most influential authors	Afrand M Hemmat Esfe M Zhao NB Adamowski J Apel H	Ahmed AN Bui DT Mosavi A El-shafie A Heddam S

Table 8
Research content of the studies about water forecasting with machine/deep learning.

Water forecasting	
Hydrological forecasting	1. River water level prediction 2. Seasonal streamflow forecasting in water reservoir operations 3. Water temperature forecasting 4. Water quality forecasting 5. Surface water flood forecasting 6. Short-term rainfall forecasting 7. River's water-level fluctuation forecasting
Water demand forecasting	1. Urban water demand forecasting 2. Residential water consumption forecasting 3. Near-term forecasting of freshwater quality

with the temporal dimension, we aim to provide readers with a comprehensive understanding of the knowledge framework and evolutionary trajectory within the water forecasting domain, particularly during this period of rapid growth.

5. Discussions and conclusions

After conducting a concise analysis of pertinent literature, the following insights were gleaned. Firstly, regarding temporal development, a notable period of advancement occurred between 2019 and 2022. During this timeframe, not only was there a marked increase in the number of published studies, but these studies also garnered significant citation rates. Secondly, concerning publishing sources, the United States, China, Iran, India, and England emerged as leading contributors to the field of water forecasting with machine learning and deep learning. These countries boast prolific authors and institutions, often engaged in collaborative academic endeavors. However, despite the prevalence of productive authors and institutions, the establishment of collaborative sub-networks within this field remains limited, highlighting the need for further development within the research community. Furthermore, co-citation analysis indicates that highly productive institutions or authors may not necessarily receive commensurate citation rates. For instance, although China ranks second in terms of publication volume, it lacks highly cited papers. This discrepancy could be attributed to limited opportunities for international collaboration in this domain, resulting in certain institutions and authors receiving less attention as compared to newcomers. Lastly, regarding research content, a majority of studies fall within the domains of hydrological forecasting and water demand forecasting (see Table 8).

As evidenced by the citation burst analysis outlined in Section 3.5, focal points in recent literature encompass climate variability, temperature fluctuations, various water indices, flood risk assessment, hybrid modeling techniques, soil erosion dynamics, and urban water demand dynamics. Moreover, there has been an emergence of novel high-frequency keywords in recent years. Coupled with a comprehensive interpretation of the keywords analysis, it becomes evident that the adoption of machine learning and deep learning methodologies has the potential to enhance both the depth and breadth of research in this field.

In light of the above, researchers are urged to remain abreast of the latest advancements in the machine learning and deep learning domains and explore their broader applications. Within literature dedicated to methodological discourse, it is noted that technological innovations facilitate enhanced problem comprehension, multifaceted problem-solving approaches, and the attainment of more effective problem-solving strategies. Technology, ultimately, is harnessed to serve practical applications; the fusion of neural networks with water time series data marks a significant milestone, with future directions pointing toward the increased application of diverse machine learning and deep learning techniques, as well as ensemble modeling approaches, in the realm of water forecasting.

Although this paper provides valuable insights, there are opportunities for further improvement in the future: (1) While CiteSpace enabled visual quantitative analysis in the ocean big data research domain, it is essential to recognize that this study represents preliminary work, considering the rapid advancements in both computer big data and oceanographic fields. (2) Solely relying on the Web of Science core database as the primary data source implies a potential limitation in data comprehensiveness. Future efforts should prioritize conducting more extensive and in-depth correlation analyses using updated datasets to enhance the richness of the findings.

CRediT authorship contribution statement

Chanjuan Liu: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jing Xu:** Writing – review & editing. **Xi'an Li:** Writing – review & editing. **Zhongyao Yu:** Writing – review & editing. **Jinran Wu:** Writing – review & editing, Visualization, Supervision, Investigation.

Declaration of competing interest

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

Acknowledgments

The funding for this study was provided by the Ministry of Education of Humanities and Social Science project in China (Project No. 22YJC630083), the 2022 Shanghai Chenguang Scholars Program (Project No. 22CGA82), the Belt and Road Special Foundation of The National Key Laboratory of Water Disaster Prevention (2021491811), and the National Social Science Fund of China (Project No. 23CGL077).

References

Bicheng, D., Adnan, N., Harji, M.B., Ravindran, L., 2023. Evolution and hotspots of peer instruction: A visualized analysis using citespace. Educ. Inf. Technol. 28, 2245–2262.
Brédy, J., Gallichand, J., Celicourt, P., Gumiere, S.J., 2020. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches. Agricult. Water. Manag. 233, 106090.
Chang, F.J., Chen, P.A., Lu, Y.R., Huang, E., Chang, K.Y., 2014. Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. J. Hydrol. 517, 836–846.
Chang, F.J., Guo, S., 2020. Advances in hydrologic forecasts and water resources management. Water 12, 1819.
Chen, C., 2006. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci. Technol. 57, 359–377.
Chen, C., 2017. Science mapping: A systematic review of the literature. J. Data Inform. Sci. 2, 1–40.
Chen, C., 2018. Visualizing and exploring scientific literature with citespace: An introduction. In: 3rd ACM SIGIR Conference on Human Information Interaction and Retrieval. CHIIR, Assoc Computing Machinery, New Brunswick, NJ.
Choi, K., Chan, S.N., Lee, J., 2023. The WATERMAN system for daily beach water quality forecasting: A ten-year retrospective. Environ. Fluid Mech. 23, 205–228.

Fang, S., Xu, L., Pei, H., Liu, Y., Liu, Z., Zhu, Y., Yan, J., Zhang, H., 2013. An integrated approach to snowmelt flood forecasting in water resource management. IEEE Trans. Ind. Inform. 10, 548–558.

Fullerton, Jr., T.M., Molina, Jr., A.L., 2010. Municipal water consumption forecast accuracy. Water Resour. Res. 46, W06515.1–W06515.8.

Latif, S.D., Ahmed, A.N., 2023. A review of deep learning and machine learning techniques for hydrological inflow forecasting. Environ. Dev. Sustain. 1–28.

Liao, H., Sun, W., 2010. Forecasting and evaluating water quality of Chao lake based on an improved decision tree method. Procedia Environ. Sci. 2, 970–979.

Liu, J., Wang, Y., Zhang, Q., Wei, J., Zhou, H., 2022a. Scientometric analysis of public health emergencies: 1994–2020. Int. J. Environ. Res. Public Health 19, 640.

Liu, X., Zhao, S., Tan, L., Tan, Y., Wang, Y., Ye, Z., Hou, C., Xu, Y., Liu, S., Wang, G., 2022b. Frontier and hot topics in electrochemiluminescence sensing technology based on CiteSpace bibliometric analysis. Biosens. Bioelectron. 201, 113932.

Paudyal, G.N., 2002. Forecasting and warning of water-related disasters in a complex hydraulic setting—the case of Bangladesh. Hydrol. Sci. J. 47, S5–S18.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., Shirzadi, A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. J. Environ. Manag. 217, 1–11.

Small, H., 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. J. Am. Soc. Inf. Sci. 24, 265–269.

Sun, L., Xu, X., Yang, Y., Liu, W., Jin, J., 2020. Knowledge mapping of supply chain risk research based on CiteSpace. Comput. Intell. 36, 1686–1703.

Wee, W.J., Zaini, N.B., Ahmed, A.N., El-Shafie, A., 2021. A review of models for water level forecasting based on machine learning. Earth Sci. Inform. 14, 1707–1728.

Wei, Y., Wang, Z., Wang, H., Yao, T., Li, Y., 2018. Promoting inclusive water governance and forecasting the structure of water consumption based on compositional data: A case study of Beijing. Sci. Total Environ. 634, 407–416.

Wu, J., Jia, D., Wei, Z., Xin, D., 2020. Development trends and frontiers of ocean big data research based on CiteSpace. Water 12, 1560.

Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.W., 2019. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387–408.

Zhang, C., Wang, S., Sun, S., Wei, Y., 2020. Knowledge mapping of tourism demand forecasting research. Tourism Manag. Perspect. 35, 100715.