Academic Education Publishing House
Occupation and Professional Education Vol. 1 No. 3, 2024

Advancing Frontiers in Data Security: A Comprehensive Bibliometric and Visual Exploration from 2013 to 2023

Jindong Hu[1], Na Li[2], Xing Liu[3,4,*], Jiamin Liang[5], Xizhou Yan[6]
[1]*Faculty of Law,Macau University of Science and Technology,Macau,*
[2]*School of Law,Henan University of Science and Technology ,Luoyang, Henan, China*
[3]*School of Law, Hunan University of Technology and Business, Changsha, China*
[4]*HunanResearch Base for the Construction of Clean Governance, Changsha, China*
[5]*School of Law,Hunan Normal University,Changsha,Hunan, China*
[6]*Nanning Municipal People's Procuratorate of Guangxi Zhuang Autonomous Region,Nanning, China*
*****Corresponding Author*

Abstract: This paper presents a bibliometric analysis of global research on data security from 2013 to 2023, highlighting the increasing challenges in cybersecurity and the evolution of new research paradigms. Through the use of bibliometric methods, this study evaluates the scope and themes of scholarly articles, pinpointing key trends and outlining potential avenues for future investigation. The research begins with a detailed review of major data breaches, which sets the stage for examining critical areas of academic interest within the field of data security. This includes the gathering of data and the selection of appropriate tools for bibliometric analysis, followed by an in-depth discussion of principal themes and prospective research directions. The results reveal a notable rise in publications related to data security, especially after 2018, indicating a heightened focus within the academic community. Significant contributions of this study include the identification of the most frequently cited works, an analysis of contributions by country and institution, and an examination of the distribution of articles across various journals. Additionally, the paper identifies leading researchers, key research hotspots, and emerging trends, highlighting the importance of a multidisciplinary approach that incorporates technological, legal, and managerial perspectives to tackle issues in data security. The conclusion of this research offers insights into the dynamic and growing field of data security studies, characterized by annual increases and an expansion into new areas such as personal data protection, blockchain technology, and artificial intelligence.

Key words: Data Security; Bibliometric Analysis; Cybersecurity Trends; Academic Research; Emerging Paradigms

1. Introduction

In recent years, the prominence of data security issues has grown significantly with the advancement of the digital age. There have been numerous large-scale data breaches, such as the Facebook data leak, security vulnerabilities at the American Q&A site Quora, and the exposure of Google+ account information, along with challenges encountered by various global commercial enterprises. These incidents reveal the inadequacies of traditional cybersecurity strategies, which primarily concentrate on defending against hacker attacks. They underscore the critical need for a shift in security system frameworks, advocating a transition from the conventional network-centric to a more robust data-centric approach to cybersecurity.

Data leaks, misuse, or damage can have severe consequences, including breaches of personaprivacy, economic losses, and legal disputes. As data increasingly becomes a crucial element of modern society, its importance to organizations, governments, businesses, and individuals becomes more evident. Data not only serves as a conduit for knowledge, information, and value but also shapes our operational methods and patterns of interaction.

In this context, data security, which ensures the integrity, confidentiality, and availability of data, has become a focal point of concern. Protective measures extend beyond technical

safeguards such as encryption and access control. They also include comprehensive strategies encompassing policy, legal frameworks, and education. Thus, understanding and maintaining data security is critical for any entity dependent on data.

With the accumulation of theoretical research and practical experience in data security, it isessential to systematically analyze and reflect on the achievements in this field. Historically, researchers have conducted detailed reviews from a professional perspective across various areas of data security. Notable studies include Meland et al. (2021) on cyber security indicators, Venkatraman and Venkatraman (2019) on big data risk mitigation, Farsi et al. (2020) on cloud computing security, and Alfawzan et al. (2022) on privacy and data security in mobile health applications. These works have significantly contributed to the knowledge within the data security domain, propelling its academic progress.

Despite the extensive focus on data security issues, the academic exploration often lacks depth. Numerous scholars have addressed the challenges, solutions, and future directions in data security from diverse perspectives. However, there remains a scarcity of thorough and quantitative analyses of these contributions. There is an evident need for more comprehensive reviews that systematically assess the current state and trends of research in data security.

This paper seeks to offer a thorough analysis of the current research landscape in data security using bibliometric methods, identifying core areas of focus and projecting future trends. Initially, we examine major data breaches and their implications on traditional cybersecurity strategies, emphasizing the critical shift from network-centric to data-centric security approaches.

We then detail the academic community's primary concerns and research trajectories within data security. By applying bibliometric techniques, this study methodically evaluates the existing literature, covering publication volumes, thematic developments, and emerging trends. Furthermore, the paper proposes forward-looking insights on prospective research directions in data security, providing a strategic guide for both scholars and practitioners in the field.

Previous research has extensively covered various aspects of data security, including cybersecurity indicators, risk mitigation in big data, cloud computing security, and issues related to data sharing. These studies have enriched the field with crucial knowledge and established a solid foundation for ongoing research. Building on this groundwork, this article utilizes bibliometric analysis and visualization tools to explore several research questions: What are the trends in publications, key terms, and authorship within data security literature? How is collaboration characterized across countries, regions, and institutions? What are the emerging hotspots in data security research? Additionally, this paper aims to forecast future research directions and themes in data security.

This article offers a detailed examination of the evolving trends and developments in the field of data security. It includes descriptive statistical analyses that identify leading authors, journals, institutions, and countries contributing to this field. Additionally, the paper employs collaborative network analysis and keyword clustering to further elucidate the research landscape.

Following this analysis, the paper discusses potential future directions for data security research and synthesizes key findings. It concludes by providing strategic recommendations for future studies, aiming to guide ongoing research efforts in data security.

The structure of this paper is organized as follows: Section 2 details the methods of data collection and the selection of tools for bibliometric visualization analysis. Section 3 presents the results of the bibliometric analysis, showcasing trends in publications and analyses based on authorship, geographic region, institution, subject category, and keywords. Section 4 provides an in-depth discussion of the key themes emerging from the bibliometric analysis and suggests directions for future research. Finally, Section 5 summarizes the findings of this study, offering a concise overview of the insights gained.

2. Materials and Methods

Bibliometric analysis plays a crucial role in providing researchers with a comprehensive view of the developmental trajectory and evolving research directions within a specific field. It also helps identify the leading authors, journals, institutions, and countries contributing

to that field. This analysis establishes a foundational framework that guides and informs future research efforts[1].This research report analyzes trends and potential future directions in global data security research over the past decade. To facilitate this analysis, scientific information maps were created using tools such as CiteSpace, VOSviewer, and RStudio. The findings from this study are poised to enhance fundamental research in data security and promote the development of innovative applications within the field.

2.1 Data source and search stratagies

The Web of Science Core Collection (WoSCC) is recognized as a high-quality, comprehensive database, highly regarded by scholars for bibliometric research. After evaluating various databases for authority, information volume, and compatibility, this study chose WoSCC as the literature source. The search criteria were set as TS = ("Data Security") covering the period from 2013 to 2023, with the language restricted to English. The starting year of 2013 marks the beginning of the big data era, reflecting the initial enthusiasm and concerns of this period. The literature search was conducted on October 20, 2023, to ensure minimal variation in publication and citation data. From the initial retrieval of 1501 articles, exclusions were made for conference abstracts (26 articles), editorial materials (18 articles), and other non-relevant materials (9 articles), as well as non-English articles (16 articles). After a meticulous screening process involving two independent researchers, 1432 papers were selected for final bibliometric analysis. Data including abstracts, publication years, keywords, authors, institutions, countries, and references were extracted for these articles. The screening process is illustrated in Figure 1.

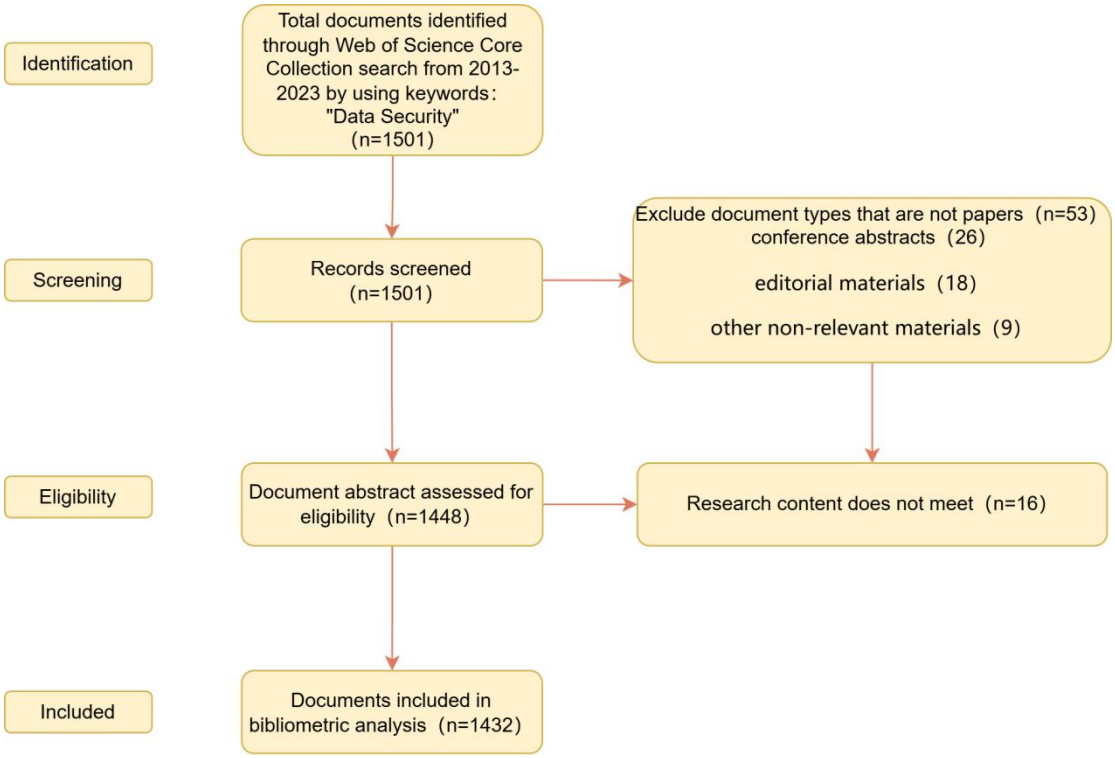

Figure 1. A PRISMA Flowchart for Bibliometric Analysis and Scoping of the Data Security

2.2 Bibliometric Analyzing Methodological Approach

The collected data were imported into Origin, VOSviewer, CiteSpace, and RStudio for comprehensive analysis. Origin was used to examine trends in annual publications by country and institution, assess citation counts for leading countries, identify top publications and journal citations, determine the most productive authors, and analyze the most frequently occurring keywords and references. This multi-tool approach allowed for a detailed exploration of the data, providing insights into various dimensions of the research landscape in data security.

CiteSpace, a citation visualization tool, was utilized to analyze fundamental article details

such as authors and keywords. This software provides a platform for visualizing co-occurrence networks, tracking the evolution of research, identifying trends in thematic shifts within specific fields, and forecasting future research directions. Using CiteSpace, we visualized collaborative networks among countries, institutions, and authors, as well as co-citation relationships between authors. Additionally, the tool facilitated burst analysis and timeline views of keywords, highlighting emerging research hotspots.

VOSviewer, created by Nees Jan van Eck and Ludo Waltman in 2009, was employed to examine bibliographic coupling indicators. This analysis tool helps map and analyze complex networks of scientific publications, offering insights into the interconnectedness of research within the field of data security[2].

3.Results

3.1 Overview of the Literature

3.1.1 Analysis of publications and citations trend

The development and growth of data security as a research field can be observed through the annual publication rates and citation counts. Between January 2013 and October 2023, the Web of Science Core Collection (WoSCC) cataloged 1432 papers on data security. Of these, 1374 (95.9%) were research articles and 58 (4.1%) were review papers. Figures 2A and 2B illustrate both the cumulative and annual quantities of papers published from 2012 to 2022. Over these 11 years, there has been a significant increase in the volume of publications. Notably, from 2018 onward, the annual output consistently exceeded 100 papers, with a peak of over 300 publications in 2022. This surge reflects a growing interest and marked expansion in the field, highlighting its rising popularity and increasing scholarly engagement.

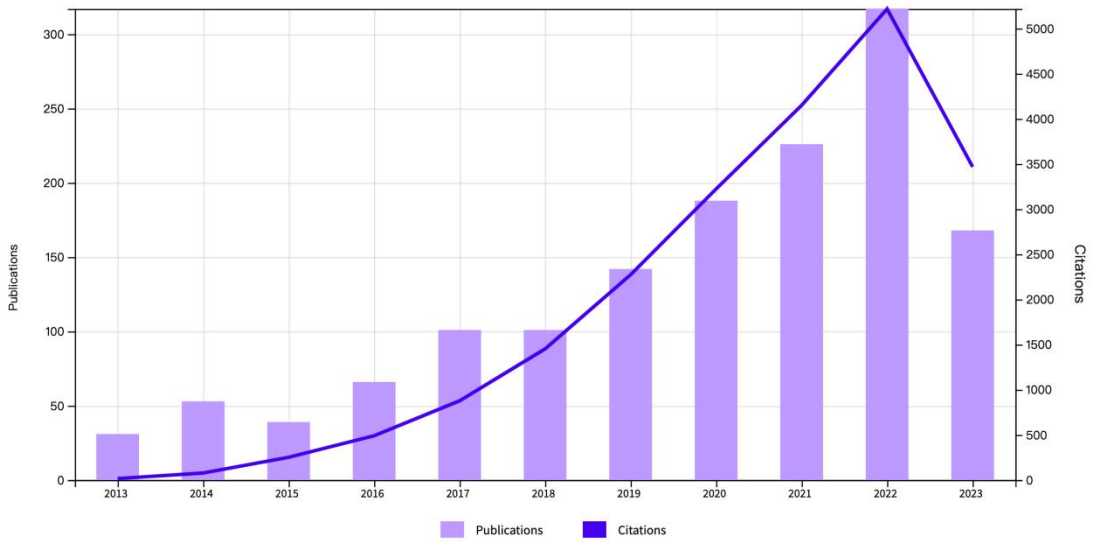

Figure 2. The Annual Number (A) and Cumulative Number(B) of Publications on Data Security from 2013 to 2023.

Next, we explore the impact of publications in the field by analyzing the most highly cited papers. As detailed in Table 1, the top ten most cited publications in data security research are presented. The leading paper, a review on the use of data mining and machine learning for detecting network security intrusions, has garnered an impressive 1258 citations, averaging about 179.71 citations per year. This highlights both the intense research activity in the field and the significant influence of this work. Conversely, the publication ranked tenth, while the least cited among the top ten, focuses on the integration of big data and machine learning in smart grid security, illustrating the wide-ranging relevance of data security research.

The publications ranked second to ninth demonstrate that data security reaches beyond cyberspace, affecting critical sectors such as smart cities, healthcare, food, and water safety. Notably, in the context of the ongoing global health crisis, the sixth-ranked article, which analyzes data from the International Health Regulations Annual Report, emphasizes the critical importance and urgency of data security research in public health. This breadth of application highlights the pervasive relevance

of data security across various essential domains.

The fifth-ranked publication, though not the highest in total citations, is focused on smart cities—a rapidly evolving field with significant implications for future urban planning and management. This research underscores the critical intersection of data security with urban development. Additionally, other papers listed in the table cover a range of pertinent topics within data security, including the application of blockchain technology, big data privacy, and data protection in edge computing. These areas are currently among the most dynamic and crucial hotspots in data security research.

Table 1. List of Top 10 Cited Publications Research on Data Security

Rank	Title	Year	Source	Citations Numbers	Average Per Year
1	A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection	2016	IEEE Communications Surveys and Tutorials	1258	179.71
2	Smart responsive phosphorescent materials for data recording and security protection	2014	Nature Communications	682	75.78
3	Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?	2018	IEEE Cloud Computing	387	77.40
4	information Security in Big Data: Privacy and Data Mining	2014	IEEE Access	331	36.78
5	Smart Cities: A Survey on Data Management, Security, and Enabling Technologies	2017	IEEE Communications Surveys and Tutorials	274	45.67
6	Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries	2020	Lancet	272	90.67
7	State estimation under false data injection attacks: Security analysis and system protection	2018	Automatica	253	50.60
8	Data Descriptor: A land data assimilation system for sub-Saharan Africa food and water security applications	2017	Scientific Data	232	38.67
9	Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues	2018	IEEE Access	214	42.80
10	Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review	2019	IEEE Access	165	41.25

3.1.2 Country and institution analysis

In recent years, institutions across various countries have notably advanced the field of data security, as reflected in their research outputs. Table 2 illustrates that among the countries contributing to this research, only the top ten have each published over 22 papers, with the majority originating from Asia, North America, Europe, and Oceania. China is at the forefront with a substantial 527 papers, followed by India with 198, the USA with 158, and South Korea with 50. Notably, while the UK has also published 50 papers, its average citation count per paper is 25.12, significantly higher than South Korea's 11.66.

The USA, however, leads in terms of average citation count at 27.25, suggesting that the quality and impact of its research may exceed that of other nations. Despite China's dominance in quantity, its average citation count stands at 14.22, which is lower than that of the USA, indicating the sustained academic influence and leadership of American research in the field.

Figure 3A depicts the international collaboration network within the field of data security, highlighting China as the central node with the most extensive collaborative links. This illustrates China's significant influence and active participation in global data security research. The USA also holds a crucial position in this network, exhibiting a wide array of international partnerships. Similarly, European countries like France, Germany, and the UK

show strong collaborative networks, indicating their active involvement and substantial contributions to the research in data security.

In the temporal dimension, multicolored rings around the nodes illustrate the varying intensity of collaborations over different years, demonstrating the dynamic evolution of data security research. The thickness of the lines between countries indicates the strength and frequency of their collaborations, underscoring the importance of specific bilateral relationships. Geographic proximity and cultural connections also appear to play roles in shaping collaboration patterns, as seen in the dense network of cooperation among European countries. Emerging research nations such as India and Saudi Arabia, though represented by smaller nodes, are becoming increasingly prominent in data security research collaborations, evidenced by the diversity of colors and thicker connecting lines. This diagram effectively reveals the complex landscape of international scientific collaboration in the field of data security.

This research spans 655 institutions worldwide, with the top 10 detailed in Table 3. Notably, the top five include the Chinese Academy of Sciences with 37 publications (2.58% of total), Egyptian Knowledge Bank (EKB) with 19 publications (1.33%), and the Indian Institutes of Technology System (IIT System), King Saud University, and Tsinghua University each contributing 17 publications (1.19%). The prominence of Chinese institutions, which occupy several top spots, highlights China's intense focus and growing interest in data security. Significantly, institutions from the People's Republic of China hold the top 15 positions, with the University of California System in the United States ranked 16th, indicating the global but concentrated nature of research efforts in this field.

Figure 3B illustrates the collaboration network among academic institutions in the field of data security. In this diagram, nodes symbolize various academic entities, with the size of each node correlating to the institution's influence or its volume of collaborations within the network. For example, the Chinese Academy of Sciences is depicted as a significantly larger node, reflecting its extensive involvement in collaborative research efforts in data security. The colors of the nodes range from deep red, representing collaborations that began in 2013, to deep blue, denoting more recent activities up to 2023. This color gradient shows the temporal dynamics of institutional collaborations. Notably, institutions like the Vellore Institute of Technology (VIT) and Anna University are marked by deep blue, indicating heightened activity in recent years.

In the diagram, lines represent collaborative relationships between academic institutions, with the thickness of these lines indicating the intensity of collaboration or the number of jointly published papers. For instance, a thicker line between the Indian Institute of Technology System (IIT System) and the Chinese Academy of Sciences indicates substantial collaborative engagement between these entities. Geographically, institutions in Asia, particularly from China and India, are highly active within the collaborative network for data security research. There is also significant collaboration with institutions from other regions, including the University of California System and University of Texas System in the United States, as well as King Saud University in Saudi Arabia, demonstrating a global interconnectivity in data security academic endeavors.

This chart provides a clear depiction of the complex collaborative relationships among global higher education institutions within the field of data security research. It highlights the central roles that certain academic institutions play within this network. This visual representation of collaborative relationships offers valuable insights into the trends and intensities of global research cooperation, underscoring the interconnected nature of scholarly efforts in data security.

Table 2. List of Top 10 Countries Publishing Research on Data Security

Ranking	Country	Number of publications	Citation	Average citation
1st	china	527	7496	14.22
2nd	india	198	1735	8.76
3rd	usa	158	4306	27.25
4th	korea	50	583	11.66
5th	united kingdom	50	1256	25.12

6th	turkey	23	125	5.43
7th	australia	33	781	23.67
8th	spain	21	257	12.24
9th	saudi arabia	25	281	11.24
10th	canada	22	431	19.59

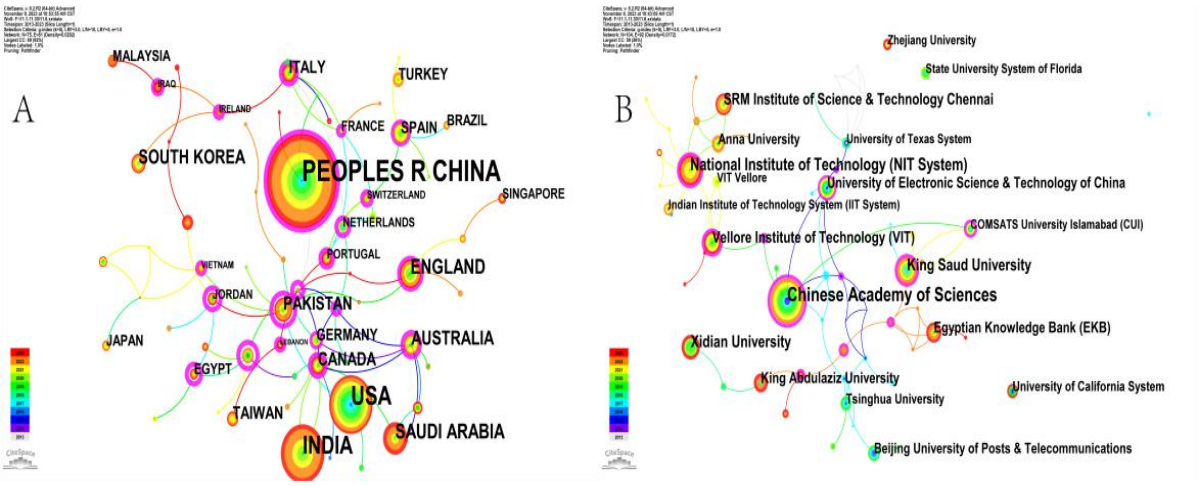

Figure 3. Analysis of Countries (A) and Institutions (B) Engaged in Ferroptosis in Data Security

Table 3 List of Top 10 Productive Institutions Related To Data Security

Randking	Institution	Country	Count	%
1st	CHINESE ACADEMY OF SCIENCES	China	37	2.58%
2nd	EGYPTIAN KNOWLEDGE BANK (EKB)	Egypt	19	2.58%
3rd	INDIAN INSTITUTE OF TECHNOLOGY SYSTEM (IIT SYSTEM)	India	17	1.33%
4th	KING SAUD UNIVERSITY	Saudi Arabia	17	1.19%
5th	TSINGHUA UNIVERSITY	China	17	1.19%
6th	XIDIAN UNIVERSITY	China	17	1.19%
7th	UNIVERSITY OF CALIFORNIA SYSTEM	United States	15	1.19%
8th	NATIONAL INSTITUTE OF TECHNOLOGY (NIT SYSTEM)	United States	15	1.05%
9th	UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA	China	15	1.05%
10th	VELLORE INSTITUTE OF TECHNOLOGY (VIT)	India	14	1.05%

3.1.3 Distribution of journals

A total of 633 journals have featured papers on data security, with 25 of these journals each contributing at least three articles. The top ten journals collectively represent approximately 23.18% of the publications in this field. This data underscores the concentration of significant research output within a select group of journals, highlighting their influential role in the dissemination of data security knowledge.

In the field of data security research, as detailed in Table 4, the journals leading in publication volume include IEEE Access, Security and Communication Networks, and Sensors. This underscores their prominent status and significant scholarly contributions to the field. Notably, the United States and Switzerland each host two publishers among the top ten, highlighting these countries' active roles in data security research. Particularly, IEEE Access, as the leading journal, stands out with an average citation count of 18.58, indicating the high impact of its published papers.

The journal with the highest average citation count (AC) is the IEEE Internet of Things Journal, with an AC of 35.24, indicating the high quality and significant impact of its published papers. This is closely followed by Future Generation Computer Systems—The International Journal of eScience, boasting an AC of 34.72. In terms of Impact Factor (IF), the IEEE Internet of Things Journal leads with an IF of 10.6, followed by Computers &

Security, which has an IF of 5.6. These metrics reflect the prestigious standing of these journals within the scientific community.

The majority of these journals are classified as Q1 and Q2, indicating their prominence in the field of data security. Notably, Multimedia Tools and Applications, though ranked fifth in terms of publication volume, has an Impact Factor (IF) of 4.27, categorizing it within the Q1 level. This demonstrates that despite a lower quantity of publications, the journal's academic quality and influence are significant. The research published in these journals is not only abundant but also of high quality, contributing significantly to the scholarly landscape of data security.

Table 4. List of Top 10 Productive Journals Related to Data Security

Randking	Source	JCR	Country	ISSN	IF(2022)	NP	NC	AC
1st	IEEE ACCESS	Q2	USA	2169-3536	3.9	67	1245	18.58
2nd	SECURITY AND COMMUNICATION NETWORKS	Q3	UK	1939-0114	1.968	39	171	4.38
3rd	SENSORS	Q2	Switzerland	1424-8220	3.9	35	340	9.71
4th	WIRELESS PERSONAL COMMUNICATIONS	Q2	Netherlands	0929-6212	3.16	34	145	4.26
5th	MULTIMEDIA TOOLS AND APPLICATIONS	Q1	USA	1380-7501	4.27	25	283	11.32
6th	APPLIED SCIENCES-BASEL	Q2/ Q3	Switzerland	2076-3417	2.7	19	103	5.42
7th	FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE	Q1	Netherlands	0167-739 X	7.307	18	625	34.72
8th	SUSTAINABILITY	Q2/ Q3	Switzerland	2071-1050	3.9	18	110	6.11
9th	IEEE INTERNET OF THINGS JOURNAL	Q1	UK	2327-4662	10.6	17	599	35.24
10th	COMPUTERS & SECURITY	Q3	England	0167-4048	5.6	16	324	20.25

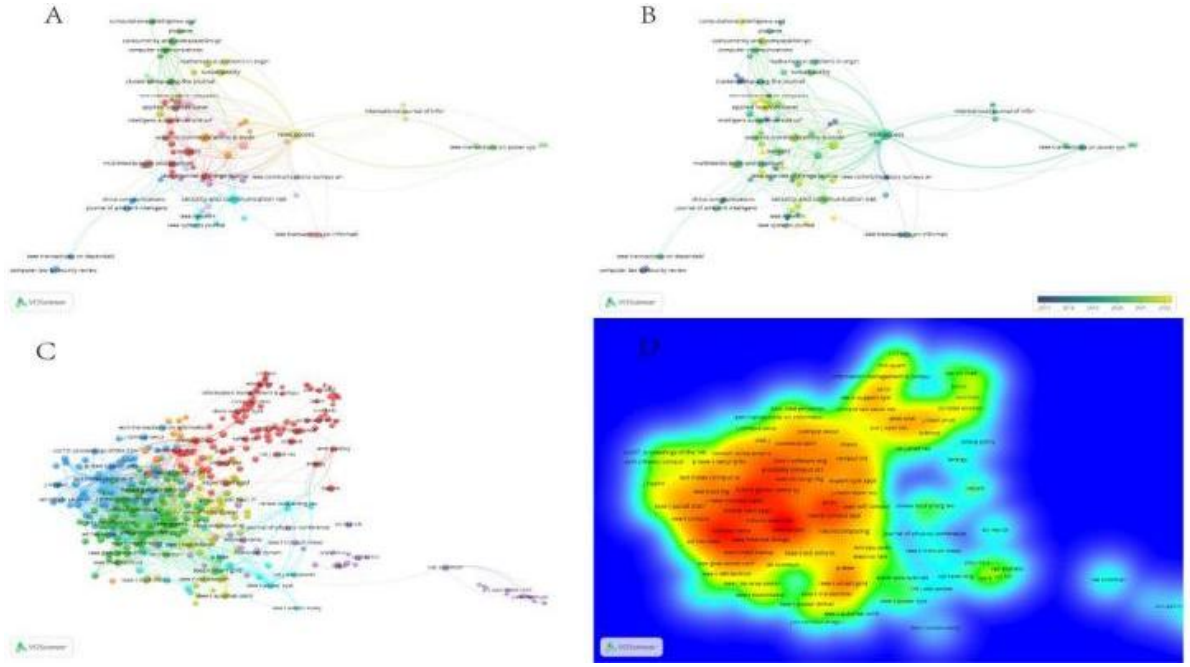

Figure 4. Visual analysis of Journals. (A) Visual Analysis of Journals Published at Least Three Publications in Each Journal. (B) The Evolution of Several Journal Publications. (C) Visualization of Co-Cited Journals that Have Been Cited at Least 20 Times. (D) Density Map of Co-Cited Journals.

3.1.4 Characterization of the authors of the publications(Distribution of authors)

This section presents a systematic analysis of authors, regions/countries, and institutions within the field of data security, aiming to identify the most prolific and influential contributors. The primary metric used to measure an author's contribution is their number of publications. Table 5 lists the top ten authors, detailing their total number of publications, number of papers (NP), total citations (NC), and the h-index (AC).

For example, Jin Wang from Hunan University of Science & Technology is ranked second with 11 papers and an impressive citation count of 505, which is reflected in a significant h-index of 49. This highlights the high relevance and recognition of his work within the field. Amit Kumar Singh from the National Institute of Technology Patna is the highest-ranked author from India, with 9 papers and 254 citations, underscoring his substantial contribution and the academic impact of his research in data security.

These findings demonstrate the dynamic and collaborative nature of research in data security, emphasizing significant contributions from various academic institutions and countries, particularly China and India, which dominate the list with their extensive academic output and impact.

Table 5. List of Top 10 Productive Authors Related to Data Security

Randking	Author	Affiliations	Country	H-index	NP	NC	AC
1st	Liu, Liang	Sichuan University	China	10	11	143	13.00
2nd	Wang, Jin	Changsha University of Science & Technology	China	49	11	505	45.91
3rd	Li, Jun	Hainan University	China	7	10	442	44.20
4th	Amit Kumar Singh	National Institute of Technology Patna	India	36	9	254	28.22
5th	Xu, Yan	Central South University	China	17	9	201	22.33
6th	Wang, Liang	Princeton University	USA	4	8	156	19.50
7th	Zheng, Yan	Hunan Institute of Science & Technology	China	7	8	301	37.63
8th	Yu, Yang	Guizhou Medical University	China	6	8	135	16.88
9th	Chen, Yuanyuan	Guangxi Normal University	China	3	7	47	6.71
10th	Kim, Jinwoo	Kwangwoon University	Korea	2	7	13	1.86

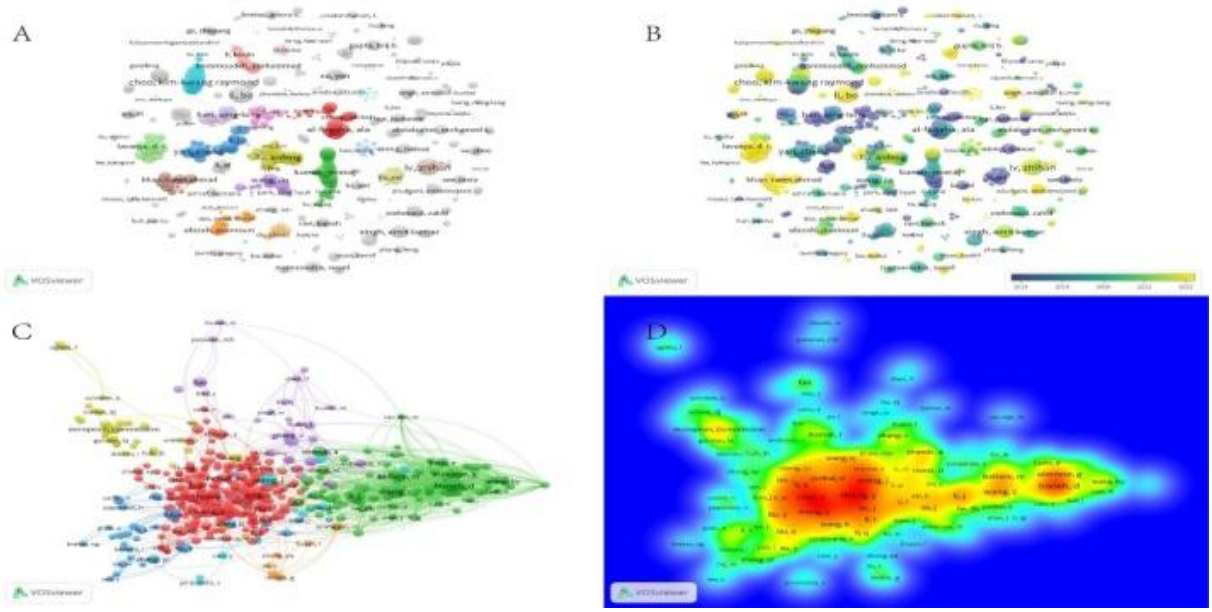

Figure 5. Visual Analysis by the Authors. (A) Visual Analysis of the Authors, with a Minimum of Two Publications Per Author. (B) The Evolution of Output Results from Different Scholars. (C) Visualization of Co-Cited Authors that Have Been Cited at Least 10 Times. (D) Density Map of Cocited Authors.

3.2 Research Hotspots

Based on the data collected, keywords were organized in descending order of frequency and analyzed using co-occurrence and clustering diagrams, as shown in Figures 6A and 6B. In these diagrams, each node represents a keyword. The size of the node corresponds to the frequency of the keyword's occurrence, with larger nodes indicating higher frequencies. For more detailed information, refer to Table 1. Additionally, in Figure 6, a pink halo around each node highlights the centrality of the keyword, where the thickness of the halo is directly proportional to the keyword's centrality in the field. This visual representation aids in identifying the most central and discussed topics within data security research.

Table 6 presents the frequency and centrality statistics of keywords within the field of data security research. Keywords such as "big data,""data security," and "cloud computing" occur frequently, indicating strong research interest and their foundational role in this domain. These terms have garnered substantial scholarly attention, highlighting their relevance in data security discussions.

In terms of centrality, "privacy" is prominent with a centrality score of 0.73, signifying its central role in research discussions and its strong interconnectivity with other key terms in data security. This reflects the integral relationship between privacy issues and other aspects of data security. "Efficiency" and "big data" also show significant centrality, emphasizing the critical need for efficient data processing and analysis in this field.

Although "data security" is highly frequent, its lower centrality score indicates that while it is a core theme, its connections with other keywords may not be as extensive as those of "privacy." Additionally, the rising importance of "edge computing,""data aggregation," and "cloud computing" in recent literature signals an increasing focus on these technologies to enhance data security. This shift underscores the evolving nature of research priorities and technological advancements in the field.

The frequency and centrality statistics of keywords in the field of data security research, as detailed in Table 6, highlight the principal areas of focus and the interconnectivity among these topics. This information reveals that privacy protection, efficiency enhancement, and the integration of emerging technologies are currently the key and most discussed topics in data security research. These insights reflect the evolving priorities and ongoing advancements within the field.

Figure 6B displays keyword clusters generated using the Log-Likelihood Ratio (LLR) algorithm, represented by irregular quadrilaterals. The clustering modularity index (Q) and the clustering silhouette index (S) are 0.7487 and 0.9094, respectively. A Q value greater than 0.3 indicates a significant and well-defined clustering structure, while an S value above 0.5 is typically seen as satisfactory. An S value exceeding 0.7 suggests highly efficient and convincing clustering results. Together, these indices validate the scientific robustness and effectiveness of the clustering in the diagram.

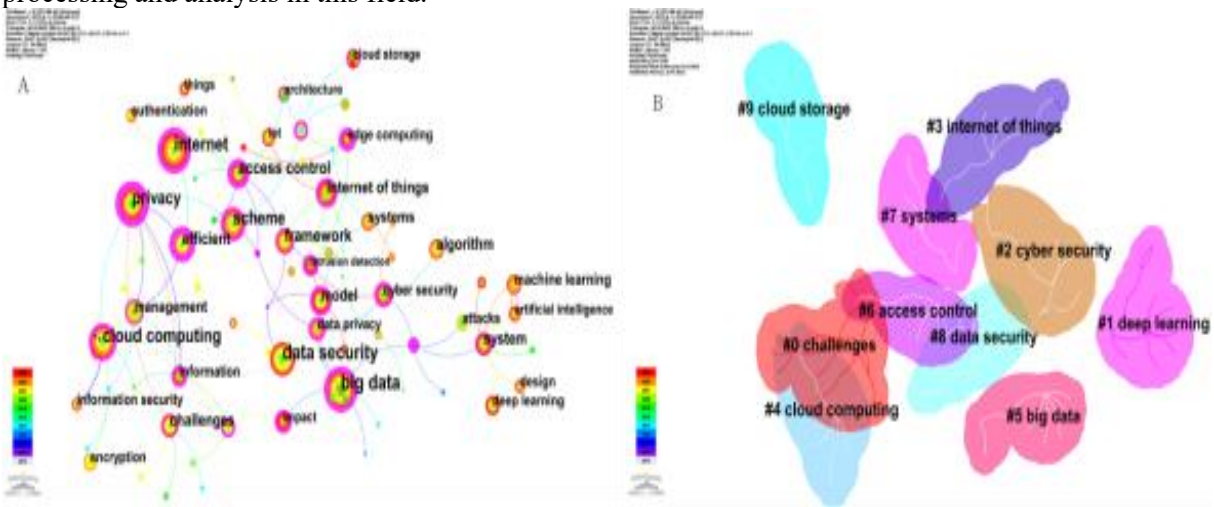

Figure 6. Keyword Co-Occurrence and Clustering Map. (a)Keyword Co-Occurrence Map. (b)Keyword Clustering Map.

Table 6. Keyword Frequency and Centrality Statistics Related To Data Security

Keyword Frequency Statistics			Keyword Centrality Statistics		
number	Keywords	Frequency	number	Keywords	Frequency
1	big data	122	1	privacy	0.73
2	data security	119	2	efficient	0.64
3	cloud computing	113	3	big data	0.51
4	internet	95	4	access control	0.5
5	privacy	85	5	internet	0.44
6	scheme	75	6	edge computing	0.43
7	access control	56	7	data aggregation	0.41
8	framework	55	8	cloud computing	0.33
9	system	53	9	cyber security	0.3
10	challenges	50	10	information	0.27

4. Discussion

4.1 Hot Literature

Building on previous scholarly work and the analysis of the knowledge maps described earlier, this study categorizes the themes of data security research into three main perspectives from a macro viewpoint: legal, management, and technological. This classification serves to clarify and systematically organize the knowledge system and developmental trajectory of data security research, providing a structured framework that enhances understanding of the field's evolution and current focus areas.

(1) Legal Perspective. In addressing the legal governance of data security, this study delves into various legal perspectives necessitated by the advent of big data technology. This new era of data management brings with it a range of legal challenges, particularly related to data ownership, application, and management. To address these challenges, the research explores key legal perspectives including personality rights, property rights, criminal law, and sovereignty. Each of these areas is crucial for understanding and navigating the complex legal landscape that surrounds modern data security issues.

Under the context of data security, personality rights primarily focus on personal privacy and data protection [. Individuals should have the authority to control how their personal data is utilized, which includes not only basic identity information but also preferences, communication records, and more. In alignment with personality rights, any entity handling personal data must maintain transparency, fairness, and legality throughout the processing procedure. Violations of these principles, such as unauthorized data marketing activities, could infringe upon an individual's personality rights.

Within the framework of property law, data is recognized as a valuable asset owned by individuals, corporations, or other organizations. This ownership grants the right to manage, utilize, and transfer the data. Unlike traditional property, data can be replicated and transferred without diminishing the value of the original asset. However, if the data is used or leaked unlawfully, the owner may suffer economic and reputational losses. This highlights the unique challenges of safeguarding data within the scope of property rights.

In the realm of data security, criminal law targets actions that breach data protection regulations, including data theft, illegal intrusion, and data tampering. The objective of these legal sanctions is to deter such offenses, thereby enhancing the overall security of data. Furthermore, criminal law also addresses associated criminal activities, such as data fabrication and fraud, ensuring comprehensive legal coverage to prevent misuse of data.

The role of sovereignty in data governance is closely linked to the reality of transnational data flows. In the globalized digital economy, data frequently crosses national borders, necessitating countries to develop laws and policies to ensure data flows are safe, fair, and lawful. This regulatory landscape also encompasses issues of data ownership and control in a transnational context. Multinational companies must navigate and adapt to varied regulations concerning data control, storage, and usage across different countries, which adds complexity to their operations.

In conclusion, as we navigate the challenges of the big data era, it is crucial to approach data security from multiple legal perspectives to ensure data is handled securely and lawfully. This comprehensive legal approach will help address the complexities introduced by the vast amounts of data now integral to our digital economy.

(2) Technical Perspective. From a technical standpoint, data security encompasses three core areas: data storage security, data transmission security, and data access control. Each of these components is crucial for safeguarding information throughout its lifecycle, from storage to transmission and access.

Firstly, in terms of data storage, with data distributed across local servers, the cloud, and various devices, it is critical to ensure its security across these environments. Encryption plays a vital role in protecting this data. By employing suitable encryption algorithms and adhering to stringent key management protocols, unauthorized access or tampering with stored data can be prevented. However, relying solely on encryption is not enough. It is also essential to implement comprehensive data backup, recovery, and disaster response strategies to maintain data integrity and ensure continuity in the event of unforeseen circumstances.

Secondly, regarding data transmission, a fundamental challenge is ensuring that data is neither intercepted nor altered during transit. To address this, secure transmission protocols like SSL/TLS are used to encrypt data, protecting its confidentiality and integrity while it moves between networks. Additionally, implementing advanced network security strategies, such as intrusion detection and prevention systems, provides further protection by monitoring and responding to potential malicious activities, thus enhancing overall security.

Finally, data access control is a crucial component of data security. In environments where multiple users and applications are present, it is essential to ensure that only authorized users and systems can access data. One effective strategy is the principle of "least privilege," which limits users to access only the data necessary for their roles. Additionally, enhancing security through mechanisms such as multi-factor authentication and role-based access control further strengthens data access management. Regular audits and continuous monitoring are also vital, as they help quickly detect and respond to unauthorized or inappropriate access to data.

In summary, data security on a technical level encompasses several interrelated key areas that collectively ensure the confidentiality, integrity, and availability of data. Given the continuous evolution of technology and the emergence of new threats, ongoing innovation and regular updates in these areas are crucial to maintaining robust data security.

(3)Managemant Perspective.In today's digital age, data is a critical asset, integral to an organization's continuous operation and competitive edge. However, data security transcends technical solutions, presenting a multifaceted management challenge that encompasses organizational strategy, human resources, and operational processes. From a management perspective, data security can be divided into three primary components: strategic management, human resource management, and operational process management.

In strategic management, it is essential for data security measures to be in sync with the broader business strategy of the organization. This alignment necessitates a profound understanding by senior management of the importance of data security, ensuring it is woven into long-term planning. Moreover, organizations must regularly perform data security risk assessments to identify potential threats and vulnerabilities, crafting tailored risk management strategies based on these findings. These steps are crucial for maintaining the integrity and security of organizational data.

In the realm of human resource management, employees can represent a significant vulnerability in data security. Consequently, organizations must prioritize regular data security training to ensure employees are well-versed in best practices for data protection. Clearly defining each employee’s role and responsibilities in safeguarding data can help mitigate risks associated with mishandling and unauthorized activities. Moreover, it is crucial for organizations to appoint dedicated teams or individuals specifically responsible for managing data security.

In terms of operational process management, it is essential for organizations to implement standardized processes that cover the entire

data lifecycle—from generation and storage to usage and eventual destruction—ensuring comprehensive protection at every stage. Furthermore, in the event of data breaches or other security incidents, having a pre-established emergency response plan is critical. Such preparedness enables quick and effective actions to address threats, thereby reducing potential damage and enhancing overall resilience.

In summary, from a management perspective, data security extends beyond the confines of the IT department, encompassing a multidisciplinary challenge that implicates every level of the organization. To effectively safeguard data, organizations must adopt a holistic approach, implementing comprehensive strategies and responses that span strategic planning, human resource management, and operational processes.

4.2 Research Trend Prognosis

An integrated analysis of keyword co-occurrence maps and clustering outcomes has identified three emerging trends in data security research.

4.2.1 Artificial Intelligence and Data Security

In examining the relationship between artificial intelligence and data security, a comprehensive and systematic analysis of their fundamental characteristics and impacts is essential..Primarily, data privacy emerges as a critical concern, necessitating strict adherence to relevant legal and normative frameworks to avert privacy violations . Additionally, machine learning models are susceptible to adversarial attacks, potentially leading to misclassification or data breaches, thereby underscoring the importance of enhancing model robustness and security. Concurrently, the widespread application of AI technologies poses risks of malicious use in cyber-attacks or misinformation propagation further accentuating the need for stringent regulation and legal compliance. Notably, the leakage of machine learning models could reveal sensitive information in training data, highlighting the urgency of privacy protection. Moreover, considerations of dataset diversity and algorithmic fairness are crucial to avoid biases and injustices. Additionally, system vulnerabilities pose latent threats to AI, making the fortification of overall system security paramount. Given automated threats such as botnet attacks, enhancing network and system security becomes even more essential. Ultimately, the accuracy and integrity of data are pivotal to decision-making precision, emphasizing the need to ensure data quality to prevent misjudgments and erroneous decisions[3-13]. In summary, achieving a harmonious integration of artificial intelligence and data security necessitates a comprehensive approach that encompasses technical, regulatory, ethical, and policy dimensions. This multidisciplinary strategy is essential to effectively safeguard data privacy and system security.

4.2.2 Blockchain and Data Security

In the modern realm of data security, blockchain technology, characterized by its distributed, immutable, and transparent nature, is perceived as a potentially revolutionary strategy for resolving data security vulnerabilities. The cornerstone of blockchain is its distributed ledger system, where data is not centralized in a single location but rather dispersed across various nodes in the network.When combined with authorized key access, this mechanism significantly enhances the security and privacy of data effectively countering unauthorized data access.Moreover, its inherent immutability ensures that once data is recorded on the blockchain, it becomes exceedingly difficult to modify or tamper with, providing an additional layer of security. Concurrently, each data block is linked to its predecessor, forming an ever-growing chain. This structured linkage not only provides a clear basis for the detection of data tampering but also enhances the integrity of the data. Notably, compared to traditional centralized storage, blockchain's data is stored across multiple dispersed nodes, significantly reducing the risk of single-point attacks. Furthermore, smart contracts on the blockchain, as self-executing programs, ensure that data exchanges and processing always adhere to established rules, preventing improper use or malicious tampering[14-21]. Coupled with its inherent transparency, blockchain ensures that all network participants can review data on the chain, thus increasing the fairness of data processing and reducing the risk of biases. Despite its many advantages in data security blockchain also faces challenges in energy consumption, scalability, and complianceNevertheless, despite these

challenges, blockchain continues to demonstrate its revolutionary potential in the field of data security[22-27].

4.2.3 Personal Information and Data Security

In the digital age, the handling of personal information and data security issues entails considerations on multiple fronts, as outlined below. Primarily, the protection of data privacy occupies a central position due to its direct impact on the security of personal information[28]. Specifically, the leakage of personal information can lead directly to illicit activities such as identity theft[29].Security assessment of the Spanish contactless identity card] and fraud[30]. Secondly, data breaches are an urgent issue, stemming from various factors including, but not limited to, cyber attacks, negligence or misconduct by insiders, and the insecurity of information storage[31]. To counter this threat, organizations must implement stringent security measures aimed at preventing, detecting, and promptly responding to any potential data breach incidents.

Concurrently, identity theft is not merely a consequence of data breaches but also an independent security threat, as attackers might use illegally obtained personal information for a range of fraudulent activities. Moreover, social engineering attacks, which often employ such tactics, manipulate and deceive targets using personal information to achieve unlawful objectives [32]. Additionally, improper data sharing and lack of appropriate compliance measures also pose risks to data security, necessitating stricter regulatory measures and ensuring transparency by industry and government bodies[33]. Furthermore, issues concerning data quality[34]and the misuse of biometric information[35]are also worthy of attention.

In sum, ensuring the accuracy and integrity of data, as well as balancing legal data use with the protection of personal privacy, is of paramount importance. Therefore, to ensure comprehensive protection of personal information in the digital era, a holistic and multi-faceted approach is essential in addressing the aforementioned issues.

5. Conclusion

This study utilizes bibliometric techniques and scientific knowledge mapping visualization to statistically analyze literature in the field of data security from 2013 to 2023, drawing from the Web of Science Core Collection database. The analysis yielded the following key conclusions:

(1) The frequent occurrence of data security incidents has spurred extensive research in this field, resulting in a robust body of work that shows a consistent year-on-year growth. Notably, since 2018, there has been a significant increase in the volume of publications from international authors in this area. Regarding the nationality of the literature's authors, China remains a dominant force, consistently leading and shaping the direction of data security research.

(2) Journal publication statistics indicate that the field of data security research is becoming increasingly diversified. Key publications such as IEEE Access, Multimedia Tools and Applications, and Security and Communication Networks have contributed significantly to this area, serving as essential platforms for its development. Regarding the disciplinary distribution, the literature in international journals primarily centers on computer science and medicine, covering areas like interdisciplinary applications, artificial intelligence theories and methodologies, and information systems cybernetics.

(3) Analysis of keywords in data security literature reveals that research predominantly focuses on three areas: technology, law, and management. Technologically, the emphasis is on the data lifecycle and the development of platforms to support data security management and operations. Legally, the research is geared towards establishing support and norms through frameworks such as sovereignty, property rights, and personality rights. From a management perspective, the literature highlights the importance of comprehensive governance, with various countries devising unique concepts and strategies. These include multi-dimensional approaches like hierarchical classification and talent cultivation to enhance data security governance.

(4) Emergent keyword analysis indicates that research themes in the field of data security are both diverse and evolving. Over the past two years, terms such as "personal information,""blockchain," and "artificial intelligence" have become increasingly prevalent. This trend suggests that these topics have gained significant traction in data security research since 2019 and are likely to shape

future research directions.

References

[1]Ma, D., Yang, B., Guan, B., Song, L., Liu, Q., Fan, Y., ... & Xu, H. (2021). A bibliometric analysis of pyroptosis from 2001 to 2021. Frontiers in immunology, 12, 731933.

[2]Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics, 84(2), 523-538.

[3]Carre, J. R., Curtis, S. R., & Jones, D. N. (2018). Ascribing responsibility for online security and data breaches. Managerial Auditing Journal, 33(4), 436-446.

[4]Yang, P., Xiong, N., & Ren, J. (2020). Data security and privacy protection for cloud storage: A survey. IEEE Access, 8, 131723-131740.

[5]Liu, X., Zhao, J., Li, J., Cao, B., & Lv, Z. (2022). Federated neural architecture search for medical data security. IEEE transactions on industrial informatics, 18(8), 5628-5636.

[6]Goldblum, M., Tsipras, D., Xie, C., Chen, X., Schwarzschild, A., Song, D., ... & Goldstein, T. (2022). Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 1563-1580.

[7]Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications surveys & tutorials, 18(2), 1153-1176.

[8]Mohammad, A. S., & Pradhan, M. R. (2021). Machine learning with big data analytics for cloud security. Computers & Electrical Engineering, 96, 107527.

[9]Zhang, N., Jia, H., Hou, Q., Zhang, Z., Xia, T., Cai, X., & Wang, J. (2022). Data-driven security and stability rule in high renewable penetrated power system operation. Proceedings of the IEEE.

[10]Coulter, R., Han, Q. L., Pan, L., Zhang, J., & Xiang, Y. (2019). Data-driven cyber security in perspective—Intelligent traffic analysis. IEEE transactions on cybernetics, 50(7), 3081-3093.

[11]Sullivan, G. (2022). Law, technology, and data-driven security: infra-legalities as method assemblage. Journal of Law and Society, 49, S31-S50.

[12]Wang, Y. G., Zhu, G., Li, J., Conti, M., & Huang, J. (2020). Defeating lattice-based data hiding code via decoding security hole. IEEE Transactions on Circuits and Systems for Video Technology, 31(1), 76-87.

[13]Yang, J., Meng, Q., Lu, W., Fu, L., & Zhang, B. (2022). Data gathering, quality and security in intelligent phytoprotection. Frontiers in Plant Science, 13, 1010525.

[14]Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A survey on blockchain for information systems management and security. Information Processing & Management, 58(1), 102397.

[15]Soltanisehat, L., Alizadeh, R., Hao, H., & Choo, K. K. R. (2020). Technical, temporal, and spatial research challenges and opportunities in blockchain-based healthcare: A systematic literature review. IEEE Transactions on Engineering Management, 70(1), 353-368.

[16]Soltanisehat, L., Alizadeh, R., Hao, H., & Choo, K. K. R. (2020). Technical, temporal, and spatial research challenges and opportunities in blockchain-based healthcare: A systematic literature review. IEEE Transactions on Engineering Management, 70(1), 353-368.

[17]Suhail, S., Hussain, R., Jurdak, R., & Hong, C. S. (2021). Trustworthy digital twins in the industrial internet of things with blockchain. IEEE Internet Computing, 26(3), 58-67.

[18]Deepa, N., Pham, Q. V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., ... & Pathirana, P. N. (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. Future Generation Computer Systems, 131, 209-226.

[19]Liang, W., Xiao, L., Zhang, K., Tang, M., He, D., & Li, K. C. (2021). Data fusion approach for collaborative anomaly intrusion detection in blockchain-based systems. IEEE Internet of Things Journal, 9(16), 14741-14751.

[20]Javed, A. R., Hassan, M. A., Shahzad, F., Ahmed, W., Singh, S., Baker, T., & Gadekallu, T. R. (2022). Integration of blockchain technology and federated learning in vehicular (iot) networks: A comprehensive survey. Sensors, 22(12), 4394.

[21]Griggs, K. N., Ossipova, O., Kohlios, C. P.,

Baccarini, A. N., Howson, E. A., & Hayajneh, T. (2018). Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. Journal of medical systems, 42, 1-7.

[22]Yaqoob, I., Salah, K., Jayaraman, R., & Al-Hammadi, Y. (2021). Blockchain for healthcare data management: opportunities, challenges, and future recommendations. Neural Computing and Applications, 1-16.

[23]Sehar, N. U., Khalid, O., Khan, I. A., Rehman, F., Fayyaz, M. A., Ansari, A. R., & Nawaz, R. (2023). Blockchain enabled data security in vehicular networks. Scientific Reports, 13(1), 4412.

[24]Deng, S., Hu, Q., Wu, D., & He, Y. (2023). BCTC-KSM: A blockchain-assisted threshold cryptography for key security management in power IoT data sharing. Computers and Electrical Engineering, 108, 108666.

[25]Mohanty, S. P., Yanambaka, V. P., Kougianos, E., & Puthal, D. (2020). PUFchain: A hardware-assisted blockchain for sustainable simultaneous device and data security in the internet of everything (IoE). IEEE Consumer Electronics Magazine, 9(2), 8-16.

[26]Khalid, M. I., Ahmed, M., & Kim, J. (2023). Enhancing data protection in dynamic consent management systems: formalizing privacy and security definitions with differential privacy, decentralization, and Zero-Knowledge proofs. Sensors, 23(17), 7604.

[27]Chintapalli, S. S. N., Paramesh, S. P., Nijaguna, G. S., Jeyaraj, J. R. A., & Subhash, P. (2023). Controlled blockchain enabled data record security for healthcare applications. Neural Computing and Applications, 1-13.

[28]Liu, X., Zhao, J., Li, J., Cao, B., & Lv, Z. (2022). Federated neural architecture search for medical data security. IEEE transactions on industrial informatics, 18(8), 5628-5636.

[29]Rodríguez, R. J., & Garcia-Escartin, J. C. (2017). Security assessment of the Spanish contactless identity card. IET Information Security, 11(6), 386-393.

[30]van Rensburg, S. K. J. (2021). End-user perceptions on information security: Pragmatic lessons on social engineering attacks in the workplace in Gauteng, South Africa. Journal of Global Information Management (JGIM), 29(6), 1-16.

[31]Mahajan, H. B., Rashid, A. S., Junnarkar, A. A., Uke, N., Deshpande, S. D., Futane, P. R., ... & Alhayani, B. (2023). Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems. Applied Nanoscience, 13(3), 2329-2342.

[32]van Rensburg, S. K. J. (2021). End-user perceptions on information security: Pragmatic lessons on social engineering attacks in the workplace in Gauteng, South Africa. Journal of Global Information Management (JGIM), 29(6), 1-16.

[33]Winter, J. S., & Davidson, E. (2022). Harmonizing regulatory regimes for the governance of patient-generated health data. Telecommunications Policy, 46(5), 102285.

[34]Shin, S. Y., & Kim, H. S. (2021). Data pseudonymization in a range that does not affect data quality: correlation with the degree of participation of clinicians. Journal of Korean Medical Science, 36(44).

[35]Paik, S., Mays, K. K., & Katz, J. E. (2022). Invasive Yet Inevitable? Privacy Normalization Trends in Biometric Technology. Social Media+ Society, 8(4), 20563051221129147.