Iraqi Journal for Computer Science and Mathematics

Volume 5 | Issue 4 Article 22

2024

Secure Healthcare Systems and Big Data: A Bibliometrics Analysis

Rasha Talal Hameed
College of Education, Al-Iraqia University, Iraq, rashatalal2015@gmail.com

Saad Ahmed Dheyab
College of Engineering, University of Information Technology and Communications, Iraq, saad.theyab@uotic.edu.iq

Saba Abdulbaqi Salman
College of Education, Al-Iraqia University, Iraq, sabasalman2019@gmail.com

Ahmed Hussein Ali
College of Education, Al-Iraqia University, Iraq, ahmed.ali@aliraqia.edu.iq

Omar Abdulwahabe Mohamad
College of Education, Al-Iraqia University, Iraq, engit2020@gmail.com

Recommended Citation

Hameed, Rasha Talal; Dheyab, Saad Ahmed; Salman, Saba Abdulbaqi; Ali, Ahmed Hussein; and Mohamad, Omar Abdulwahabe (2024) "Secure Healthcare Systems and Big Data: A Bibliometrics Analysis," *Iraqi Journal for Computer Science and Mathematics*: Vol. 5: Iss. 4, Article 22.
DOI: https://doi.org/10.52866/2788-7421.1225
Available at: https://ijcsm.researchcommons.org/ijcsm/vol5/iss4/22

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2024;5:274–298

RESEARCH ARTICLE

Secure Healthcare Systems and Big Data: A Bibliometrics Analysis

Rasha Talal Hameed [a], Saad Ahmed Dheyab [b], Saba Abdulbaqi Salman [a], Ahmed Hussein Ali [a], Omar Abdulwahabe Mohamad [a,*]

[a] College of Education, Al-Iraqia University, Iraq
[b] College of Engineering, University of Information Technology and Communications, Iraq

ABSTRACT

This study conducts a bibliometrics analysis of research on secure healthcare systems and big data, aiming to identify trends, key contributors, and thematic areas within the field. By examining a comprehensive database of academic publications, we highlight the evolution of research from foundational concepts to contemporary innovations in data security and privacy management in healthcare. Key metrics such as publication volume, citation impact, and coauthorship networks are analyzed to uncover the most influential authors and institutions. Additionally, we explore the integration of big data analytics in enhancing healthcare delivery while addressing security challenges. The findings provide valuable insights for researchers and practitioners, emphasizing the importance of developing robust, secure frameworks that leverage big data to improve patient outcomes and operational efficiency in healthcare systems. This analysis serves as a foundational resource for future studies and strategic initiatives in securing health information systems amid the growing reliance on big data technologies.

Keywords: Healthcare, Big data, Secure healthcare, Bibliometrics

1. Introduction

Big Data refers to the vast volumes of data that are generated, collected, and analyzed to uncover patterns, trends, and insights that were previously impossible to detect with traditional data processing tools. Big Data is revolutionizing the healthcare industry by offering new ways to enhance patient care, optimize operational efficiency, and drive medical research [1, 2]. An introduction to how big data is being applied in healthcare [3, 4]:

a. **Volume:** The healthcare industry generates vast amounts of data daily, including electronic health records (EHRs), medical imaging, genomics, wearable device data, and patient feedback.
b. **Velocity:** Data is created and needs to be processed in real-time or near-real-time. For example, wearable devices constantly monitor and transmit patient data.
c. **Variety:** Healthcare data comes in various forms, such as structured data (e.g., patient demographics, lab results), semi-structured data (e.g., clinical notes), and unstructured data (e.g., medical images, voice recordings).
d. **Veracity:** Ensuring the accuracy and reliability of data is crucial for effective decision-making in healthcare.
e. **Value:** The ultimate goal is to derive actionable insights that can improve patient outcomes, enhance efficiency, and reduce costs.

Received 8 October 2024; accepted 20 November 2024.
Available online 16 December 2024

* Corresponding author.
E-mail addresses: rashatalal2015@gmail.com (R. T. Hameed), saad.theyab@uotic.edu.iq (S. A. Dheyab), sabasalman2019@gmail.com (S. A. Salman), ahmed.ali@aliraqia.edu.iq (A. H. Ali), engit2020@gmail.com (O. A. Mohamad).

https://doi.org/10.52866/2788-7421.1225

2. Applications of big data in healthcare

Big data is having a profound impact on the healthcare industry, enhancing patient care, improving operational efficiency, and driving medical research. Here are some key applications of big data in healthcare [5–7]:

a. **Personalized Medicine:**
   Genomics and Precision Medicine: By analyzing genetic information alongside patient data, healthcare providers can tailor treatments to individual genetic profiles, potentially improving efficacy and reducing adverse effects.
   Risk Prediction: Big data can help identify individuals at high risk for certain conditions, allowing for early interventions.
b. **Predictive Analytics:**
   Disease Outbreaks: Analyzing patterns and trends in data can help predict and manage outbreaks of diseases.
   Patient Readmissions: Predictive models can identify patients at risk of readmission, allowing for targeted interventions to prevent it.
c. **Operational Efficiency:**
   Resource Management: Big data helps hospitals and clinics optimize the use of resources, such as staff and equipment, by analyzing patient flow and treatment patterns.
   Supply Chain Management: Analyzing data related to inventory and usage can improve the management of medical supplies and reduce waste.
d. **Clinical Decision Support:**
   Evidence-Based Guidelines: Integrating data from various sources can help in developing and updating clinical guidelines based on the latest research and outcomes.
   Decision-Making: Data analytics tools can provide real-time support to clinicians by suggesting treatment options and flagging potential issues.
e. **Patient Engagement and Satisfaction:**
   Behavioral Insights: Analyzing patient feedback and behavior data helps healthcare providers improve patient experiences and tailor communications.
   Self-Management Tools: Wearable devices and mobile health apps collect data that helps patients manage chronic conditions and adhere to treatment plans.
f. **Medical Research:**
   Clinical Trials: Big data accelerates the identification of suitable candidates for clinical trials and analyzes results more comprehensively.
   Drug Discovery: Large datasets are used to identify potential drug candidates and understand their effects more quickly.

3. Challenges and considerations

The application of big data in healthcare offers transformative potential but also comes with several challenges and considerations. Addressing these challenges is essential to fully leverage big data for improving patient care, operational efficiency, and research outcomes. A detailed look at the key challenges and considerations [7–11]:

1. **Data Privacy and Security:** Ensuring patient data is protected against breaches and unauthorized access is a major concern. Compliance with regulations such as HIPAA (Health Insurance Portability and Accountability Act) is essential.
2. **Data Integration:** Combining data from various sources (e.g., EHRs, wearable devices) can be complex, requiring interoperability between different systems.
3. **Data Quality:** The accuracy and completeness of healthcare data are crucial for reliable analysis and decision-making.
4. **Ethical Issues:** The use of big data in healthcare raises ethical concerns regarding consent, data ownership, and the potential for discrimination based on health data.
5. **Skill Requirements:** Analyzing big data requires specialized skills in data science, statistics, and domain knowledge in healthcare.

4. Key components of the bibliometric analysis

A bibliometric analysis of our paper "Secure Healthcare Systems and Big Data" involves evaluating and summarizing the scholarly literature related to the intersection of cyber security in healthcare systems and the application of big data. This analysis aims to provide insights into the research trends, key authors, influential papers, and thematic areas within this field [10, 12–14].

Scope and Objectives: To understand the evolution of research related to securing healthcare systems in the context of big data.

Data Collection: Databases: Commonly used databases include PubMed, IEEE Xplore, Scopus, and Google Scholar. **Search Terms:** Relevant search terms might include "secure healthcare systems," "Big Data in healthcare," "cybersecurity in healthcare," "healthcare data security," and "data protection in healthcare."

Key Metrics: Publication Count: Number of publications over time. **Authors:** Leading authors and their contributions. **Journals:** Key journals where significant research is published. **Citations:** Highly cited papers indicating influential work. **Keywords:** Commonly used terms and emerging topics.

Analysis and Visualization: Trends Over Time: Track the growth of publications in this area to identify increasing interest or emerging trends. **Authorship Analysis:** Identify prolific authors and research groups. **Journal Analysis:** Determine which journals are most influential in publishing research on this topic. **Citation Analysis:** Identify seminal papers and influential research. **Keyword Analysis:** Examine frequent keywords to understand research focus areas and shifts.

5. Key findings from bibliometric analysis

Research Trends:

Growth: There is a growing body of literature on the intersection of Big Data and cyber-security in healthcare, reflecting increased awareness and concern about data breaches and privacy issues. **Emerging Topics:** Research may focus on encryption methods, access control, threat detection, and compliance with regulations such as GDPR or HIPAA [13–15].

Influential Papers and Authors:

Seminal Papers: Key papers that have shaped the field, often addressing foundational concepts or novel approaches to securing healthcare data. **Leading Authors:** Researchers who have published extensively and have made significant contributions to the field.

Journal Analysis:

Top Journals: Journals like Journal of Biomedical Informatics, IEEE Transactions on Information Forensics and Security, and Health Information Science and Systems might feature prominently. **Impact Factor:** High-impact journals that publish influential studies.

Keyword and Thematic Analysis:

Keywords: Commonly used terms might include "data encryption," "privacypreserving techniques," "cyber-attack prevention," and "health information security." **Themes:** Major themes could involve securing Big Data infrastructure, regulatory compliance, risk management, and integrating security measures with data analytics.

Geographical and Institutional Analysis:

Research Institutions: Institutions with significant contributions to the field. **Geographical Distribution:** Regions or countries leading in research, reflecting differences in regulatory environments and healthcare systems.

Table 1. Main information.

Description	Results
Main information about data	
Timespan	2020:2023
Sources (Journals, Books, etc)	252
Documents	377
Annual growth rate %	−1.3
Document average age	1.5
Average citations per doc	12.59
References	1
Document Contents	
Keywords plus (ID)	2318
Author's keywords (DE)	1072
Authors	
Authors	1393
Authors of single-authored docs	25
Authors collaboration	
Single-authored docs	26
Co-authors per doc	4.13
International co-authorships %	36.34
Document types	
Article	155
Book	3
Book chapter	66
Conference paper	109
Conference paper book chapter	2
Conference paper conference paper	1
Note	1
Review	40

6. A bibliometrics analysis

Table 1 shows the main information about our paper provided as:

Main Information:

Timespan: The data covers the period from 2020 to 2023. This indicates that the dataset is relatively recent, which is important for research and analysis as it includes current publications. **Sources:** There are 252 sources that contribute to this dataset. This suggests a diverse set of references, including journals, books, and other types of publications, which can enrich the variety of information and perspectives available for analysis. **Documents:** There are a total of 377 documents in the dataset. This is a substantial number and provides a significant amount of data for analysis, indicating a comprehensive collection of research material. **Annual Growth Rate %:** The dataset has an annual growth rate of −1.3%. A negative growth rate suggests a slight decrease in the number of documents over time. It's essential to consider whether this decline is due to a reduction in relevant research or other factors. **Document Average Age:** The average age of documents in the dataset is 1.5

years. This means that, on average, the documents are relatively recent, which is beneficial for staying up-to-date with the latest research. **Average Citations per Document:** The dataset has an average of 12.59 citations per document. A high average citation count indicates that the documents in the dataset are well-cited, potentially reflecting their significance in the field. **References:** On average, there is 1 reference per document. While this number appears low, it might be due to a specific type of document or research area where fewer references are typically included.

Document Contents:

Keywords Plus (ID): There are 2,318 Keywords Plus (ID) in the dataset. These could be used to perform more granular searches and analyze trends in research topics and themes. **Author's Keywords (DE):** There are 1,072 Author's Keywords (DE) in the dataset. These keywords represent the topics and themes that authors themselves have associated with their research. They can be valuable for understanding the primary focus of the documents.

Authors:

Authors: There are 1,393 authors contributing to the dataset, indicating a broad range of researchers and scholars involved in this body of work. **Authors of Single-authored Docs:** There are 25 single-authored documents, showing that a small portion of the documents has a single author.

Authors Collaboration:

Single-authored Docs: There are 26 single-authored documents, suggesting that a majority of the documents involve collaboration among authors. **CoAuthors per Doc:** On average, there are 4.13 co-authors per document. This indicates a relatively high level of collaboration among authors, which is common in academic research. **International Co-authorships %:** Approximately 36.34% of the collaborations involve international co-authors. This international collaboration can lead to diverse perspectives and expertise.

Document Types:

The dataset includes various types of documents, including articles, books, book chapters, conference papers, reviews, and others. This diversity can provide a wide range of information sources for analysis [15].

In summary, this dataset is relatively recent, contains a substantial number of documents from various sources, has well-cited research, and represents a diverse set of research topics. The collaborative nature of the research, including international collaboration, suggests that it covers a broad range of perspectives and experts.

Table 2. Number of articles per year with growth rate.

Year	Number of articles	Annual growth rate
2020	78	–
2021	107	37.18%
2022	117	9.35%
2023	75	−36.75%

6.1. Annual scientific production

An analysis of the annual scientific production based on the provided data as shown in Table 2 and Fig. 1.

The Analysis:

1. In 2020, there were 78 articles published.
2. In 2021, the number of articles increased to 107, showing a significant annual growth rate of 37.18%.
3. In 2022, the number of articles continued to increase to 117, with a more moderate annual growth rate of 9.35%.
4. In 2023, there was a decrease in the number of articles to 75, resulting in a negative annual growth rate of −36.75%.

6.2. Average citations per year

Table 3 shows an analysis of the "Average Citations per Year".

The Analysis:

1. In 2020, the mean number of citations per article was 29.09. This resulted in a total of 7.27 mean citations per year when considering the number of articles published in that year. This is based on 4 citable years.
2. In 2021, the mean number of citations per article decreased to 14.35. This led to a mean of 4.78 citations per year, considering 3 citable years.
3. In 2022, the mean citations per article further decreased to 7.29, resulting in a mean of 3.64 citations per year based on 2 citable years.
4. In 2023, the mean number of citations per article was the lowest at 1.2, resulting in 1.2 citations per year, considering only 1 citable year.

Table 3. Average citations per year.

Year	Mean citations per article	Total number of articles (N)	Mean citations per article	Citable years
2020	29.09	78	7.27	4
2021	14.35	107	4.78	3
2022	7.29	117	3.64	2
2023	1.2	75	1.2	1

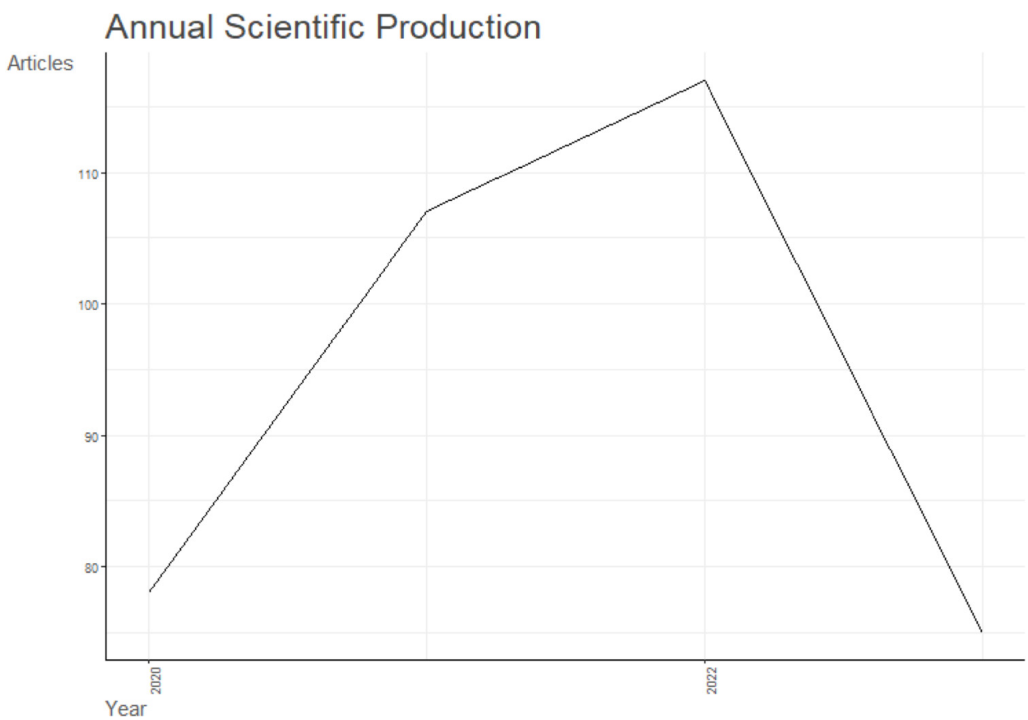

Fig. 1. Annual scientific production.

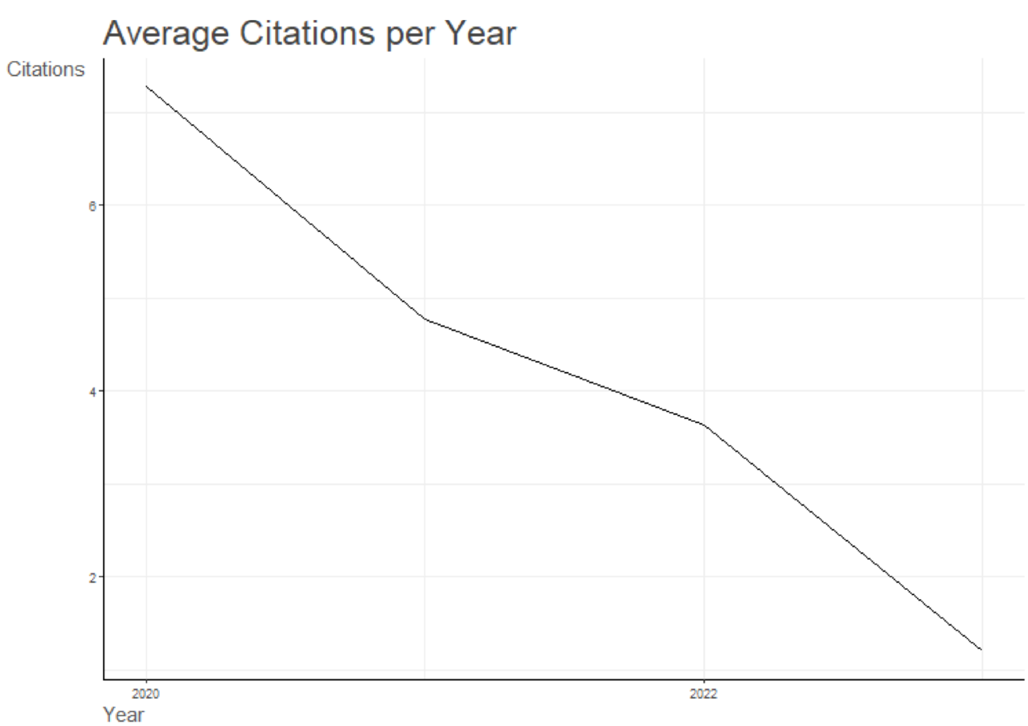

Fig. 2. Average citations per year.

The data shows a decreasing trend in mean citations per article and mean citations per year over the four years. This could indicate changes in the significance or impact of the articles over time.

6.3. Most relevant sources

It appears that a list of sources and the number of articles associated with each source. To analyze the most relevant sources from the list of papers provided. To do this, we would typically need to define a specific research topic or area of interest. There are a general approach to analyze and identify relevant sources for a research topic [16, 17]:

Define Your Research Topic: Clearly define the research area or topic you are interested in. For example, if your research topic is "Internet of Things (IoT) in Healthcare," you would look for sources related to that topic.

Keywords and Search Terms: Identify relevant keywords and search terms related to your research topic. For the example above, keywords might include "IoT in Healthcare," "Healthcare IoT Applications," "IoT in Medical Devices," etc.

Database Search: Use academic databases like PubMed, IEEE Xplore, Google Scholar, or your institution's library resources to search for articles, papers, and publications related to your research topic.

Filter and Sort: When you get search results, use filters and sorting options to identify the most relevant sources. You can often sort by relevance, date, citation count, or other criteria.

Review Abstracts: Read the abstracts of the articles to get an overview of their content. This will help you determine if the source is directly related to your research.

Citation Count: Check the number of citations for each source. Highly cited sources are often considered more relevant.

Author Reputation: Consider the reputation of the authors. Well-known researchers in the field may produce more reliable and relevant work.

Peer-Reviewed Journals: Sources published in reputable, peer-reviewed journals are generally more reliable and relevant.

Research Focus: Consider the focus of the source. Does it address your research question or provide information that is directly related to your topic

Cross-Referencing: Look at the references within the sources. This can lead us to additional relevant sources.

6.4. Core sources by Bradford's law

Bradford's Law is a bibliometric and information retrieval concept that describes the distribution of scientific literature in a particular field. It is often used to identify core sources, or journals, in a given subject area. The law suggests that a small number of journals (core) will contribute a significant portion of the published articles in that field. These core sources are surrounded by a larger number of peripheral journals. The distribution can often be approximated by a geometric progression. Table 4 shows the examples of average citations per year [18].

Bradford's Law is typically illustrated in Fig. 3. By using the Bradford's Zones or Bradford's Graph, which shows the distribution of articles across the core, the first zone, and the second zone. Here's a basic explanation of how it works:

Bradford's Zones: In Bradford's Zones, the core sources are the most productive and important journals, while the first zone and second zone represent journals with diminishing importance. **Bradford's Graph:** A Bradford's Graph is a visual representation of this concept. It typically shows the number of journals in each zone, with the core zone being the smallest but containing a significant portion of the articles, followed by the first and second zones. **Geometric Progression:** Bradford's Law suggests that the number of core journals is about one-third of the total number of articles in a specific field. The first zone will contain about two-thirds of the articles, and the second zone will contain even fewer.

A typical representation might look something like this:

The Copy code:

Zone 1: Core Journals - 1/3 of articles Zone 2: First Supplementary Zone - 2/3 of articles Zone 3: Second Supplementary Zone - Much smaller share of articles

The simplified example:

Suppose you want to identify the core journals in the field of artificial intelligence. You analyze a large database of AI articles and find that there are 1000 articles in the field. According to Bradford's Law, you might expect to find approximately one-third of the articles in the core journals, so around 333 articles would be published in the core journals. The remaining two-thirds would be distributed among the first and second supplementary zones.

Bradford's Law can help librarians, researchers, and information professionals make decisions about which journals to subscribe to or which sources to consult when conducting literature reviews in a specific field. It's important to note that Bradford's Law is a heuristic and may not hold precisely in all cases, but it provides a useful guideline for organizing and prioritizing information sources.

7. Sources' local impact by H index

The term "Sources' local impact by the H index" likely refers to the evaluation of the influence or impact of academic sources (such as journals or articles) within a specific context or community, using the

Table 4. Average citations per year.

Source	Rank	Freq	cumFreq	Zone
Lecture Notes in Networks and Systems	1	11	11	Zone 1
IEEE Access	2	10	21	Zone 1
Internet of Things	3	8	29	Zone 1
Communications in Computer and Information Science	4	7	36	Zone 1
IEEE Internet of Things Journal	5	6	42	Zone 1
Advances in Intelligent Systems and Computing	6	5	47	Zone 1
Electronics (Switzerland)	7	5	52	Zone 1
Future Generation Computer Systems	8	5	57	Zone 1
Intelligent Systems Reference Library	9	5	62	Zone 1
Wireless Personal Communications	10	5	67	Zone 1
ACM International Conference Proceeding Series	30	2	129	Zone 2
Artificial Intelligence In Medicine	31	2	131	Zone 2
Block Chain for 5G-Enabled IoT: The New Wave for Industrial Automation	32	2	133	Zone 2
Ceur Workshop Proceedings	33	2	135	Zone 2
Computational and Mathematical Methods	35	2	139	Zone 2
Computer Science Review	36	2	141	Zone 2
Computers	37	2	143	Zone 2
Computers and Electrical Engineering	38	2	145	Zone 2
Computers, Materials and Continua	39	2	147	Zone 2
Artificial Intelligence Review	85	1	210	Zone 2
Handbook of Global Health: With 362 Figures and 152 Tables	130	1	255	Zone 3
Health and Technology	131	1	256	Zone 3
Health Informatics Journal	132	1	257	Zone 3
Healthcare Analytics	133	1	258	Zone 3
Heliyon	134	1	259	Zone 3
Heredity	135	1	260	Zone 3
Iberian Conference on Information Systems and Technologies, Cisti	136	1	261	Zone 3
Ict Express	138	1	263	Zone 3
IEEE Design And Test	139	1	264	Zone 3
IEEE Systems Journal	141	1	266	Zone 3

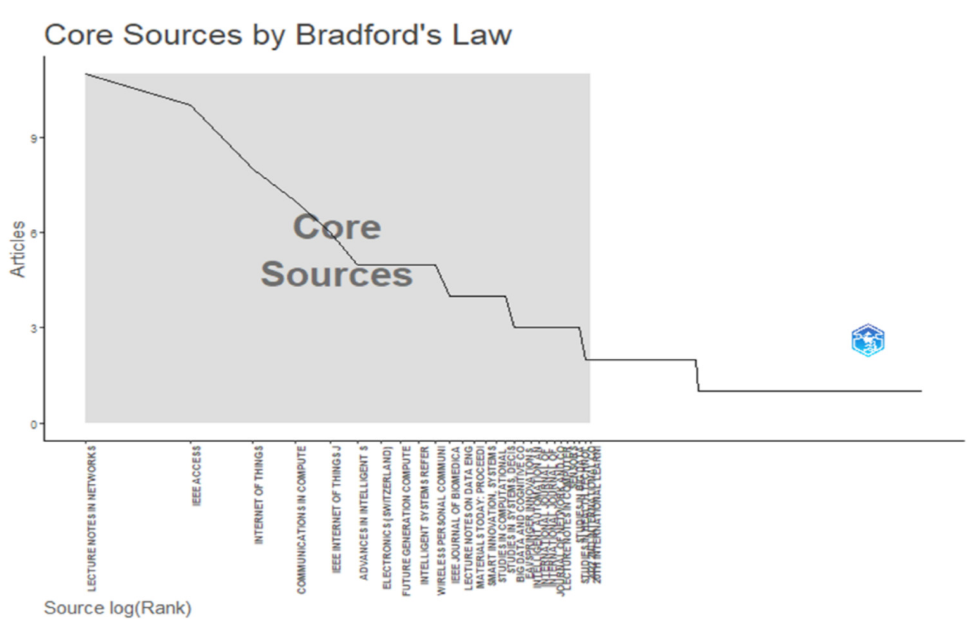

Fig. 3. Core sources by Bradford's law.

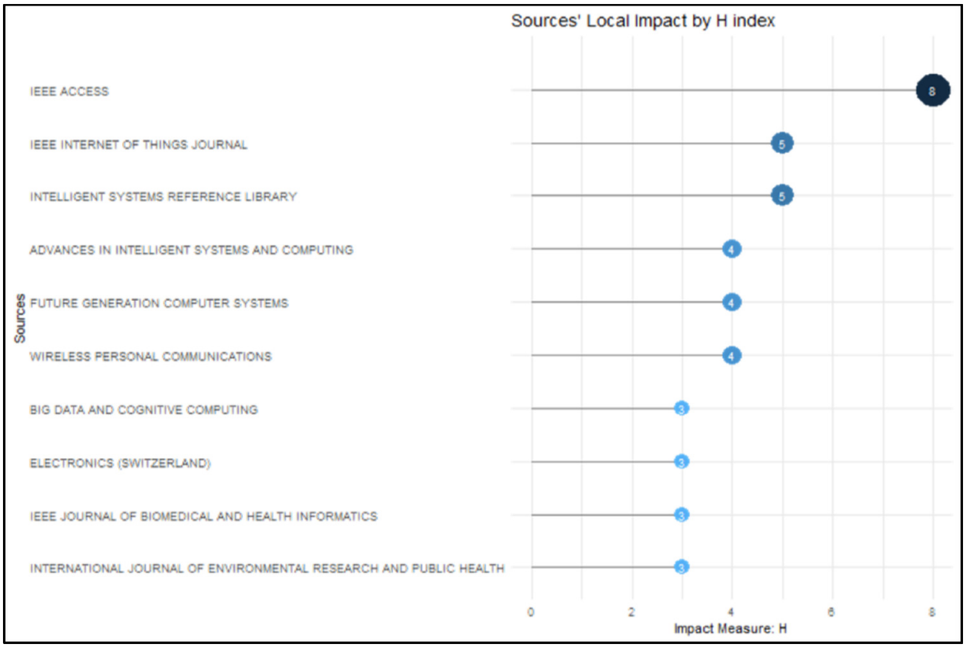

Fig. 4. Sources' local impact by the H index.

H-index as a metric. The Hindex is a quantitative measure that aims to assess both the productivity and citation impact of a researcher or a body of work. An H-index of "h" means that the author has "h" papers that have been cited at least "h" times each. When applied to sources, it gives an idea of how many articles from a particular source have received a significant number of citations [19, 20].

The "Local impact" could refer to the influence of these sources within a specific field, region, or institution, highlighting how well-regarded or impactful they are in that context. It suggests a more focused assessment compared to broader metrics that consider global citation patterns. In summary, the phrase implies analyzing the H-index of sources to understand their significance and influence within a particular local context as shown the examples in Table 5.

It appears that we have provided a table of information about various academic journals and conferences, including their H-index, G-index, M-index, total citations (TC), number of publications (NP), and the publication year (PY_start). The H-index is a metric that measures the productivity and impact of academic journals or researchers. For example, the data about **IEEE Access** as follow : (H-index: 8, G-index: 10, M-index: 2, Total Citations (TC): 635, Number of Publications (NP): 10, Publication Year (PY_start): 2020). The data about **IEEE INTERNET OF THINGS JOURNAL** as follow : (H-index: 5, Gindex: 6, M-index: 1.67, Total Citations (TC): 128, Number of Publications (NP): 6, Publication Year (PY_start): 2021). The data about **ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING** as follow : (H-index: 4, Gindex: 4, M-index: 1, Total Citations (TC): 22, Number of Publications (NP): 5, Publication Year (PY_start): 2020).

To analyze the local impact of these sources using the H-index, we can consider the following aspects:

Top Journals by H-index: Identify the journals with the highest H-index, as this is an indicator of their overall impact. In your dataset, "IEEE ACCESS" has the highest H -index of 8, making it one of the most impactful journals.

Impact Over Time: Consider the publication year (PY_start) and H-index to see how the impact of a source has evolved over the years. Newer sources may have lower H -indices because they haven't been around as long.

Comparison with G-index and M-index: While the H-index is a common measure of impact, also consider the G-index and M-index, which provide additional insights. For example, if a source has a much higher G-index than Hindex, it may have a few highly cited papers that significantly contribute to its impact.

Total Citations (TC): Look at the total citations to get a sense of the overall influence of a source. A source with a high H-index and a large number of total citations is likely to be highly influential.

Number of Publications (NP): Consider the number of publications, as this can affect the H -index. A source with a high H -index and a relatively low

Table 5. Sources' local impact by the H index.

m_index	TC	NP	PY_start
2	635	10	2020
1.66666667	128	6	2021
1	693	5	2020
1.33333333	44	5	2021
0.75	28	3	2020
1	27	5	2021
0.75	196	4	2020
1	245	3	2021
0.75	28	3	2020
0.66666667	13	2	2021
0.5	13	7	2020
0.66666667	37	2	2021
0.66666667	14	2	2021
1	36	2	2022
1	34	3	2022
0.66666667	9	2	2021
0.5	12	4	2020
1	10	2	2022
0.5	29	3	2020
0.66666667	8	2	2021
0.25	9	1	2020
0.3333333	7	1	2021
0.5	1	1	2022
1	2	1	2023
0.25	50	1	2020
0.3333333	1	1	2021
0.33333333	27	1	2021
1	5	1	2023
0.25	28	1	2020
1	1	1	2023
0.25	7	2	2020
0.33333333	22	1	2021
1	3	2	2023
0.5	2	1	2022
0.25	25	1	2020
0.5	19	1	2022
0.33333333	5	1	2021
1	3	1	2023
0.5	12	1	2022
0.33333333	2	1	2021
1	2	1	2023
0.25	2	1	2020
0.5	2	1	2022
0.5	3	1	2022
1	2	1	2023
0.33333333	4	1	2021
0.25	17	1	2020
0.33333333	1	1	2021
0.33333333	1	1	2021
0.25	3	1	2020
0.25	54	1	2020
1	4	1	2023
0.5	12	1	2022
0.33333333	5	1	2021
0.33333333	5	2	2021
0.33333333	5	4	2021
1	2	1	2023
0.5	5	1	2022

number of publications may have a more significant impact per paper.

Subject Area: It's essential to consider the subject area of each source. Different fields may have different expectations for impact. Some sources may be specialized and highly influential within their domain.

Recent Sources: Pay attention to recently established sources as they may take some time to accumulate citations and build their impact.

Comparative Analysis: Compare sources within the same subject area to understand their relative impact. Some fields may have more highly cited sources than others.

Remember that the H-index, G-index, and M-index are just a few of many metrics used to evaluate the impact of academic sources. They provide valuable information about the influence and productivity of sources, but a comprehensive assessment may involve considering various factors and context within a specific field of study.

8. Sources' production over time

The Table 6 and Fig. 5 show data on the production of articles or papers in various academic journals over a few years.

The analyze the data which provided on the production of articles or papers in different academic journals over the years 2020 to 2023.

The Analysis:

1. **Total Production by Year:**

- In 2020, there were a total of 6 publications.
- In 2021, there were a total of 18 publications.
- In 2022, there were a total of 29 publications.
- In 2023, there were a total of 42 publications.

2. **Growth Over Time:**

- The total production of articles increased significantly from 2020 to 2023, showing a clear upward trend.

3. **Journal Comparison:**

- In 2023, "IEEE INTERNET OF THINGS JOURNAL" had the highest number of publications (6), followed by "IEEE ACCESS" (10).
- "LECTURE NOTES IN NETWORKS AND SYSTEMS" showed significant growth from 1 publication in 2021 to 11 publications in 2023, the highest growth rate among these journals.

4. **Publication Trends:**

- "IEEE ACCESS" has been consistently producing a high number of publications over the years.
- "INTERNET OF THINGS" and "COMMUNICATIONS IN COMPUTER AND INFORMATION SCI-

Table 6. Sources' production over time.

Year	Lecture notes in networks and systems	IEEE access	Internet of thinks	Communications in computer and information science	IEEE internet of things journal
2020	0	4	0	2	0
2021	1	6	5	4	2
2022	3	9	6	6	5
2023	11	10	8	7	6

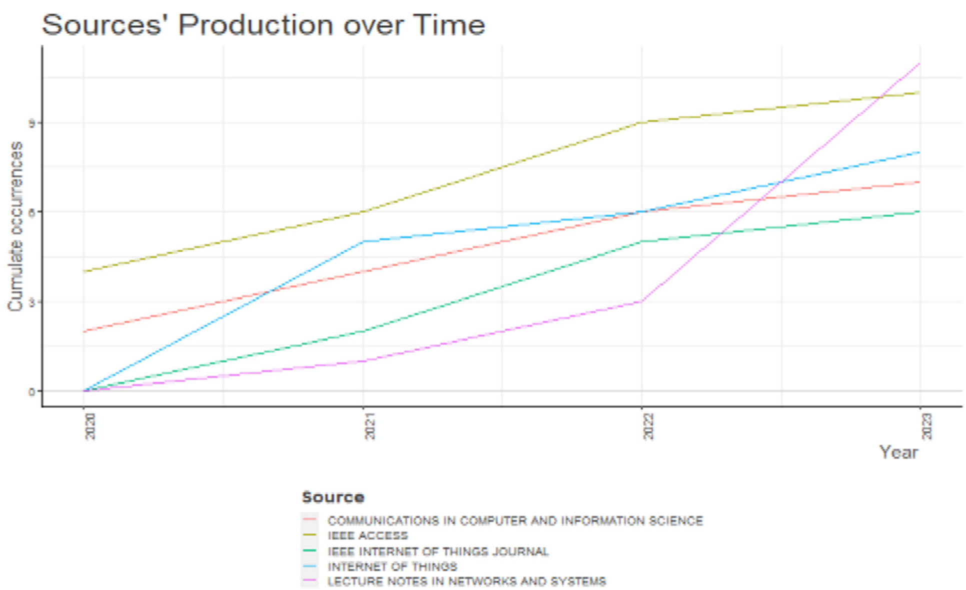

Fig. 5. Sources' production over time.

ENCE" have shown moderate growth over the years.
- "IEEE INTERNET OF THINGS JOURNAL" showed significant growth in the number of publications from 2021 to 2023.

9. Author productivity through lotka's law

The provided data on the distribution of the number of articles authored by a certain number of authors, along with the corresponding frequency [19]. This data is typically analyzed in the context of Lotka as shown in Table 7 and Fig. 6 respectively.

To analyze the author productivity data using Lotka's Law, we can calculate the exponent (alpha) for the power-law distribution. The exponent can be determined using the following equation:

$$f(x) = k \cdot x^{-\alpha}$$

where:

- (f(x)) is the frequency of authors producing (x) works,
- (k) is a constant,
- (x) is the number of works,
- (α) is the exponent that characterizes the distribution.

To estimate (α), we can use methods such as:

Maximum Likelihood Estimation (MLE): This involves fitting the data to the power-law model and finding the value of (α) that maximizes the likelihood function.

Least Squares Fitting: Transform the equation to a linear form (taking logarithms) and perform linear regression on the log-transformed data.

Empirical Distribution: Count the number of authors producing a specific number of works and fit

Table 7. Author productivity through lotka's law.

Number of articles	Number of authors	Frequency
1	1259	0.90380474
2	111	0.07968413
3	18	0.01292175
4	3	0.00215363
5	2	0.00143575

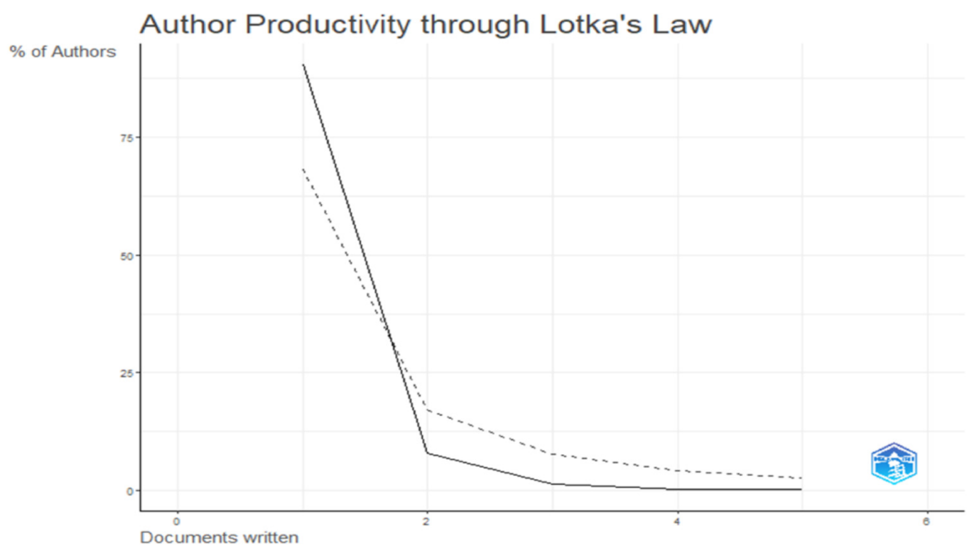

Fig. 6. Author productivity through lotka's law.

the resulting frequency distribution to the power-law model.

10. Affiliation production overtime

This approach will help in effectively analyzing how different affiliations contribute to research output over time. The Fig. 7 provides information about the number of articles published by different affiliations in the field of computer science and engineering over the years 2020 to 2023. Here's an analysis:

The Analysis:

Department Of Computer Science And Engineering:
2020: 1 article, 2021: 4 articles, 2022: 7 articles, 2023: 8 articles.

King Abdulaziz University:
2020: 2 articles, 2021: 2 articles, 2022: 6 articles, 2023: 6 articles

King Saud University:
2020: 3 articles, 2021: 4 articles, 2022: 7 articles, 2023: 7 articles.

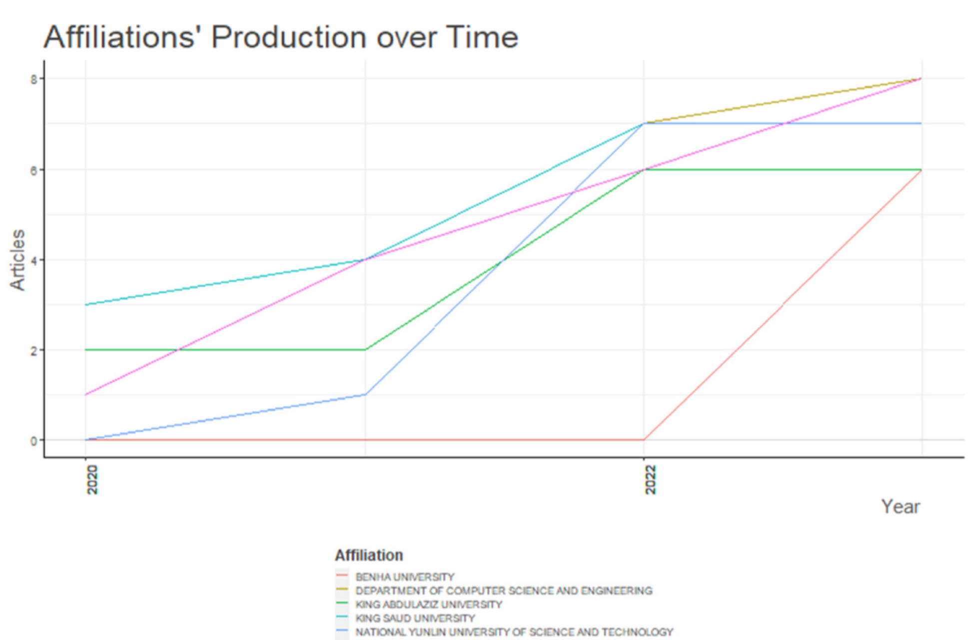

Fig. 7. Affiliations' production overtime.

SRM Institute Of Science And Technology:
2020: 1 article, 2021: 4 articles, 2022: 6 articles, 2023: 8 articles.

National Yunlin University Of Science And Technology:
2020: 0 articles, 2021: 1 article, 2022: 7 articles, 2023: 7 articles.

Benha University:
2020: 0 articles, 2021: 0 articles, 2022: 0 articles, 2023: 6 articles.

Observations:

- The number of articles generally increased for most affiliations over the years.
- "National Yunlin University Of Science And Technology" and "Benha University" started publishing articles from 2021 onward.
- "Department Of Computer Science And Engineering" consistently increased its publications each year.
- "King Saud University" consistently published a significant number of articles each year.
- "Benha University" started publishing articles in 2023, with a relatively high number (6) in that year.

11. Corresponding author's countries

The Fig. 8 and Table 8 provide information on the number of articles, SCP (Source Code Paper) counts, MCP (Methodology/Case Study Paper) counts, and corresponding frequencies and ratios for different countries.

The Analysis:

India: Articles: 102, SCP: 75, MCP: 27, MCP Ratio: 0.265 (Approximately 26.5%).

China: Articles: 33, SCP: 22, MCP: 11, MCP Ratio: 0.333 (Approximately 33.3%).

USA: Articles: 14, SCP: 8, MCP: 6, MCP Ratio: 0.429 (Approximately 42.9%).

United Kingdom: Articles: 12, SCP: 4, MCP: 8, MCP Ratio: 0.667 (Approximately 66.7%).

Italy: Articles: 8, SCP: 8, MCP: 0, MCP Ratio: 0.021 (Approximately 2.1%).

Korea: Articles: 7, SCP: 3, MCP: 4, MCP Ratio: 0.571 (Approximately 57.1%).

Saudi Arabia: Articles: 7, SCP: 4, MCP: 3, MCP Ratio: 0.429 (Approximately 42.9%).

Australia: Articles: 5, SCP: 1, MCP: 4, MCP Ratio: 0.8 (Approximately 80%).

Malaysia: Articles: 5, SCP: 2, MCP: 3, MCP Ratio: 0.6 (Approximately 60%).

Morocco: Articles: 5, SCP: 4, MCP: 1, MCP Ratio: 0.2 (Approximately 20%).

Observations:

- India and China have a high number of articles, with China having a higher MCP ratio.
- The USA has a significant number of articles, and a relatively high MCP ratio.

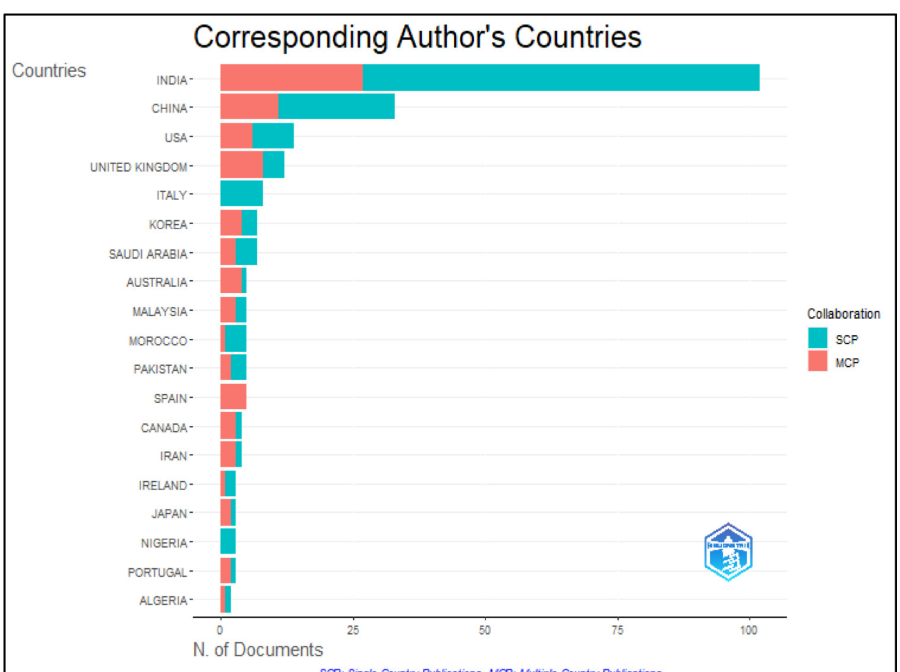

Fig. 8. Corresponding author's countries.

Table 8. Corresponding author's countries.

Country	Articles	SCP	MCP	Freq	MCP Rating
India	102	75	27	0.27055703	0.26470588
	99	70	29	0.26259947	0.29292929
China	33	22	11	0.08753316	0.33333333
USA	14	8	6	0.03713528	0.42857143
United Kingdom	12	4	8	0.03183024	0.66666667
Italy	8	8	0	0.02122016	0
Korea	7	3	4	0.01856764	0.57142857
Saudi Arabia	7	4	3	0.01856764	0.42857143
Australia	5	1	4	0.0132626	0.8
Malaysia	5	2	3	0.0132626	0.6
Morocco	5	4	1	0.0132626	0.2
Pakistan	5	3	2	0.0132626	0.4
Spain	5	0	5	0.0132626	1
Canada	4	1	3	0.01061008	0.75
Iran	4	1	3	0.01061008	0.75
Ireland	3	2	1	0.00795756	0.33333333
Japan	3	1	2	0.00795756	0.66666667
Nigeria	3	3	0	0.00795756	0
Portugal	3	1	2	0.00795756	0.66666667
Algeria	2	1	1	0.00530504	0.5
Bahrain	2	0	2	0.00530504	1
Bangladesh	2	1	1	0.00530504	0.5
Czech Republic	2	1	1	0.00530504	0.5
Egypt	2	1	1	0.00530504	0.5
France	2	1	1	0.00530504	0.5
Germany	2	2	0	0.00530504	0
Greece	2	2	0	0.00530504	0
Iraq	2	2	0	0.00530504	0
Macedonia	2	0	2	0.00530504	1
Netherlands	2	2	0	0.00530504	0
Poland	2	1	1	0.00530504	0.5
Sri Lanka	2	1	1	0.00530504	0.5
Switzerland	2	2	0	0.00530504	0
United Arab Emirates	2	2	0	0.00530504	0
Austria	1	0	1	0.00265252	1
Belgium	1	1	0	0.00265252	0
Bosnia	1	1	0	0.00265252	0
Brazil	1	1	0	0.00265252	0
Cyprus	1	0	1	0.00265252	1
Denmark	1	1	0	0.00265252	0
Ecuador	1	1	0	0.00265252	0
Guinea	1	0	1	0.00265252	1
Hong Kong	1	0	1	0.00265252	1.
Hungary	1	0	1	0.00265252	1
Jordan	1	0	1	0.00265252	1
Kenya	1	0	1	0.00265252	1
Kyrgyzstan	1	1	0	0.00265252	0
Malawi	1	0	1	0.00265252	1
Mauritius	1	1	0	0.00265252	0
Norway	1	1	0	0.00265252	0
Qatar	1	0	1	0.00265252	1
Serbia	1	0	1	0.00265252	1
Singapore	1	0	1	0.00265252	1
Tunisia	1	0	1	0.00265252	1

- The United Kingdom has a lower number of articles but a high MCP ratio, indicating a focus on methodology/case study papers.
- Italy has a lower overall contribution, with a negligible MCP ratio.
- Some countries, like Korea and Saudi Arabia, have a balanced distribution between SCP and MCP.
- Australia and Malaysia have a high MCP ratio, suggesting a focus on methodology/case study papers.
- Morocco has a higher SCP ratio compared to MCP.

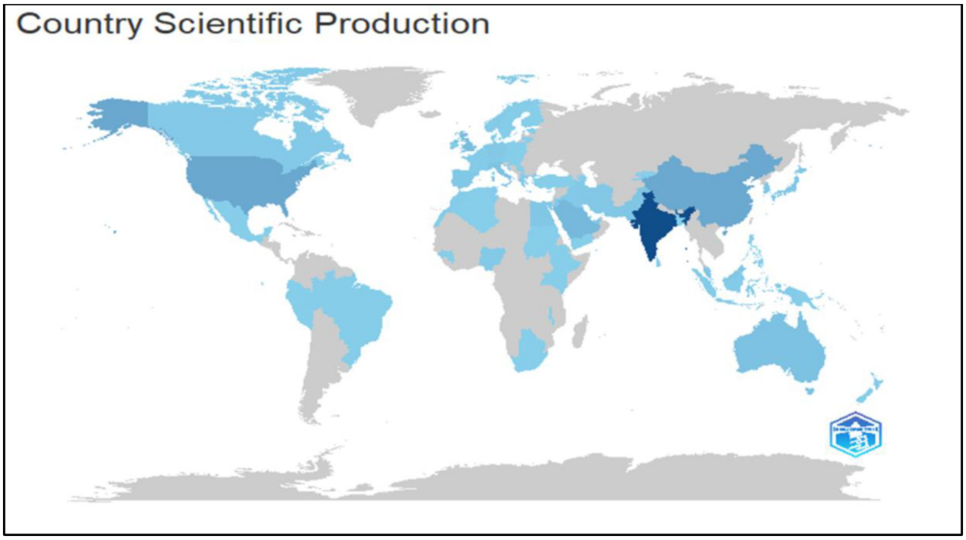

Fig. 9. Country scientific production.

12. Country scientific production

Fig. 9 and Table 10 show the analyze scientific production by country.

The Table 9 provides a frequency count of corresponding authors' countries.

The Analysis:

Top Contributor Countries:

India (INDIA): 332, **USA:** 93, **China:** 91

Significant Contributors:

Saudi Arabia, UK, Australia, Italy, Malaysia, Pakistan, South Korea, United Arab Emirates, Egypt, Japan, Spain, Germany, Greece, Iran, Portugal, Canada, Iraq, Netherlands, Nigeria, Switzerland, France.

Moderate Contributors:

Jordan, Ireland, Norway, Czech Republic, Indonesia, Morocco, Belgium, Sri Lanka, Tunisia, Brazil, Singapore.

Minor Contributors:

Austria, Bahrain, New Zealand, Qatar, Sweden, Algeria, Bangladesh, Croatia, Denmark, Ecuador, Finland, Guinea, Hungary, North Macedonia, Papua New Guinea, Poland, Slovakia, South Africa, Turkey.

Countries with a Single Contributor:

Azerbaijan, Botswana, Cyprus, Estonia, Ethiopia, Kenya, Kuwait, Kyrgyzstan, Lithuania, Malawi, Mauritius, Mexico, Peru, Philippines, Serbia, Sudan, Uganda, Yemen.

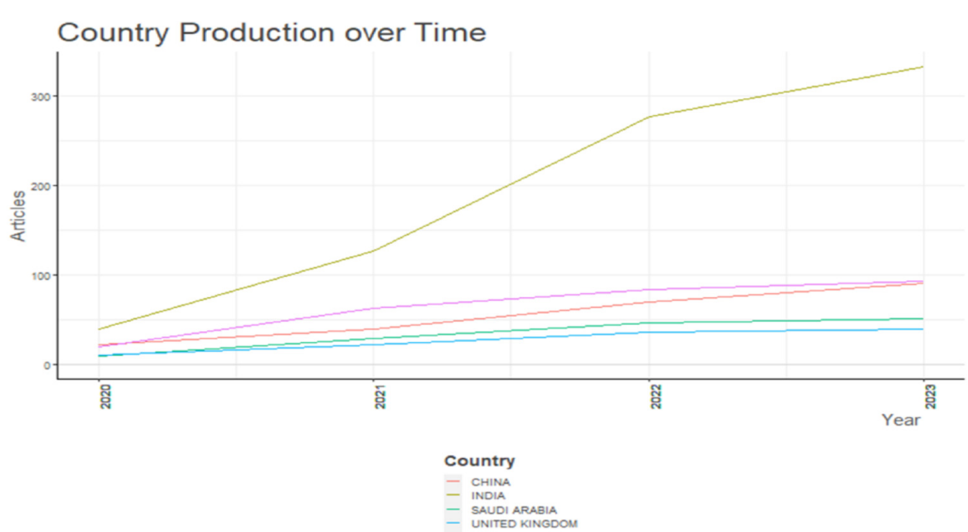

Fig. 10. Country production over time.

Table 9. Country scientific production.

Region	Frequency
India	332
USA	93
China	91
Saudi Arabia	51
UK	39
Australia	29
Italy	28
Malaysia	21
Pakistan	21
South Korea	18
United Arab Emirates	18
Egypt	17
Japan	16
Spain	16
Germany	15
Greece	13
Iran	13
Portugal	12
Canada	11
Iraq	10
Netherlands	10
Nigeria	10
Switzerland	10
France	9
Jordan	8
Ireland	7
Norway	7
Czech Republic	6
Indonesia	6
Morocco	6
Belgium	5
Sri Lanka	5
Tunisia	5
Brazil	4
Singapore	4
Austria	3
Bahrain	3
New Zealand	3
Qatar	3
Sweden	3
Algeria	2
Bangladesh	2
Croatia	2
Denmark	2
Ecuador	2
Finland	2
Guinea	2
Hungary	2
North Macedonia	2
Papua New Guinea	2
Poland	2
Slovakia	2
South Africa	2
Turkey	2
Azerbaijan	1
Botswana	1
Cyprus	1
Estonia	1
Ethiopia	1
Kenya	1
Kuwait	1
Kyrgyzstan	1

(*continued on next column*)

Table 9. Continued.

Region	Frequency
Lithuania	1
Malawi	1
Mauritius	1
Mexico	1
Peru	1
Philippines	1
Serbia	1
Sudan	1
Uganda	1
Yemen	1

Observations:

The majority of contributions come from India, followed by the USA and China. There's a diverse range of countries represented, indicating a global distribution of corresponding authors. Some countries have a higher concentration of contributors, reflecting a strong presence in the academic community.

13. Country production over time

The Table 10 and Fig. 10 provide the number of articles published in different years for specific countries.

The Analysis:

1. **China:**
   2020: 22 articles, 2021: 40 articles, 2022: 70 articles, 2023: 91 articles. China shows a consistent increase in the number of articles over the years.

2. **India:**
   2020: 39 articles, 2021: 127 articles, 2022: 277 articles, 2023: 332 articles. India also

Table 10. Country production over time.

Country	Year	Articles
China	2020	22
China	2021	40
China	2022	70
China	2023	91
India	2020	39
India	2021	127
India	2022	277
India	2023	332
Saudi Arabia	2020	9
Saudi Arabia	2021	29
Saudi Arabia	2022	47
Saudi Arabia	2023	51
United Kingdom	2020	11
United Kingdom	2021	22
United Kingdom	2022	36
United Kingdom	2023	39
USA	2020	20
USA	2021	63
USA	2022	84
USA	2023	93

exhibits a substantial increase in article publication over the years.

3. **Saudi Arabia:**
   2020: 9 articles, 2021: 29 articles, 2022: 47 articles, 2023: 51 articles. Saudi Arabia demonstrates a growth in article production, with the highest number in 2023.
4. **United Kingdom:**
   2020: 11 articles, 2021: 22 articles, 2022: 36 articles, 2023: 39 articles. The United Kingdom also shows a gradual increase in the number of published articles.
5. **USA:**
   2020: 20 articles, 2021: 63 articles, 2022: 84 articles, 2023: 93 articles. The USA exhibits a consistent growth pattern in article production over the years.

Overall Trends:

- All countries (China, India, Saudi Arabia, the United Kingdom, and the USA) show an upward trend in the number of published articles from 2020 to 2023.
- China and India contribute the highest number of articles, with India surpassing other countries in terms of the sheer volume of publications.

This analysis provides a snapshot of the publication trends for the specified countries over the given years. Further insights could be derived by considering additional factors or specific research areas.

14. Most citied countries

The Table 11 provides information about the total number of articles (TC) and the average number of citations per article for various countries [21].

The analysis of the data:

1. **Top Performers:**
   Australia: With a total of 454 articles, Australia has the highest average citations per article (90.8). This indicates a high impact and recognition of Australian research in the academic community.
   Canada: Similar to Australia, Canada also shows a high average of 113.5 citations per article, with a total of 454 articles.
2. **High Citations:**
   China: While China has the highest total number of articles (595), it also maintains a high average of 18 citations per article, indicating a significant impact of Chinese research.
   USA: The USA, with 360 articles, has an average of 25.7 citations per article, reflecting the strong influence of American research.
3. **Balanced Impact:**
   Tunisia: Despite having a lower total number of articles (144), Tunisia has a very high average

Table 11. Most citied countries.

Country	TC	Average article citations
India	651	6.4
China	595	18
Australia	454	90.8
Canada	454	113.5
USA	360	25.7
Iran	260	65
United Kingdom	181	15.1
Korea	177	25.3
Malaysia	145	29
Tunisia	144	144
Italy	92	11.5
Saudi Arabia	90	12.9
Nigeria	79	26.3
Japan	60	20
Qatar	36	36
Jordan	35	35
Spain	35	7
Singapore	34	34
Guinea	26	26
Brazil	25	25
Switzerland	24	12
Portugal	23	7.7
Hungary	18	18
Greece	16	8
Ireland	13	4.3
Malawi	13	13
Algeria	11	5.5
Austria	11	11
Czech Republic	11	5.5
Netherlands	11	5.5
Egypt	9	4.5
Norway	8	8
Bosnia	7	7
Pakistan	7	1.4
Poland	6	3
Kyrgyzstan	5	5
Hong Kong	4	4
Bangladesh	3	1.5
France	3	1.5
Iraq	3	1.5
Cyprus	2	2
Mauritius	2	2
Serbia	2	2
Belgium	1	1
Germany	1	0.5
Macedonia	1	0.5
Sri Lanka	1	0.5
Bahrain	0	0
Denmark	0	0
Ecuador	0	0
Kenya	0	0
Morocco	0	0
United Arab Emirates	0	0

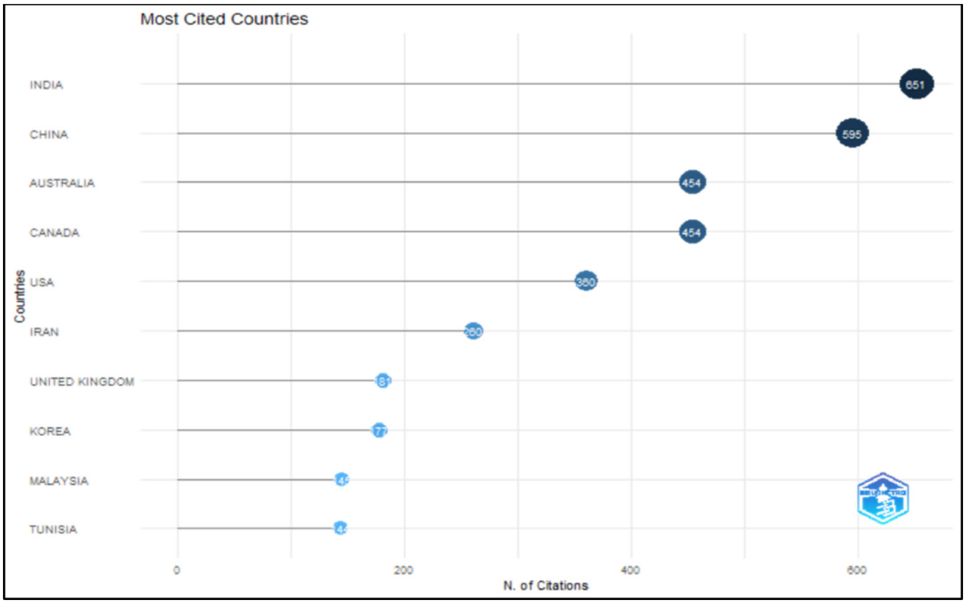

Fig. 11. Most citied countries.

of 144 citations per article, suggesting a concentrated impact in specific research areas.

Qatar, Jordan, Singapore, Guinea, Brazil: These countries also show a balanced impact with the average number of citations equal to the total number of articles.

4. **Moderate Impact:**

   United Kingdom, Iran, Korea, Malaysia: These countries demonstrate a moderate impact with average citations ranging from 15.1 to 29.

5. **Low Impact:**

   Several countries, including Germany, Macedonia, Sri Lanka, Bahrain, Denmark, Ecuador, Kenya, Morocco, and the United Arab Emirates, have either zero or very low average citations per article. This suggests that research from these countries may have a lower visibility or impact in terms of citations.

Overall Trends:

The number of citations per article is a key indicator of research impact and influence. Some countries excel in achieving a high average impact despite a lower total number of articles. It's essential to consider the research field, publication venue, and citation practices when interpreting these numbers.

This analysis provides insights into the relative impact of research from different countries based on citation metrics.

15. Words' frequency over time

The Table 12 and Fig. 12 provide data on the growth or adoption of various technology-related trends across different years. An analyze the trends in each category [22–24]:

1. **Big Data:**

   The adoption of Big Data has steadily increased over the years, with significant growth observed from 2020 to 2023.

2. **Health Care:**

   The Health Care sector has shown consistent growth, and it has nearly doubled from 2020 to 2023. This indicates an increasing focus on technology in the healthcare industry.

3. **Internet of Things (IoT):**

   IoT has experienced substantial growth, more than doubling from 2020 to 2023. This suggests a broader integration of IoT devices in various sectors.

Table 12. Words' frequency over time.

Year	Big data	Health care	Internet of things (IoT)	Data analytics	Blockchain	Artificial intelligence (AI)	Data privacy	Network security	Human privacy	Privacy
2020	50	30	19	16	12	9	15	12	17	16
2021	110	80	51	37	33	28	29	28	30	25
2022	176	137	86	64	56	48	42	44	43	39
2023	216	162	106	82	71	54	52	51	50	47

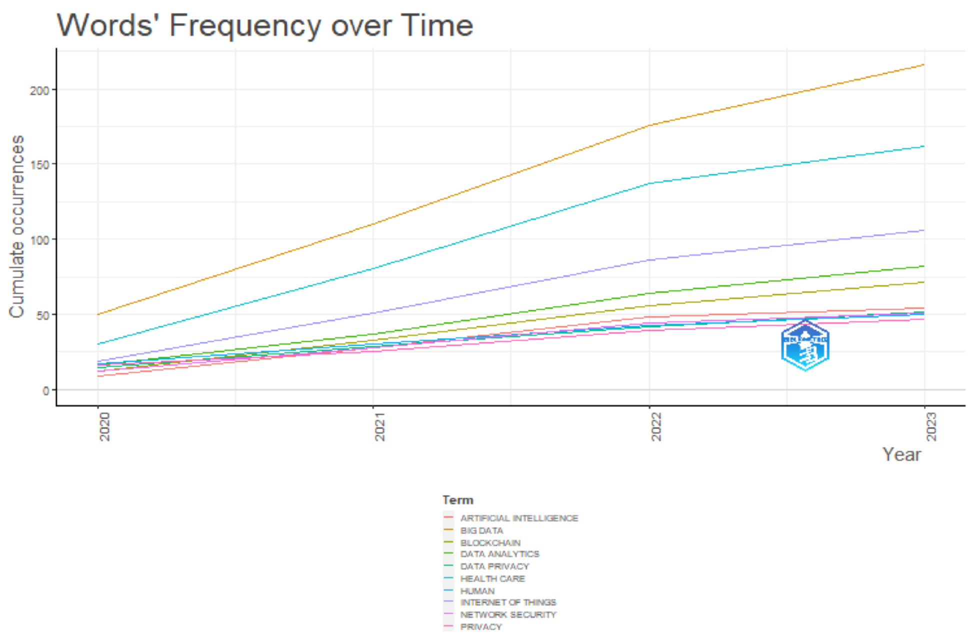

Fig. 12. Words' frequency over time.

4. **Data Analytics:**
   Data Analytics has also shown significant growth, more than doubling from 2020 to 2023. This reflects the increasing importance of extracting insights from large datasets.
5. **Blockchain:**
   Blockchain has shown steady growth, although the rate is comparatively slower than some other trends.
6. **Artificial Intelligence (AI):**
   AI has experienced consistent and substantial growth, more than doubling from 2020 to 2023. This aligns with the increasing integration of AI technologies across various domains.
7. **Data Privacy:**
   Data Privacy has seen a notable increase, especially in the earlier years. This suggests a growing awareness and emphasis on protecting user data.
8. **Network Security:**
   Network Security has shown steady growth, reflecting the ongoing importance of securing digital networks.
9. **Human Privacy:**
   The category "Human Privacy" seems to represent a more general concept, and it has shown an increase, reflecting a growing concern for preserving individual privacy.

In summary, the overall trend across these technology domains is one of growth and increasing adoption. This data indicates a continued and expanding reliance on technologies such as Big Data, IoT, AI, and Data Analytics, with a simultaneous focus on addressing privacy and security concerns.

16. Trend topics

The Table 13 provides information on the frequency of certain items in different years, along with the year's corresponding to the first quartile (year_q1), median (year_med), and third quartile (year_q3) of the data. Let's analyze the trends for each item [24–27]:

1. **Health-care system:**
   Appears 15 times in 2020, 2020, and 2021. Indicates a consistent presence throughout the analyzed period.

Table 13. Trend topics over years.

Item	Freq	Year_q1	Year_med	Year_q3
Health-care system	15	2020	2020	2021
Europe	12	2020	2020	2020
Privacy preserving	12	2020	2020	2022
Big data	216	2021	2021	2022
Artificial intelligence	54	2021	2021	2022
Data privacy	52	2020	2021	2022
Health care	162	2021	2022	2022
Internet of things	106	2021	2022	2022
Data analytics	82	2021	2022	2022
Healthcare	20	2022	2023	2023
Edge computing	15	2022	2023	2023
Decentralized	7	2022	2023	2023

2. **Europe:**
   Appears 12 times in 2020. Indicates a focus on Europe specifically in the early part of the analyzed period.

3. **Privacy preserving:**
   Appears 12 times in 2020, 2020, and 2022. Suggests a consistent but sporadic mention with a gap in 2021.

4. **Big Data:**
   Appears 216 times in 2021 and 2021, and 2022. Indicates a significant emphasis on Big Data throughout the analyzed period.

5. **Artificial Intelligence:**
   Appears 54 times in 2021, 2021, and 2022. Indicates a substantial focus on Artificial Intelligence.

6. **Data Privacy:**
   Appears 52 times in 2020, 2021, and 2022. Indicates a consistent focus on data privacy.

7. **Health care:**
   Appears 162 times in 2021, 2022, and 2022. Indicates a strong and consistent emphasis on health care.

8. **Internet of Things (IoT):**
   Appears 106 times in 2021, 2022, and 2022. Indicates a notable focus on IoT.

9. **Data Analytics:**
   Appears 82 times in 2021, 2022, and 2022. Indicates a substantial emphasis on data analytics.

10. **Healthcare:**
    Appears 20 times in 2022 and 2023. Indicates a focus on healthcare specifically in later years.

11. **Edge Computing:**
    Appears 15 times in 2022 and 2023. Indicates a mention of edge computing in the later years.

12. **Decentralised:**
    Appears 7 times in 2022 and 2023. Indicates a mention of decentralization in the later years.

In Fig. 13 the analysis shows consistent and increasing mentions of key terms like Big Data, Artificial Intelligence, Data Privacy, Health Care, and Internet of Things throughout the analyzed period. Some terms, like Europe and Privacy Preserving, have sporadic mentions, while others, like Healthcare,

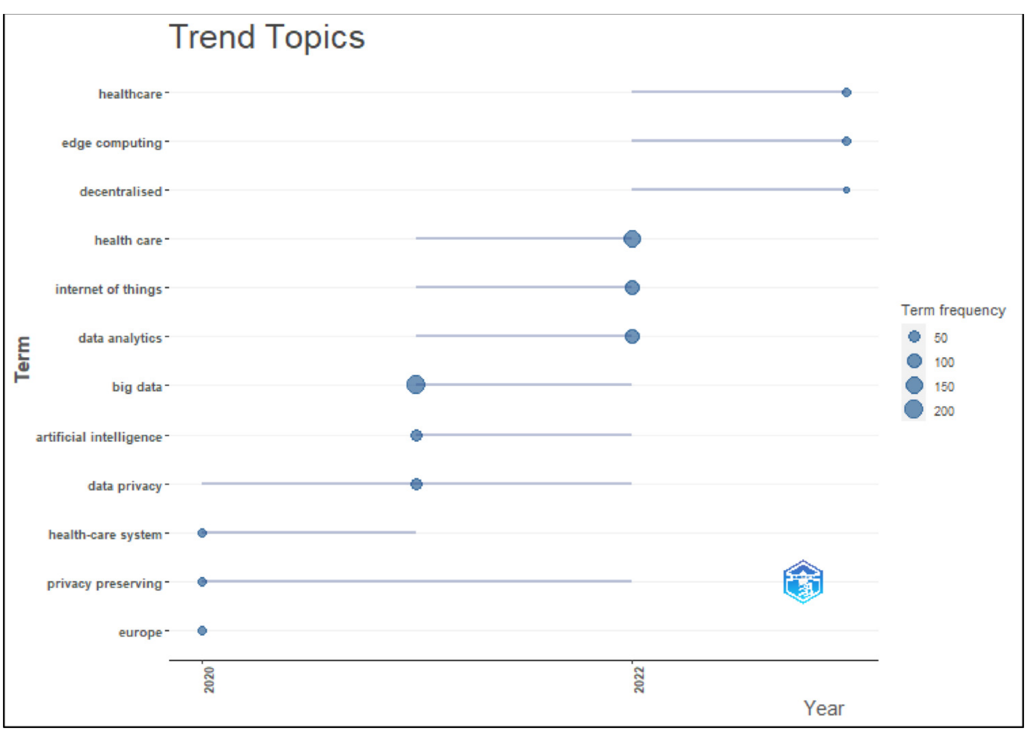

Fig. 13. Trend topics over years.

Table 14. CoWordNet analysis.

Node	Cluster	Betweenness	Closeness	PageRank
Artificial intelligence	1	6.69208363	0.01724138	0.0241147
Human	1	7.87134645	0.01639344	0.03069521
Privacy	1	5.09079656	0.01639344	0.0242223
Computer security	1	0.78067616	0.01388889	0.01535003
Humans	1	3.06950939	0.01470588	0.02339336
Cloud computing	1	2.79928869	0.01470588	0.01763065
Article	1	0.68212035	0.01315789	0.01494404
Health care delivery	1	0.86419889	0.01351351	0.01677028
Delivery of health care	1	0.78498061	0.01351351	0.01609219
Review	1	0.24641312	0.01282051	0.01140195
Security and privacy	1	0.60280744	0.01408451	0.0113833
Big data	2	160.150413	0.02040816	0.09522055
Health care	2	113.662333	0.02083333	0.0770129
Internet of things	2	52.4958465	0.02083333	0.05666181
Data analytics	2	17.6971046	0.02040816	0.03612567
Blockchain	2	22.2092207	0.02	0.03987742
Data privacy	2	10.5171622	0.01923077	0.02949849
Network security	2	8.79058261	0.01851852	0.02622946
Digital storage	2	6.77447667	0.01785714	0.02496782
Block-chain	2	7.22548936	0.01724138	0.02904794
Information management	2	6.68845098	0.01818182	0.02693273
Security	2	4.66754198	0.01694915	0.02521801
Deep learning	2	1.38847784	0.01428571	0.01532983
Machine learning	2	2.9585825	0.01587302	0.01861203
Diagnosis	2	2.85324521	0.01639344	0.01935309
Cryptography	2	1.989839	0.01470588	0.01772954
Learning systems	2	0.28168294	0.01315789	0.01219632
Cloud-computing	2	1.3255719	0.01470588	0.015523
Decision making	2	1.1937447	0.01449275	0.0123382
Medical computing	2	1.29031344	0.01449275	0.01492074
Data mining	2	0.59543813	0.01408451	0.01127181
Healthcare industry	2	0.40875756	0.01369863	0.01225839
Machine-learning	2	1.71693732	0.01470588	0.01444274
Healthcare	2	0.25562964	0.01282051	0.01096179
Security of data	2	0.79562555	0.01333333	0.01094832
Access control	2	0.27807845	0.01351351	0.01104375
Advanced analytics	2	1.46427419	0.015625	0.01409895
Healthcare services	2	0.77505945	0.01515152	0.01153275
Data sharing	2	0.34270088	0.01333333	0.01014808
Smart city	2	0.05470586	0.01234568	0.00813362
Fog computing	2	0.78817707	0.01408451	0.01154245
Healthcare systems	2	0.14023677	0.01298701	0.00913492
Internet of things (iot)	2	0.58757395	0.01351351	0.01042759
Privacy by design	2	0.72930514	0.01388889	0.00983394
Sensitive data	2	0.09958795	0.01234568	0.0074395
Edge computing	2	0.08321613	0.01282051	0.00833199
Health-care system	2	0.62008456	0.01351351	0.01007015
Hospitals	2	0.34695696	0.01388889	0.00940409
Authentication	2	0.27335416	0.01369863	0.0101816

Edge Computing, and Decentralized, become more prominent in the later years.

17. CoWordNet (Node Degrees) analysis

CoWordNet is a linguistic resource that maps the co-occurrence of words in a specific context, often used in natural language processing and semantic analysis [27–29]. The Table 14 provides information about the node, Cluster, Betweenness, Closeness and PageRank.

The Fig. 14 appears to contain information related to network analysis metrics for nodes in two clusters, with various attributes such as betweenness, closeness, and PageRank. Analyze the information as follow:

Cluster 1:

- **Nodes:**
  artificial intelligence, human, privacy, computer security, humans, cloud computing, article,

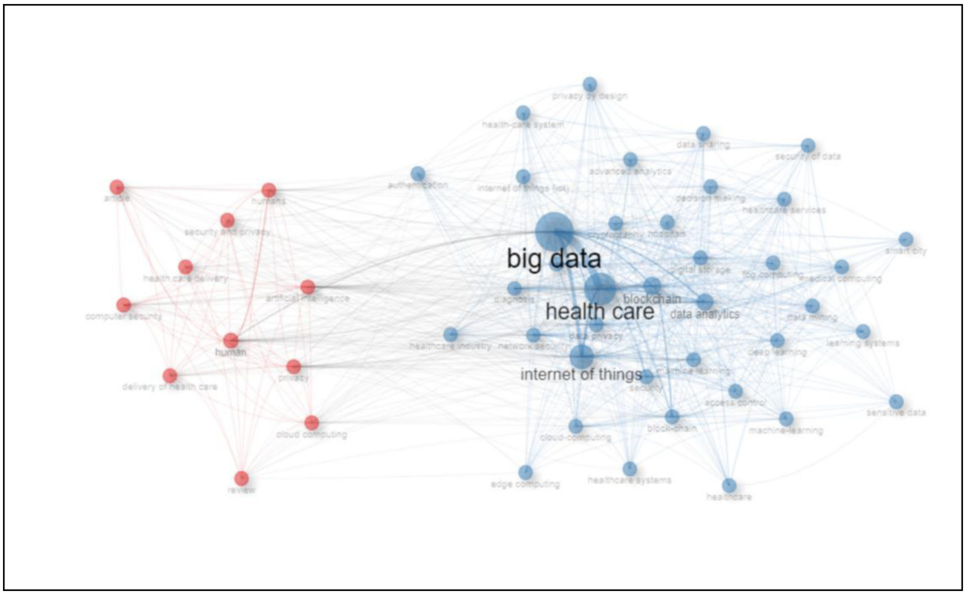

Fig. 14. CoWordNet information.

health care delivery, delivery of health care, review, security and privacy.

- **Observations:**
  These nodes have varying values for betweenness, closeness, and PageRank. Betweenness values suggest the importance of a node in connecting different parts of the network. Closeness values indicate how close a node is to other nodes. PageRank values give an indication of the importance of a node based on its connections.

Cluster 2:

- **Nodes:**
  big data, health care, internet of things, data analytics, blockchain, data privacy, network security, digital storage, block-chain, information management, security, deep learning, machine learning, diagnosis, cryptography, learning systems, cloud-computing, decision making, medical computing, data mining, healthcare industry,

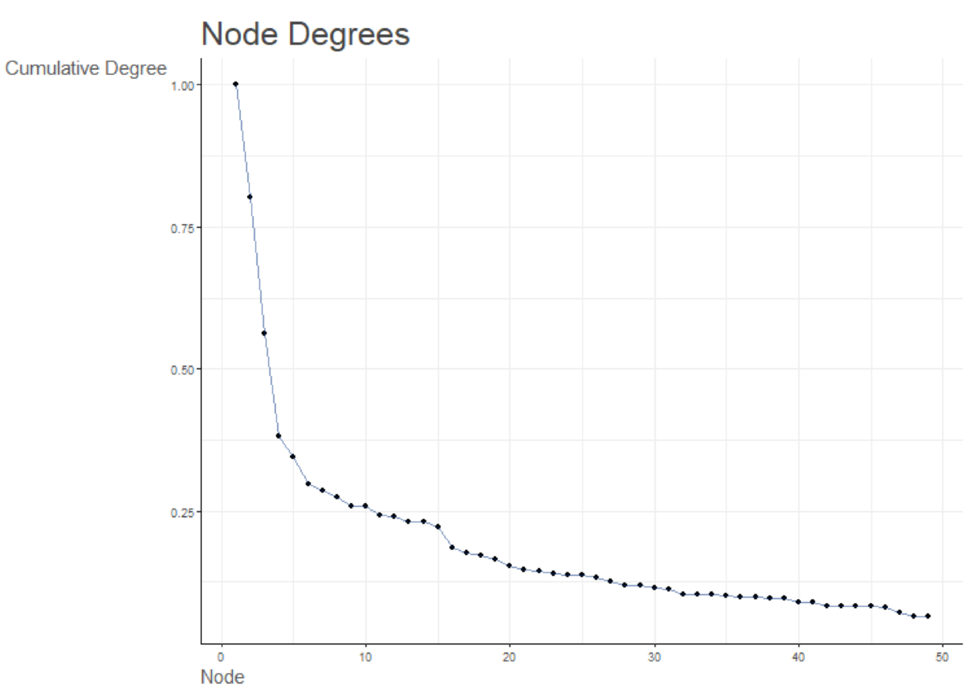

Fig. 15. CoWordNet information.

machine-learning, healthcare, security of data, access control, advanced analytics, healthcare services, data sharing, smart city, fog computing, healthcare systems, internet of things (IOT), privacy by design, sensitive data, edge computing, health-care system, hospitals, authentication.

- **Observations:**
  Nodes in this cluster also have varying values for betweenness, closeness, and PageRank. Big data, health care, and internet of things appear to have high betweenness, indicating their importance in connecting different parts of the network. Closeness values show how well-connected nodes are to others. PageRank values suggest the importance of nodes based on their connections.

Overall Analysis:
Both clusters contain nodes related to various technology and healthcare-related terms. Nodes like artificial intelligence, human, privacy, big data, health care, and internet of things appear to be central in their respective clusters. The between ness, closeness, and PageRank metrics provide insights into the network structure and the importance of specific nodes in facilitating communication or information flow.

18. Tree map analysis

Tree map analysis is a visual representation technique used to display hierarchical data using nested rectangles. Each rectangle represents a branch of the hierarchy, with its size proportional to a specific metric, such as quantity or value. Table 15 and Fig. 16 show the information terms with frequency.

The analysis of the frequency table:

Dominant Themes:
The most frequently occurring terms include “big data,” “health care,” “internet of things,” and “data analytics,” suggesting that these are major themes within the dataset.

Technology and Innovation:
Terms like “blockchain,” “artificial intelligence,” “machine learning,” “cloud computing,” and “data mining” indicate a focus on technological advancements and data-driven approaches.

Security and Privacy:
The presence of terms like “network security,” “data privacy,” “cryptography,” “security,” and “security and privacy” emphasizes the importance of securing data and maintaining privacy.

Human-Centric Focus:
The terms “human” and “humans” suggest a consideration of human factors, while “privacy by design” underscores the significance of privacy in the design of systems.

Table 15. Tree map analysis.

Terms	Frequency
Big data	216
Health care	162
Internet of things	106
Data analytics	82
Block chain	71
Artificial intelligence	54
Data privacy	52
Network security	51
Human	50
Privacy	47
Digital storage	46
Block-chain	45
Information management	45
Security	41
Deep learning	40
Machine learning	37
Computer security	36
Diagnosis	35
Humans	34
Cloud computing	33
Cryptography	31
Article	24
Learning systems	24
Cloud-computing	23
Decision making	22
Health care delivery	22
Medical computing	22
Data mining	21
Healthcare industry	21
Machine-learning	21
Delivery of health care	20
Healthcare	20
Security of data	20
Access control	19
Advanced analytics	19
Healthcare services	18
Data sharing	17
Review	17
Smart city	17
Fog computing	16
Healthcare systems	16
Internet of things (iot)	16
Privacy by design	16
Security and privacy	16
Sensitive data	16
Edge computing	15
Health-care system	15
Hospitals	15
Authentication	14
Health care application	14

Healthcare Emphasis:
The terms “health care,” “healthcare industry,” “healthcare services,” and “healthcare systems” highlight a strong emphasis on healthcare-related topics, including delivery and applications.

Data Management:
Terms like “digital storage,” “information management,” “data sharing,” and “sensitive data” indicate a concern for effective data management and sharing practices.

Fig. 16. Tree map analysis.

Learning Systems:

The terms “deep learning,” “machine learning,” “learning systems,” and “advanced analytics” point towards a focus on machine learning algorithms and advanced analytical methods.

IoT and Connectivity:

“Internet of things,” “smart city,” “fog computing,” and “edge computing” suggest an interest in IoT, smart city technologies, and edge/fog computing.

Healthcare Delivery and Application:

Terms like “health care delivery,” “delivery of health care,” “health care application,” and “hospitals” indicate a focus on improving healthcare delivery and applications.

Variations and Standardization:

Some terms have variations in spelling or formatting (“blockchain” vs. “block-chain,” “cloud computing” vs. “cloud-computing”). Standardizing these variations may enhance consistency in analysis.

Decision Making:

“Decision making” is mentioned, suggesting a consideration of decision support systems or decision-making processes.

Frequency Distribution:

The frequency distribution provides insights into the relative importance of different concepts within the dataset.

Fig. 17. WordCloud analysis.

Potential Research Areas:

The terms could represent potential research areas or topics of interest, with their frequency reflecting the focus of the dataset.

In summary, the Table 16 indicates a multidimensional dataset covering various aspects, including technology, healthcare, security, and human-centered considerations. Further exploration or context about the dataset would be necessary for a more in-depth analysis and interpretation.

Table 16. WordCloud analysis.

Terms	Frequency
Big data	216
Health care	162
Internet of things	106
Data analytics	82
Blockchain	71
Artificial intelligence	54
Data privacy	52
Network security	51
Human	50
Privacy	47
Digital storage	46
Block-chain	45
Information management	45
Security	41
Deep learning	40
Machine learning	37
Computer security	36
Diagnosis	35
Humans	34
Cloud computing	33
Cryptography	31
Article	24
Learning systems	24
Cloud-computing	23
Decision making	22
Health care delivery	22
Medical computing	22
Data mining	21
Healthcare industry	21
Machine-learning	21
Delivery of health care	20
Healthcare	20
Security of data	20
Access control	19
Advanced analytics	19
Healthcare services	18
Data sharing	17
Review	17
Smart city	17
Fog computing	16
Healthcare systems	16
Internet of things (iot)	16
Privacy by design	16
Security and privacy	16
Sensitive data	16
Edge computing	15
Health-care system	15
Hospitals	15
Authentication	14
Health care application	14

19. WordCloud analysis

Word cloud analysis is a visual representation of text data, where words are displayed in varying sizes based on their frequency or importance in a given dataset as shown in Table 16. The more frequently a word appears, the larger it is illustrated in the cloud in Fig. 17.

The provided table lists terms and their respective frequencies. Here's an analysis:
big data: Frequency: 216.
health care: Frequency: 162.
internet of things: Frequency: 106.
data analytics: Frequency: 82.
blockchain: Frequency: 71.
artificial intelligence: Frequency: 54.
data privacy: Frequency: 52.
network security: Frequency: 51.
human: Frequency: 50.
privacy: Frequency: 47.

Observations:

As shown in Fig. 17, the "Big data" is the most frequently mentioned term, indicating a significant focus on large-scale data in the context of the articles. "Health care" and "internet of things" also have high frequencies, suggesting a substantial emphasis on these topics. Terms like "blockchain," "data analytics," "artificial intelligence," and "network security" reflect a focus on cutting-edge technologies and methods. Some terms, such as "block-chain" and "cloud-computing," appear to have variations in spelling, which might need standardization for consistency in analysis.

20. Conclusion and future work

This bibliometric analysis has illuminated the critical intersection of secure healthcare systems and big data, revealing significant trends and influential contributors in the field. The findings underscore the growing importance of data security in healthcare as organizations increasingly leverage big data technologies to enhance patient care and operational efficiency. Notably, the analysis highlights the need for ongoing research to address emerging security challenges, such as data breaches and compliance with regulatory frameworks. As healthcare systems continue to evolve in the digital age, the integration of robust security measures will be essential to maintain trust and protect sensitive information.

Future research should focus on several key areas. First, a deeper exploration of specific case studies showcasing successful implementations of secure big data frameworks in healthcare settings will provide practical insights. Second, further investigation into the evolving regulatory landscape and its impact on data security practices in healthcare is crucial. Additionally, interdisciplinary collaborations among healthcare professionals, data scientists, and cybersecurity experts can foster innovative solutions to emerging threats.

Acknowledgement

Not provided.

Funding

Not provided.

References

1. S. Gunalan B. Pharm., "Big data for secure healthcare system: a review," AMET University. Guest Faculty, BBA Department, D. G. Vaishnav College, Chennai, 2021.
2. Sangjukta Das and Suyel Namasudra, "A Lightweight and Anonymous Mutual Authentication Scheme for Medical Big Data in Distributed Smart Healthcare Systems," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 21, no. 4, Dec. 2022.
3. Inam Ullah, Rehmat Ullah, Rehmat, Ur Ateeq, and Ateeq Ur Rehman, "Big Data Analytics and Artificial Intelligence in Electronics," *Journal of Electronics*, June 2023.
4. Mohit Kumar, Hritu Raj, and Nisha Chaurasia, "Blockchain inspired secure and reliable data exchange architecture for cyber-physical healthcare system 4.0," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 309–322, Jun 2023.
5. A. Angel Cerli, Kalaiselvi K, and Vijayakumar Varadarajan, "Secure Healthcare Systems Based on Big Data Analytics," In book: *Handbook of Intelligent Healthcare Analytics*, May 2022.
6. Vivek Sharma, "A Survey: Secure Indexing & Storage of Big Data," Conference Proceedings of ICDLAIR, February 2021.
7. Tahsien Al-Quraishi, Naseer Al-Quraishi, Hussein AlNabulsi, Hussein ALQarishey, and Ahmed Hussein Ali, "Big Data Predictive Analytics for Personalized Medicine: Perspectives and Challenges," *Applied Data Science and Analysis*, Apr 11, 2024.
8. K. Rajeshkumar, Dhanasekaran Subbiah, and Vinod Vasudevan, "Efficient and secure medical big data management system using optimal map-reduce framework and deep learning," *Journal of Multimedia Tools and Applications*, October 2023.
9. Ismail Ameskour, Houssam Benbrahim, and Aouatif Amine, "Health Data Security in a Big Data Environment," In book: *Artificial Intelligence and Green Computing*, pp. 215–231, November 2023.
10. Mahmood A. Bazel, Fathey Mohammed, and Mazida Ahmad, "Blockchain Technology in Healthcare Big Data Management: Benefits, Applications and Challenges," 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), August 2021.
11. Ahmed J. Kadhim and Tayseer S. Atia, "Strengthening Security and Confidentiality in E-Health Systems through Quantum Encryption of Healthcare," *AL-Iraqia Journal for Scientific Engineering Research*, vol. 2, no. 3, 2023.
12. Smita Sanjay Ambarkar and Narendra Shekokar, "Toward Smart and Secure IoT Based Healthcare System," In book: *Internet of Things, Smart Computing and Technology: A Roadmap Ahead*, Fab. 2020.
13. Seok-Jae Moon and Namju Lee, "A review of big data analytics and healthcare," February 2020.
14. Ahmed Hussein Ali, Saad Ahmed Dheyab, Abdullah Hussein Alamoodi, Aws Abed Al Raheem Magableh, and Yuantong Gu, "Leveraging AI and Big Data in Low-Resource Healthcare Settings," *Mesopotamian Journal of Big Data*, Feb 14, 2024.
15. A. Shukla and Nitin Lodha, "Role Of Cloud Security in Big Data Processing for Healthcare System," *International Conference on Advanced Computing and Communication Systems (ICACCS)*, March 2022.
16. Aijaz Reshi, Arif Shah, Shabana Shafi, and Majid Hussain Qadri, "Big Data in Healthcare-A Comprehensive Bibliometric Analysis of Current Research Trends," *Journal of Scalable Computing Practice and Experience*, Sep. 2023.
17. M. Bahadır Kalıpçı and Erkan Kadir Şimşek, "Exploring The Intersection of Big Data and Hospitality: A Bibliometric Analysis," *Journal of Business Research Turk*, October 2023.
18. Maryam Moradi, "Big Data Analysis in Healthcare," *1st International Conference on Statistical Data Analysis*, August 2023.
19. Pooja Agarwal and Virendra Kushwaha, "Revolutionizing Healthcare Through Advanced Analytics: Big Data," *International Journal of Pharma and Bio Sciences*, Oct. 2023.
20. Xiaoping Qin, Richard Wang, and Yu-Ni Huang, "Organisational Culture Research in Healthcare: A Big Data Bibliometric Study," *Journal of Healthcare*, vol. 11, no. 2, p. 169, January 2023.
21. Dila Ram Bhandari and Ron Kenett, "Paradigm Shift of Statistical Big Data in Healthcare: Management, Analysis and Future Prospects," *Journal of Pravaha*, vol. 28, no. 1, pp. 35–44, December 2022.
22. Svetlana Polubinskaya, "Big Data in healthcare: cybersecurity and legal protection of personal data," *Journal of Gosudarstvo i pravo*, January 2023.
23. Bhuvanendra Putchala, Lasya Sreevidya Kanala, Devi Prasanna Donepudi, and Hari Kishan Kondaveeti, "Applications of Big Data Analytics in Healthcare Informatics," In book: *Health Informatics and Patient Safety in Times of Crisis*, December 2022.
24. G. Owen Schaefer, "Big data in healthcare," In book: *Routledge Encyclopedia of Philosophy*, April 2021.
25. Pijush Kanti Dutta Pramanik, Saurabh Pal, and Moutan Mukhopadhyay, "Healthcare Big Data: A Comprehensive Overview," In book: *Research Anthology on Big Data Analytics, Architectures, and Applications*, January 2022.
26. M. Pavithra, E.S.Shamila, and G. Krishna Priya, "Importance of Big Data In Healthcare System A Survey Approach," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, November 2020.
27. Oluwaseun Olarewaju, "Ethical Considerations in the use of big data in Healthcare: Privacy and Ownership as a case study," University of Derby, May 2023.
28. Christos L. Stergiou, Maria Koidou, and Konstantinos E. Psannis, "IoT-Based Big Data Secure Transmission and Management over Cloud System: A Healthcare Digital Twin Scenario," *Journal of Applied Sciences*, vol. 13, no. 16, p. 9165, August 2023.
29. Guma Ali and Maad M. Mijwil, "Cybersecurity for Sustainable Smart Healthcare: State of the Art, Taxonomy, Mechanisms, and Essential Roles," *Mesopotamian Journal Of CyberSecurity*, May 23, 2024.