Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 6 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i06.48085

PAPER

Unveiling the Landscape of Big Data Analytics in Healthcare: A Comprehensive Bibliometric Analysis

Mohd Amran Mohd Daril[1](✉), Fozia Fatima[2], Samar Raza Talpur[3], Alhamzah F. Abbas[4]

[1]Quality Engineering Research Cluster, Quality Engineering Section, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Johor, Malaysia

[2]Department of Health Professions Education, National University of Medical Sciences, PWD Campus Islamabad, Rawalpindi, Pakistan

[3]Sukkur IBA University, Sukkur, Pakistan

[4]Marketing and Entrepreneurship, Universiti Teknologi Malaysia (UTM), Johor, Malaysia

mamran@unikl.edu.my

ABSTRACT

In the rapidly evolving landscape of healthcare, the digital transformation marked by healthcare 4.0 has spurred a surge in data generation, giving rise to 'big data'. Big data analytics has become an effective tool in the healthcare industry, revolutionising medical research, patient care, and healthcare management. This study undertakes a meticulous bibliometric analysis, drawing upon a dataset of 2212 articles from the Scopus database spanning 2014 to 2023, to unravel the trajectory of big data analytics in healthcare. The research explores diverse dimensions, from the distribution of studies across years to the productivity rankings of journals, countries, and institutions, elucidating the evolving trends and key contributors. Co-authorship networks and keyword co-occurrence analysis reveal thematic clusters and intellectual structures, contributing to a nuanced understanding of the field. The results underscore the escalating global interest in the fusion of big data and healthcare, illuminating collaborations, and identifying influential players. Additionally, the study identifies pressing challenges, including security concerns and skill shortages, emphasizing the imperative of overcoming these barriers for effective big data applications in healthcare. Serving as a valuable resource for researchers, practitioners, and policymakers, this research not only captures the current landscape but also provides insights for future exploration, contributing to strategic planning in this dynamic domain.

KEYWORDS

big data, healthcare, bibliometric analysis, research trends, literature review

1 INTRODUCTION

The healthcare business is undergoing a third revolution characterized by digital transformation, leading healthcare organizations to aggressively pursue methods to improve value for patients, popularly known as healthcare 4.0 [1] [2]. The adoption of health information technologies signifies a notable paradigm shift in the healthcare sector, evolving from initial digital tracking infrastructures to increasingly complex systems, including electronic health records, wearable and implantable

Mohd Daril, M.A., Fatima, F., Talpur, S.R., Abbas, A.F. (2024). Unveiling the Landscape of Big Data Analytics in Healthcare: A Comprehensive Bibliometric Analysis. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(6), pp. 4–24. https://doi.org/10.3991/ijoe.v20i06.48085

Article submitted 2023-12-08. Revision uploaded 2024-02-06. Final acceptance 2024-02-06.

devices, and more recently, the integration of cloud, fog, edge computing, and artificial intelligence [3]. These digital technologies play a pivotal role in healthcare [4] [5] optimizing patient care, waste management, and bolstering supply chain resilience [6] among various other contributions [4] [5]. Among the myriad techniques and approaches to improve healthcare, considerable academic attention has been directed toward big data [7].

The field of healthcare and medicine deals with large and complex information, known as big data, which present difficulties in analysis and management using traditional software or technology [8] [9]. Big data analytics include the integration of various data kinds, maintaining data integrity, doing analysis, developing models, interpreting outcomes, and confirming conclusions [10]. Big data analytics enables the discovery of valuable insights from large amounts of data. In the field of medicine and healthcare, big data analytics enables the examination of large datasets collected from many patients. The process entails finding patterns and relationships in datasets, as well as creating models that can make predictions using data mining techniques [11]. The field of big data analytics in medicine and healthcare combines the analysis of several scientific disciplines, including bioinformatics, medical imaging, sensor informatics, medical informatics, and health informatics.

Efforts have been made to investigate the potential of big data in supporting clinical treatments and interventions, improving decision-making for patients and health professionals, enhancing diagnostic accuracy, and optimizing healthcare systems [12] [13] [14]. Data analytics has proven instrumental in predicting heart disease [15] [16] conducting psychological analyses of language on social media [17] [18] [19] [20] and enhancing pandemic management [21]. The academic literature highlights many benefits linked to the use of big data in the healthcare sector [22]. However, there remain significant challenges and barriers that require attention, encompassing concerns such as the security of personal health records [7], increased financial investments [23], a shortage of relevant skill sets [24] and the need for a robust information technology infrastructure [25] among others. In order to successfully use big data in healthcare, it is crucial to recognise and address these problems. Understanding the significance of big data in healthcare is crucial for this objective.

In recent times, scholars have been actively involved in studying and analysing the utilisation of big data in the field of healthcare using bibliometric analysis (refer to Table 1). The table presents a concise summary of several academic research that examine the intersection of big data and healthcare. The text provides an overview of the databases used, the time frames encompassed, and the particular topics investigated in each research endeavour. Liao et al. [26] utilized the Web of Science (WoS) database spanning from 1991 to 2018 to analyse documents related to "medical big data" using GraphPad Prism 5, VOSviewer, and CiteSpace bibliometric software. Oravec et al. [27] investigated the use of large amounts of data in the area of paediatric neurosurgery by analysing data from the Web of Science database from 2000 to 2018. Wang et al. [28] examined the use of cutting-edge technologies such as big data, cloud computing, and the Internet of Things (IoT) in cancer research. The study focused on the period from 2000 to 2017 and analysed data from the Web of Science (WoS) database. Galetsi and Katsaliaki [29] examined the use of big data analytics in healthcare by using the NVivo software. The study focused on the time frame spanning from 2000 to 2016. The study conducted by dos Santos et al. [30] examined the use of data mining and machine learning methods in the field of public health, specifically focusing on articles published in the Web of Science database between

2009 and 2018. In their study, X. Chen et al. [31] conducted an extensive analysis utilising the WoS, PubMed, and Scopus databases, covering the period from 2016 to 2020. Their research specifically examined the integration of smart health technologies with artificial intelligence. Furstenau and his colleagues [22] performed a performance analysis utilising Web of Science (WoS) data from 2009 to 2020, with the assistance of a scoping review. Sikandar et al. [32] investigated the utilisation of digital technologies in the healthcare sector using Scopus database for the period between 2017 and 2021. In a separate study, Jimma [33] examined research on artificial intelligence in healthcare using Scopus database from 2000 to 2021. This collection provides a comprehensive overview of the many study areas and approaches used in the field of big data and healthcare. This research specifically examines the intersection of big data analytics and healthcare, providing an analysis that spans from 2014 to 2023, ensuring a more up-to-date perspective. The research uses Vos Viewer for bibliometric analysis, offering a distinct viewpoint on the developing terrain of big data in healthcare.

Table 1. Previous bibliometric studies on big data and healthcare research

Study	Database	Coverage	Focus
[26]	Web of Science	1991–2018	Analysed “medical big data” papers using GraphPad Prism 5, VOSviewer, and CiteSpace bibliometric software.
[27]	Web of Science	2000–2018	Exploration of large-scale data analysis in the domain of paediatric neurosurgery.
[28]	Web of Science	2000–2017	An analysis of cutting-edge technology (such as big data, cloud computing, and the Internet of Things) in the area of cancer research using the CiteSpace tool.
[29]	Web of Science	2000–2016	An investigation on the use of NVivo software for big data analytics in the healthcare sector.
[30]	Web of Science	2009–2018	An examination of the use of data mining and machine learning methods in the field of public health.
[31]	Web of Science (WoS), PubMed, and Scopus databases	2016–2020	Emphasise the integration of artificial intelligence in smart health.
[22]	Web of Science	2009–2020	The text discusses the study of a Business Process Network study (BPNA) that is supported by a Strategic Roadmap (SR). It focuses on performance analysis, strategic themes, thematic evolution structure, primary challenges, and future prospects. The analysis is conducted using the SciMAT programme and is supported by a scoping review.
[32]	Scopus	2017–2021	The use of Vos viewer in healthcare facilitates the examination of digital technology.
[33]	Scopus	2000–2021	Examination of studies pertaining to the use of artificial intelligence in the field of healthcare.
This study	Scopus	2014–2023	Analysed published literature on big data analytics and healthcare using the Scopus database and Vos Viewer.

Through bibliometric analysis, the researchers statistically examine the authors, institutions, countries, and cooperation among authors, subject areas, keywords and published articles in the given period in the selected field. The study is designed to address the following research questions (RQ):

RQ1. What is the publication trend of research on big data during the past ten years?
RQ2. What are the most prolific countries, institutions, journals and authors associated with literature on big data in healthcare?
RQ3. Which authors have made significant contributions through collaborative research?
RQ4. Which countries have made the most substantial contributions through collaborative research efforts?
RQ5. What are the most prevalent keywords linked to research on big data in healthcare?

2 METHODOLOGY

This study used the Scopus database. This database was selected as a research platform for many reasons. The Scopus database is a vast collection of peer-reviewed literature, including journals, books, and conference proceedings from many countries. It is renowned for its extensive abstracts and citations [34] [35]. Furthermore, Scopus includes the abstracts of the majority of publications published in other databases, ensuring extensive inclusion of articles. By using the snowball sampling technique, it is possible to extend the sample to other databases derived from the current database [36]. Scopus is a user-friendly platform that offers support for a range of software tools to get data for bibliometric research. These tools allow users to access information such as authors, titles, publication years, cited references, abstracts, institutions, and countries [37]. Additionally, it enables a preliminary examination of the progression of citation counts over a period of time and the identification of the writers and publications with the highest number of citations.

The research collected data on December 30th utilising the SCOPUS database, specifically focusing on data from the years 2014 to 2023. The primary search terms were "big data" and "healthcare". The search query used was "(TITLE-ABS-KEY ("big data") AND TITLE-ABS-KEY ("healthcare"))". Retrieve articles published between 2014 and 2023, inclusive, that are in English, of type "ar" (article), and in the final publication stage. Using the specified research parameters, we discovered a total of 6499 search results in the SCOPUS database. In order to prevent any confusion caused by similar language and to maintain a focused research aimed at achieving certain goals, we deliberately refrained from seeking out any comparable phrases.

The search and selection criteria for the articles are shown in Figure 1. Articles were selected using the Scopus database, as stated in the preceding section. A total of 6499 items were obtained from the initial search string, and further refined using other search criteria. To examine the latest publication trend in the specific field of research, we only chose publications that were published during the last ten years, resulting in a total of 6402 articles. In order to maintain strict adherence to the methodology, the selection procedure was further improved by only identifying the document type as "article," resulting in a total of 2,380 articles. Furthermore, the source type was limited only to journals, leading to a subset of 2,313 articles. After a careful and thorough selection process, a final dataset of 2,212 articles published in English was acquired. This dataset will serve as the foundation for the forthcoming bibliometric study. The use of the Prisma Statement facilitated the recovery of the specific group of documents, with the results of the selection criteria shown in Figure 1.

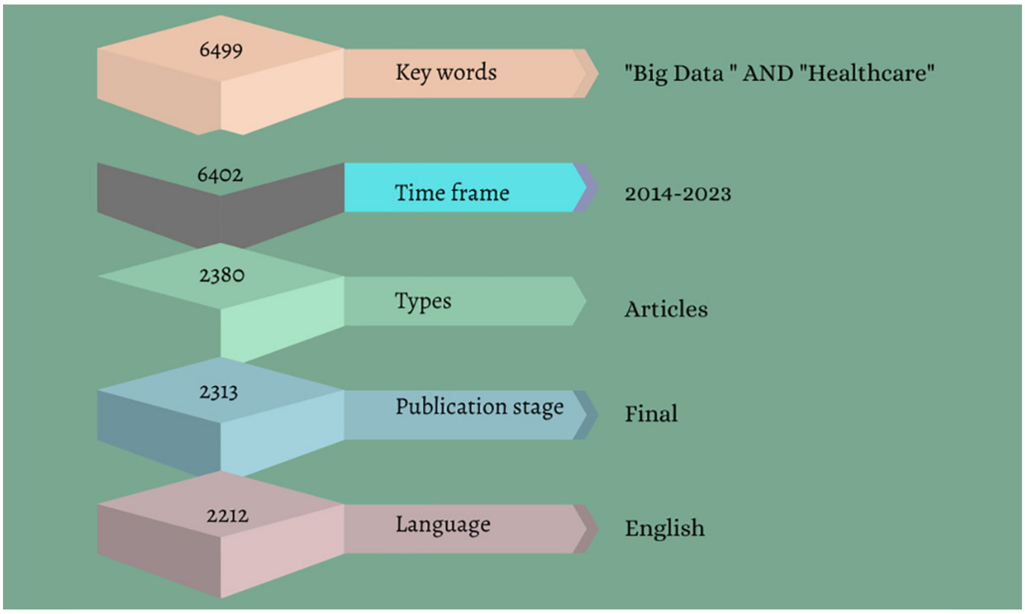

Fig. 1. Search and selection process of articles

3 ANALYSIS AND DISCUSSION OF RESULTS

3.1 Publication trend

Figure 2 reveals a consistent upward trend in the annual publication of documents related to big data in healthcare research from 2014 to 2023. In the early years, the field showed moderate growth, with 47 publications in 2014 and 82 in 2015. However, from 2016 onwards, there was a substantial increase, reaching its peak in 2018 with 223 publications. Although there was a slight decrease in 2020, the overall trajectory remained positive, with the number of documents consistently surpassing previous years. The trend reflects a growing interest and recognition among researchers of the significance and potential impact of big data in healthcare research. The sustained growth in publications each year indicates the continuous exploration and advancement of this field within the academic and research community.

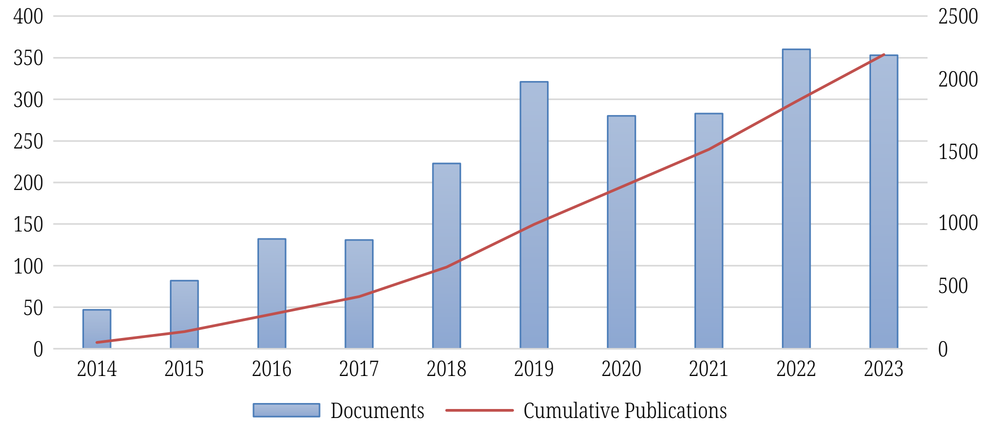

Fig. 2. Publication trend

3.2 Subject areas of the published literature

Figure 3 outlines the distribution of publications related to big data in healthcare across various academic disciplines. The numbers represent the count of publications within each discipline. The highest number of publications is found in the field of Computer Science, with 1098 publications indicating a significant emphasis on computational aspects in healthcare research. Medicine follows closely with 735 publications, highlighting the substantial contributions from the medical field. Engineering and Social Sciences also show considerable engagement, with 664 and 200 publications, respectively. Business, Management, and Accounting, as well as Mathematics, demonstrate interest with 194 and 190 publications, respectively. The distribution spans a diverse range of disciplines, including Biochemistry, Genetics, and Molecular Biology, Materials Science, Decision Sciences, Environmental Science, and Health Professions, among others. This diversity underscores the interdisciplinary nature of big data in healthcare research, with contributions from fields such as physics, nursing, chemical engineering, and economics, reflecting the multifaceted approaches and collaborations within this dynamic and evolving research domain.

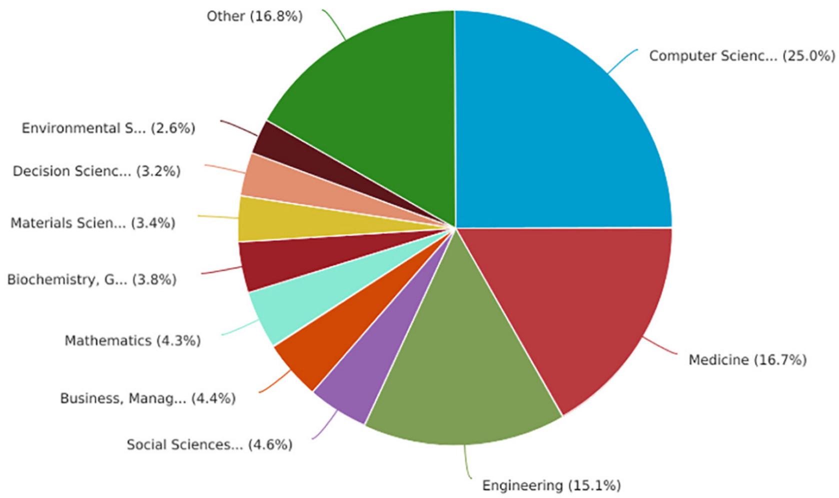

Fig. 3. Documents by subject area

3.3 Most productive countries

Figure 4 provides a detailed breakdown of the distribution of documents related to big data in healthcare research across several countries and territories. India emerges as a leading contributor with 487 documents, showcasing a substantial and active involvement in the exploration of big data applications in healthcare. Following closely, the United States demonstrates its significant role in the field, contributing 479 documents and leveraging its robust research infrastructure. China, with 299 documents, reflects a noteworthy presence and growing influence in the global discourse on big data in healthcare. The United Kingdom, South Korea, and Saudi Arabia contributed 207, 165, and 131 documents, respectively, indicating their considerable interest and active participation in big data healthcare research. Italy, Australia, Canada, and Spain also make notable contributions with 103, 96, 84, and 78 documents, respectively, underscoring their commitment to advancing knowledge in this interdisciplinary field. The diverse geographical representation in this table highlights the global engagement and collaborative efforts across countries to explore the potential and challenges of integrating big data into healthcare research.

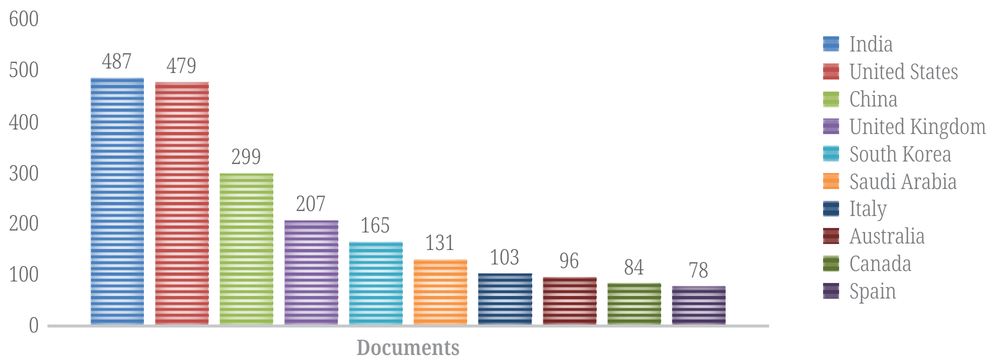

Fig. 4. Top productive countries in terms of publications

3.4 Most prolific institutions

A comprehensive ranking of academic institutions based on their prolific contributions to the research domain of big data in healthcare is provided in Table 2. King Saud University, situated in Saudi Arabia, secures the top position with an impressive count of 35 documents, affirming its pivotal role in advancing knowledge at the intersection of big data and healthcare within the region. Following closely, the Vellore Institute of Technology in India claims the second spot with 33 documents, underscoring India's substantial scholarly output and active engagement in this evolving field. Huazhong University of Science and Technology in China holds the third position with 29 documents, showcasing China's significant and growing influence in the global discourse on big data applications in healthcare research. The inclusion of King Abdulaziz University from Saudi Arabia in the fourth spot, with 26 documents, further emphasizes the nation's sustained commitment to and leadership in big data healthcare research. Harvard Medical School, representing the United States, secures the fifth position with 23 documents, highlighting its prominent role in shaping the global landscape of big data research in the medical domain. The subsequent institutions in the ranking, including those from India, South Korea, Singapore, England, and China, contribute to the diverse and collaborative international efforts aimed at advancing knowledge in the dynamic field of big data in healthcare.

Table 2. Most prolific institutions

Rank	Affiliation	Country	Documents
1	King Saud University	Saudi Arabia	35
2	Vellore Institute of Technology	India	33
3	Huazhong University of Science and Technology	China	29
4	King Abdulaziz University	Saudi Arabia	26
5	Harvard Medical School	United States of America	23
6	SRM Institute of Science and Technology	India	17
7	Kyonggi University	South Korea	17
8	National University of Singapore	Singapore	16
9	University College London	England	16
10	Tsinghua University	China	16

3.5 Most productive journals

We have provided a comprehensive overview of the distribution of documents across various sources in the context of big data analytics in healthcare. The results of our analysis revealed that IEEE Access emerges as the leading source with 74 documents, followed by the Journal of Big Data with 30 documents, showcasing their substantial contributions to the field. Other influential sources include Future Generation Computer Systems, the International Journal of Environmental Research and Public Health, and Sustainability Switzerland, each contributing significantly with 28, 26, and 26 documents, respectively. The list encompasses a diverse array of journals and publications, such as the IEEE Internet of Things Journal, Applied Sciences Switzerland, and BMJ Open, each playing a crucial role in disseminating research findings in the intersection of big data and healthcare. This diverse array of sources highlights the multidisciplinary nature of the field, reflecting contributions from journals spanning computer science, healthcare, and engineering, among others. The list of top influential journals is shown in Table 3.

Table 3. Top influential journals

Source	Documents
IEEE Access	74
Journal of Big Data	30
Future Generation Computer Systems	28
International Journal of Environmental Research and Public Health	26
Sustainability Switzerland	26
IEEE Internet of Things Journal	24
Applied Sciences Switzerland	22
International Journal of Innovative Technology and Exploring Engineering	20
International Journal of Recent Technology and Engineering	20
Journal of Advanced Research in Dynamical and Control Systems	18
Journal of Medical Systems	18
BMJ Open	17
International Journal of Advanced Computer Science and Applications	16
Sensors Switzerland	15
Healthcare Switzerland	14
International Journal of Engineering and Technology UAE	14
Journal of Computational and Theoretical Nanoscience	14
Yearbook of Medical Informatics	14

3.6 Most influential authors

Table 4 offers an insightful glimpse into the top 10 influential authors in the realm of big data analytics within the healthcare domain. The authors are ranked based on the number of documents they have contributed to the selected field.

The results of our analysis revealed that Chung, K. from Kyonggi University in South Korea, is leading the list with an impressive 17 publications. Notably, his work has garnered a significant total citation count of 3625, and his H Index stands at 35, reflecting substantial impact and influence. Following closely is Khoshgoftaar, T.M. from Florida Atlantic University, USA, with 11 documents. His remarkable total citation counts of 30,633 and an H Index of 66 underscore the substantial impact of his contributions to the field. Similarly, Chen, M. from South China University of Technology in China has made significant strides with 8 documents and an H Index of 80. Rodrigues, J.J.P.C. from Portugal, affiliated with Instituto de Telecomunicações, has also made a notable impact with 8 documents, a total citation count of 29,858, and an H Index of 88. The list further includes prolific authors such as Shankar, K. from India, Elhoseny, M. from the United Arab Emirates, Farahani, B. from Iran, and others, each with their unique contributions and impact in the field.

These top authors (Table 4), identified through bibliometric analysis, play pivotal roles in advancing knowledge and shaping the landscape of big data analytics in healthcare. Their prolific output and influence, as measured by citations and the H Index, contribute significantly to the scholarly discourse and progress in this dynamic interdisciplinary field.

Table 4. Top 10 influential authors in terms of publications

Author	Documents Published in the Selected Field	Affiliation	Scopus ID	TC	Total Published Documents	H Index
Chung, K.	17	Kyonggi University, Suwon, South Korea	25927027500	3625	205	35
Khoshgoftaar, T.M.	11	Florida Atlantic University, Boca Raton, United States	7006211475	30,633	739	66
Chen, M.	8	South China University of Technology, Guangzhou, China	25821032200	25,617	408	80
Rodrigues, J.J.P.C.	8	Instituto de Telecomunicações, Aveiro, Portugal	25930566300	29,858	1,181	88
Shankar, K.	8	Saveetha School of Engineering, Chennai, India	56884031900	6,695	265	46
Elhoseny, M.	7	Info University of Sharjah, Sharjah, United Arab Emirates	57148260400	11,846	310	63
Farahani, B.	7	Shahid Beheshti University, Tehran, Iran	57225781040	1,654	53	15
Firouzi, F.	7	Duke University, Durham, United States	36056471000	1,945	69	20

(Continued)

Table 4. Top 10 influential authors in terms of publications *(Continued)*

Author	Documents Published in the Selected Field	Affiliation	Scopus ID	TC	Total Published Documents	H Index
Chang, V.	6	Aston University, Birmingham, United Kingdom	56926234700	13,217	450	59
Hossain, M.S.	6	King Saud University, Riyadh, Saudi Arabia	24066717900	17,065	376	66

4 BIBLIOMETRIC MAPS

This section presents a bibliometric analysis of various aspects, including the co-authorship of countries, the co-authorship of authors and the co-occurrence of author keywords.

4.1 Co-authorship (Authors)

The co-authorship analysis quantifies the number of articles on which two researchers have cooperated. Co-occurrence TLS refers to the frequency at which two words are referenced together in a set of publications. [38] [39]. This research employs a co-authorship analysis to ascertain the collaboration network and writers who have made noteworthy contributions to the area by working together. Co-authorship analysis is the predominant research tool used to examine research cooperation (RC).

To conduct the analysis of co-authorship, authors were selected based on a minimum threshold of 2 documents produced by each participant. Out of the 2173 writers, only 45 met the qualifying criterion. Out of the total of 45 writers, 22 authors had strong connections with each other, resulting in the formation of 4 separate clusters (as shown in Figure 5). The 22 writers exhibit a high level of interconnectivity and have made substantial contributions to the literature via their combined efforts.

This research employs a co-authorship analysis to ascertain the collaboration network and writers who have made significant contributions to the area by working together. The TLS represents the overall strength of the author's connections with other authors [40]. The co-authorship analysis of authors involved setting a threshold of 2 documents as the minimum requirement for authors, and a minimum of 0 citations for an author. Among the total of 2173 authors, only 45 authors satisfied the threshold criteria. Among the 50 authors, 22 were highly connected and formed 4 distinct clusters, as depicted in Figure 7. These 22 authors are highly interconnected and have made significant contributions to the literature through their collaborative work.

This research does a co-authorship analysis to discover the collaborative network and writers who have contributed to the field via their joint effort. Co-authorship analysis is the primary research methodology used to examine research cooperation (RC). The TLS represents the overall strength of an author connections with other authors [41]. The examination of co-authorship among authors involved establishing a criterion where each individual must have authored at least two documents and received a minimum of zero citations.

Among the total of 2173 writers, only 45 authors fulfilled the threshold criteria. Among the 50 writers, 22 of them were well related and formed 4 distinct clusters, as shown in Figure 5. This signifies that these 22 writers possess strong networks and have made valuable contributions to the field of literature via collaborative efforts.

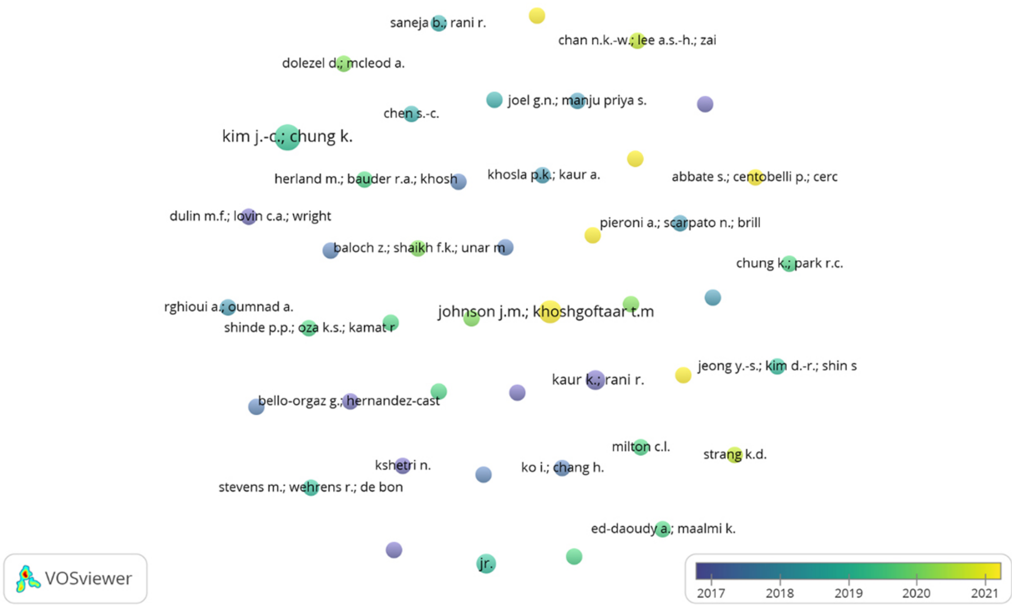

Fig. 5. Snapshot of the bibliometric map representing co-authorship analysis of authors in overlay visualisation mode

The co-authorship analysis, as revealed in the provided Table 5 from Vos Viewer, illuminates the collaborative dynamics and scholarly influence of several author pairs in the field of big data analytics and healthcare. Notably, the collaboration between Hossain M.S. and Muhammad G. has resulted in two publications, amassing an impressive citation count of 290, underscoring the impact of their joint contributions. Similarly, the collaborative efforts of Jain P., Gyanchandani M., and Khare N. have yielded two documents with a notable citation count of 245, showcasing their scholarly recognition. Authors Kim J.-C. and Chung K. stand out with a prolific collaboration, producing five documents and accumulating 154 citations, indicative of substantial influence in the academic community. These findings provide valuable insights into collaborative networks and the scholarly influence of authors within this research domain. Key authors doing collaborative research are shown in Table 5.

Table 5. Key authors doing collaborative research

Author	Documents	Citations
Hossain M.S.; Muhammad G.	2	290
Jain P.; Gyanchandani M.; Khare N.	2	245
Kim J.-C.; Chung K.	5	154
Manogaran G.; Lopez D.	2	150
Sahoo P.K.; Mohapatra S.K.; Wu S.-L.	2	144

4.2 Co-authorship (Countries)

To assess the co-authorship among countries and illustrate the collaboration among authors from different geographic regions, a criterion of 10 countries was set.

This implies that each country must have a minimum of 10 published documents. Consequently, among the 120 countries, 51 successfully reached the required level. In Figure 6, the proximity of the countries indicates a strong affiliation between them.

The co-authorship analysis of countries, as presented in the data from Vos Viewer, reveals the collaborative dynamics and impact of various nations in the field of big data analytics and healthcare research. The United States emerges as a prominent contributor, leading in both document count (480) and citations (17,620), resulting in a total link strength of 400. The United Kingdom follows closely, demonstrating a robust collaborative network with 207 documents, 7,493 citations, and a total link strength of 324. China, with 299 documents and 10,858 citations, showcases its global influence, accompanied by a total link strength of 258. India contributes notably with 486 documents and 9,618 citations, reflecting a strong collaborative network and a total link strength of 247. Saudi Arabia, with 131 documents and 4,529 citations, reflects significant collaborative efforts, resulting in a total link strength of 194. Spain, Germany, and Australia exhibit collaborative research endeavours, contributing 78 documents (2,137 citations, total link strength: 162), 65 documents (1,616 citations, total link strength: 158), and 96 documents (3,036 citations, total link strength: 155), respectively. The Netherlands, Italy, and South Korea also contribute significantly, reflecting the diverse and impactful nature of collaborative research in big data analytics and healthcare.

India, Saudi Arabia, and Spain also exhibit substantial collaborative efforts, contributing significantly to the field. Notably, Australia, the Netherlands, and Italy demonstrate noteworthy contributions in terms of both document count and citations. The overall analysis underscores the global nature of collaborative research in big data analytics and healthcare, with countries playing key roles in shaping the scholarly landscape and fostering international partnerships. The variations in total link strength indicate the strength and depth of collaborative ties among these nations. Figure 6 shows the results of co-authorship analysis of countries and Table 6 lists the countries according to their ranks as per their total link strength.

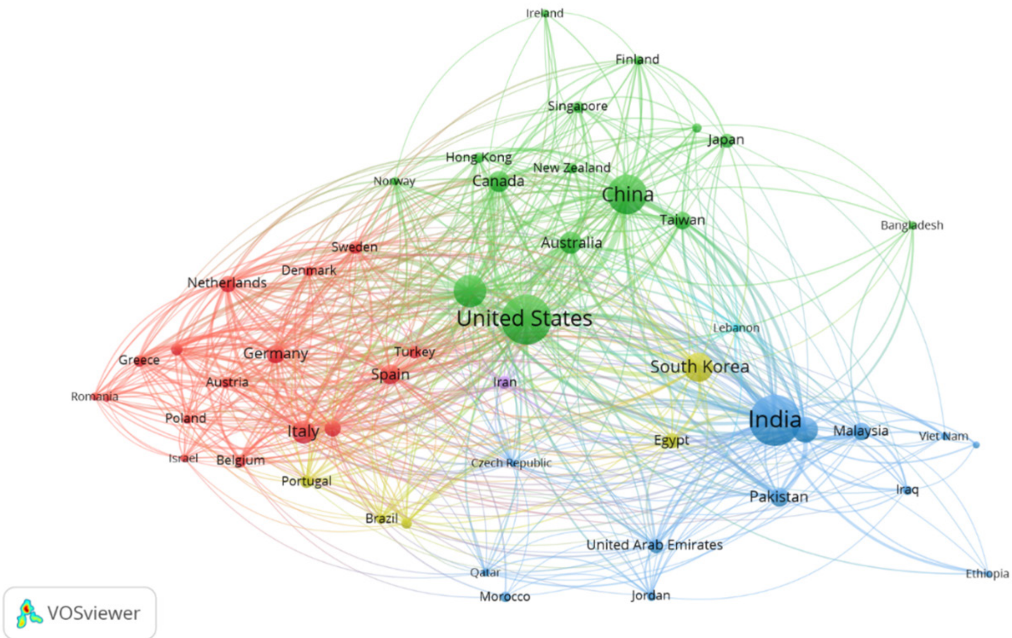

Fig. 6. Snapshot of the bibliometric map representing co-authorship analysis of countries in overlay visualisation mode

Table 6. Co-authorship analysis of countries

Country	Documents	Citations	Total Link Strength
United States	480	17620	416
United Kingdom	207	7493	340
China	299	10858	268
India	486	9618	257
Saudi Arabia	131	4529	196
Spain	78	2137	176
Germany	65	1616	168
Netherlands	54	1487	166
Australia	96	3036	159
Italy	103	2519	143
South Korea	165	4187	131
Canada	84	3527	128
France	51	1545	119
Switzerland	29	1008	111
Pakistan	68	2223	106
Sweden	31	1147	101
Austria	23	572	94
Greece	30	667	91
Portugal	35	1393	81
Egypt	44	1341	80

The countries listed, including Hungary, Ethiopia, the Czech Republic, Indonesia, Vietnam, Bangladesh, Ireland, Norway, Lebanon, Qatar, and Romania, appear to be engaging in collaborative work in big data analytics and healthcare to a lesser extent compared to the top contributors (see Table 7). Each of these countries has published a relatively modest number of documents, ranging from 10 to 15, indicating a lower volume of research output in comparison to the leading nations. The cited documents and total link strength values for these countries also reflect a less extensive collaborative network, suggesting that there might be room for increased international cooperation and partnerships in the field. These countries may benefit from fostering more collaborative efforts to enhance the impact and reach of their research in the intersection of big data analytics and healthcare.

Table 7. Countries with least collaborative research in the area of big data and healthcare

Country	Documents	Citations	Total Link Strength
Hungary	10	249	27
Ethiopia	11	105	15
Czech Republic	12	559	54
Indonesia	12	64	10
Viet Nam	13	278	22

(Continued)

Table 7. Countries with least collaborative research in the area of big data and healthcare *(Continued)*

Country	Documents	Citations	Total Link Strength
Bangladesh	13	374	18
Ireland	13	132	13
Norway	14	762	41
Lebanon	14	303	37
Qatar	14	363	30
Romania	15	382	29

4.3 Co-occurrence (Keywords)

The concept of "co-occurrence" refers to the frequency of occurrence of a keyword within a particular document [42]. The total link strength (TLS) denotes the frequency of occurrence of a word within a document. The larger the node, the more frequently occurring keyword it is, according to node size. The closer two keywords are to each other, the thicker the line between them; and the stronger the association [40] [43].

In the process of keyword analysis, upon importing Scopus data into the VOS viewer, a minimum requirement of 10 occurrences for keywords was set. This led to the identification of 94 keywords among the initial 5545. Subsequently, identical keywords were merged, resulting in a final count of 5529 keywords, of which only 78 satisfied the specified criterion. To clarify, only 78 keywords achieved the prescribed threshold of 10 occurrences per keyword. The individual keyword occurrence was consistently set at 10, generating 8 distinct clusters of keywords. The outcome of the keyword co-occurrence is illustrated in Figure 7.

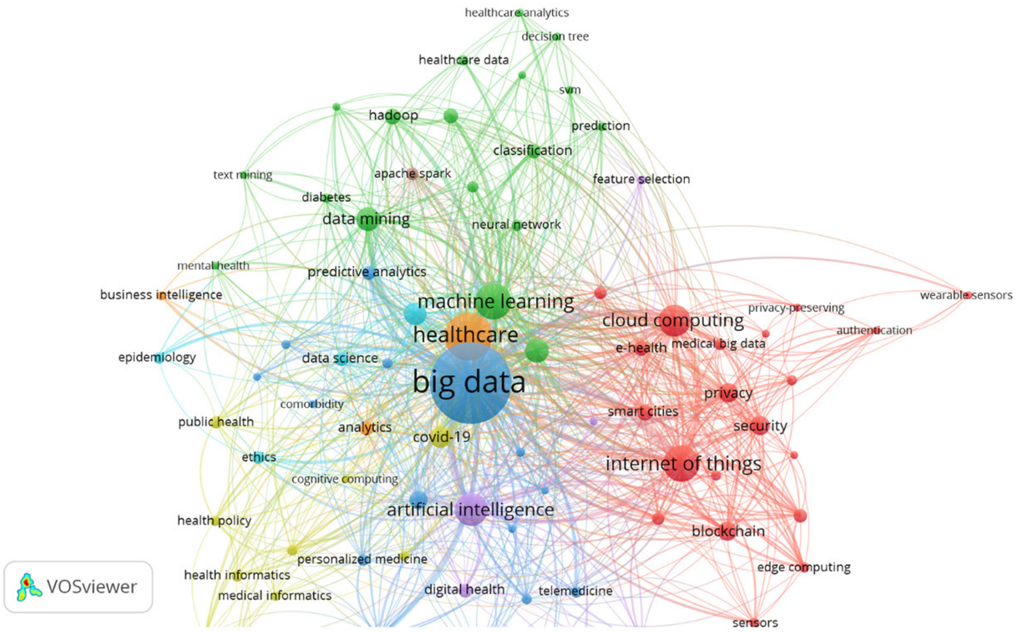

Fig. 7. Snapshot of the bibliometric map representing the co-occurrence of keyword analysis in network visualisation mode

The keyword "Big Data" demonstrates a high total link strength (TLS) of 1448, suggesting strong associations with other keywords, and it occurs frequently (975 times) in the analysed literature. Similarly, "Healthcare" exhibits a considerable TLS of 644, indicating substantial connections with other keywords, and it occurs 334 times. "Internet of Things" demonstrates a TLS of 444 and occurs 188 times, showcasing its relevance and frequent appearance. "Machine Learning" has a TLS of 424 and occurs 195 times, indicating its prominence in the literature. "Artificial Intelligence" has a TLS of 347 and occurs 157 times, underlining its significant role. "Cloud Computing" has a TLS of 342 and occurs 148 times, emphasizing its interconnectedness and frequency. Other keywords, such as "Deep Learning," "Data Mining," "Data Analysis," "Security," "Covid-19," "Blockchain," "Privacy," and "Smart Cities," also exhibit noteworthy TLS and occurrences, illustrating their importance in the analysed literature. Top keywords in the selected research field are shown in Table 8 along with their TLS, occurrences and links.

Table 8. Results of the co-occurrence of keywords as per their Total Link Strength (TLS)

Keywords	Links	Total Link Strength (TLS)	Occurrences
Big Data	77	1448	975
Healthcare	63	644	334
Internet Of Things	49	444	188
Machine Learning	64	424	195
Artificial Intelligence	51	347	157
Cloud Computing	52	342	148
Deep Learning	45	204	85
Data Mining	44	172	88
Data Analysis	43	147	76
Security	30	138	53
Covid-19	35	127	69
Blockchain	26	122	55
Privacy	28	111	53
Smart Cities	28	93	37

We have conducted an analysis of the literature to identify underexplored keywords, pointing towards potential areas for future research in this field. This exploration offers valuable insights that could serve as a foundation for scholars seeking opportunities to contribute further to this area of study. These keywords are listed in Table 9.

Table 9. Least explored keywords

Keywords	Links	Total Link Strength (TLS)	Occurrences
Wearable Sensors	6	11	10
Data Sharing	7	12	10
Privacy-Preserving	9	12	10

(Continued)

Table 9. Least explored keywords *(Continued)*

Keywords	Links	Total Link Strength (TLS)	Occurrences
Security And Privacy	11	12	11
Comorbidity	9	13	10
Innovation	6	14	10
Clinical Decision Support	8	15	10
Mhealth	10	15	10
Random Forest	8	16	10
Healthcare Analytics	11	17	11
Svm	11	18	10
Text Mining	12	18	10
Business Intelligence	8	19	13
Decision Tree	10	19	11
Feature Selection	10	20	14
Decision Making	11	21	12
Digital Transformation	13	21	10
Mental Health	13	21	11
Diabetes Mellitus	14	22	10
Federated Learning	13	22	15

The presented table outlines keywords that have been relatively less explored in existing literature on healthcare technology, indicating potential avenues for future research. Each keyword is accompanied by the number of links, total link strength (TLS), and occurrences within the analysed literature. Notable keywords include "Wearable Sensors," with moderate link strength and occurrences, suggesting room for further exploration in health monitoring through wearable devices. "Data Sharing" and "Privacy-Preserving" exhibit moderate link strength and occurrences, indicating a need for deeper investigations into the implications of data sharing and privacy in healthcare. The combined focus on "Security and Privacy" highlights a critical aspect with moderate link strength and occurrences. Keywords like "Comorbidity" and "Clinical Decision Support" also show moderate link strength and occurrences, presenting areas deserving of additional research attention. Other keywords such as "Innovation," "Mhealth," and "Random Forest" display moderate link strength and occurrences, indicating ongoing interest with potential for more comprehensive studies. The keywords with higher TLS and occurrences, such as "Business Intelligence," "Feature Selection," and "Federated Learning," suggest areas with more established research but still open to nuanced exploration. In essence, this analysis identifies specific keywords that, despite some attention, present opportunities for researchers to delve deeper into less-explored aspects of healthcare technology.

5 CONCLUSION AND DISCUSSION

To summarise, the comprehensive bibliometric analysis carried out in this work has shed light on several aspects of big data analytics in healthcare research.

The study conducted a comprehensive analysis of many academic articles sourced from the Scopus database. Its objective was to identify patterns, prominent authors, significant works, and theme groups in this rapidly evolving subject. By using bibliometric techniques, the study output's influence was measured and assessed, offering a comprehensive grasp of the intellectual framework related to big data analytics in healthcare.

An investigation of co-authorship revealed collaborative networks, highlighting the worldwide scope of research efforts in this field. The examination of co-authorship across nations demonstrated the preeminence of the United States, the United Kingdom, China, India, and Saudi Arabia in their joint research endeavours. The linked global network represents the collective dedication of these nations to promoting the comprehension and use of big data analytics in healthcare.

The co-authorship analysis identified many authors who made significant contributions, including Hossain M.S., Muhammad G., Jain P., Gyanchandani M., Khare N., Kim J.-C., Chung K., Manogaran G., Lopez D., Sahoo P.K., Mohapatra S.K., Wu S.-L., Kshetri N., and Rubí J.N.S. The writers have shown great involvement and influence in the collaborative network, indicating their important contributions to influencing the discussion on big data analytics in healthcare. These co-authorship analysis findings provide useful insights into the collaborative dynamics among authors and nations, confirming the cooperative character of research efforts in this topic within a wider framework. The results not only recognise the present level of research but also underscore the significance of worldwide cooperation in propelling developments and discoveries at the convergence of big data and healthcare.

The keyword co-occurrence analysis yielded a complete perspective on the dominant topics and primary areas of focus in the literature. Key ideas such as "Big Data," "Healthcare," "Machine Learning," and "Artificial Intelligence" have arisen as crucial fields of study, highlighting the interrelatedness of various subjects in the research field.

In addition, the selection of terms with lower frequencies and overall connection strength emphasised possible topics for further investigation. The keywords "Wearable Sensors," "Data Sharing," and "Privacy-Preserving" indicate areas that should be further explored to provide significant insights. These underexplored keywords provide academics the chance to fill important knowledge gaps in the field of healthcare technology.

This study adds to the continuing discussion on the intersection of big data analytics and healthcare in a wider sense. The results provide a comprehensive understanding of the changing environment, offering scholars, practitioners, and policy-makers a useful resource. This study synthesises information from several sources, acknowledging the existing level of research and paving the path for future inquiry and improvements in using big data to improve healthcare outcomes.

6 LIMITATIONS AND FUTURE DIRECTIONS

1. Although our bibliometric study has yielded significant information about the state of big data analytics in healthcare, it is crucial to recognise certain limits and outline future research priorities in this rapidly evolving subject.
2. A disadvantage of the research is that it only considers papers from 2014 to 2023. Subsequent investigations should expand the duration of this period to include developing patterns and guarantee the study stays up to date.

3. Another restriction is related to the intrinsic limitations of bibliometric analysis, which depends on measurable data derived from academic publications. Thorough evaluations and detailed examinations of the specific details and complexities of big data analytics in healthcare are essential for complete comprehension. Incorporating qualitative approaches, such as case studies or interviews, in future research projects may provide detailed insights into the influence of big data analytics on healthcare practices.
4. The ethical aspects of using big data analytics in healthcare need more investigation. It is essential to examine issues of privacy, data security, and the ethical ramifications of using big data for healthcare purposes. Researchers should thoroughly investigate the socioeconomic ramifications and ethical implications linked to the growing dependence on big data in the healthcare sector.
5. Future research should investigate geographical disparities in the implementation and effects of big data analytics, considering the worldwide character of healthcare concerns. An investigation comparing various locations and healthcare systems would provide insight into differing research priorities, patterns of cooperation, and methods for implementation. Implementing this method might assist in customising therapies for certain circumstances and contribute to a more detailed comprehension of the worldwide ramifications of big data analytics in the healthcare sector.
6. Collaborative research networks have been crucial in increasing knowledge in this field. Future study should investigate the intricacies of interdisciplinary cooperation, exploring how academics from different areas join forces to contribute to the progress in big data analytics in healthcare.
7. Continuous surveillance of keyword trends and new topics is essential for keeping up with developing priorities in big data analytics in healthcare. Researchers must be alert and responsive to new problems and opportunities, ensuring that their study aligns with the constantly evolving healthcare and technology field. To enhance our comprehension of the transformational power of big data analytics in healthcare, it is crucial to acknowledge these constraints and embrace future research avenues.

7 ACKNOWLEDGEMENT

The research reported in this paper was supported by the Short-Term Research Grant (STRG) No UniKL/CoRI/str19057 awarded by Universiti Kuala Lumpur, Malaysia.

8 REFERENCES

[1] M. S. Dohan, C. B. Califf, K. Ghosh, and J. Tan, "Digital transformation in healthcare: New value for a new movement," *Health Policy and Technology*, vol. 9, no. 2, 2020. https://doi.org/10.1016/j.hlpt.2020.04.003

[2] P. P. Jayaraman, A. R. M. Forkan, A. Morshed, P. D. Haghighi, and Y. Bin Kang, "Healthcare 4.0: A review of frontiers in digital health," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 2, 2020. https://doi.org/10.1002/widm.1350

[3] M. L. Kolling et al., "Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, 2021. https://doi.org/10.3390/ijerph18063099

[4] É. Maillet et al., "Laboratory testing in primary care: A systematic review of health IT impacts," *International Journal of Medical Informatics*, vol. 116, 2018. https://doi.org/10.1016/j.ijmedinf.2018.05.009

[5] R. D. Kindle, O. Badawi, L. A. Celi, and S. Sturland, "Intensive care unit telemedicine in the era of big data, artificial intelligence, and computer clinical decision support systems," *Critical Care Clinics*, vol. 35, no. 3, 2019. https://doi.org/10.1016/j.ccc.2019.02.005

[6] L. B. Furstenau, C. Zani, S. X. Terra, M. K. Sott, K. K. R. Choo, and T. A. Saurin, "Resilience capabilities of healthcare supply chain and supportive digital technologies," *Technology in Society*, vol. 71, 2022. https://doi.org/10.1016/j.techsoc.2022.102095

[7] L. Chen, N. Zhang, H. M. Sun, C. C. Chang, S. Yu, and K. K. R. Choo, "Secure search for encrypted personal health records from big data NoSQL databases in cloud," *Computing*, vol. 102, no. 6, 2020. https://doi.org/10.1007/s00607-019-00762-z

[8] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: Promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, 2014. https://doi.org/10.1186/2047-2501-2-3

[9] A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: Introduction to the special section," *Information Systems Frontiers*, vol. 18, no. 2, 2016. https://doi.org/10.1007/s10796-016-9641-2

[10] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, and M. D. Wang, "-Omic and electronic health record big data analytics for precision medicine," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, 2017. https://doi.org/10.1109/TBME.2016.2573285

[11] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: Big data for personalized healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, 2015. https://doi.org/10.1109/JBHI.2015.2406883

[12] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: Using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, 2014. https://doi.org/10.1377/hlthaff.2014.0041

[13] M. Ghassemi, L. A. Celi, and D. J. Stone, "State of the art review: The data revolution in critical care," *Critical Care*, vol. 19, no. 1, 2015. https://doi.org/10.1186/s13054-015-0801-4

[14] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Future Generation Computer Systems*, vol. 82, pp. 375–387, 2018. https://doi.org/10.1016/j.future.2017.10.045

[15] R. Khamisy-Farah, L. B. Furstenau, J. D. Kong, J. Wu, and N. L. Bragazzi, "Gynecology meets big data in the disruptive innovation medical era: State-of-art and future prospects," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, 2021. https://doi.org/10.3390/ijerph18105058

[16] N. J. L. Brown and J. C. Coyne, "Does Twitter language reliably predict heart disease? A commentary on Eichstaedt et al. (2015a)," *PeerJ*, vol. 2018, no. 9, 2018. https://doi.org/10.7717/peerj.5656

[17] P. M. Valkenburg, "Social media use and well-being: What we know and what we need to know," *Current Opinion in Psychology*, vol. 45, p. 101294, 2022. https://doi.org/10.1016/j.copsyc.2021.12.006

[18] J. Gamble, A. Gilmore, D. Mccartan-Quinn, and P. Durkan, "The marketing concept in the 21st century: A review of how marketing has been defined since the 1960s," *The Marketing Review*, vol. 11, no. 3, pp. 227–248, 2011. https://doi.org/10.1362/146934711X589444

[19] J. C. Eichstaedt et al., "Facebook language predicts depression in medical records," in *Proceedings of the National Academy of Sciences of the United States of America*, 2018, vol. 115, no. 44, pp. 11203–11208. https://doi.org/10.1073/pnas.1802331115

[20] K. Jaidka, S. Giorgi, H. A. Schwartz, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods," in *Proceedings of the National Academy of Sciences of the United States of America*, 2020, vol. 117, no. 19, pp. 10165–10171. https://doi.org/10.1073/pnas.1906364117

[21] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How big data and artificial intelligence can help better manage the covid-19 pandemic," *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, 2020. https://doi.org/10.3390/ijerph17093176

[22] L. B. Furstenau et al., "Big data in healthcare: Conceptual network structure, key challenges and opportunities," *Digital Communications and Networks*, vol. 9, no. 4. 2023. https://doi.org/10.1016/j.dcan.2023.03.005

[23] C. Schaeffer, L. Booton, J. Halleck, J. Studeny, and A. Coustasse, "Big data management in US hospitals: Benefits and barriers," *Health Care Manager*, 2016. https://doi.org/10.1097/HCM.0000000000000139

[24] G. L. Tortorella et al., "Effects of contingencies on healthcare 4.0 technologies adoption and barriers in emerging economies," *Technological Forecasting and Social Change*, 2020. https://doi.org/10.1016/j.techfore.2020.120048

[25] P. Ajmera and V. Jain, "Modelling the barriers of health 4.0–the fourth healthcare industrial revolution in India by TISM," *Operations Management Research*, vol. 12, no. 3–4, 2019. https://doi.org/10.1007/s12063-019-00143-x

[26] H. Liao, M. Tang, L. Luo, C. Li, F. Chiclana, and X. J. Zeng, "A bibliometric analysis and visualization of medical big data research," Sustainability (Switzerland), vol. 10, no. 1, 2018. https://doi.org/10.3390/su10010166

[27] C. S. Oravec, M. Motiwala, K. Reed, T. L. Jones, and P. Klimo, "Big data research in pediatric neurosurgery: Content, statistical output, and bibliometric analysis," *Pediatric Neurosurgery*, vol. 54, no. 2, 2019. https://doi.org/10.1159/000495790

[28] X. Wang, J. Guo, D. Gu, Y. Yang, X. Yang, and K. Zhu, "Tracking knowledge evolution, hotspots and future directions of emerging technologies in cancers research: A bibliometrics review," *Journal of Cancer*, vol. 10, no. 12, 2019. https://doi.org/10.7150/jca.32739

[29] P. Galetsi and K. Katsaliaki, "Big data analytics in health: An overview and bibliometric study of research activity," *Health Information and Libraries Journal*, vol. 37, no. 1, 2020. https://doi.org/10.1111/hir.12286

[30] B. S. dos Santos, M. T. A. Steiner, A. T. Fenerich, and R. H. P. Lima, "Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018," *Computers and Industrial Engineering*, vol. 138, 2019. https://doi.org/10.1016/j.cie.2019.106120

[31] X. Chen, H. Xie, Z. Li, G. Cheng, M. Leng, and F. L. Wang, "Information fusion and artificial intelligence for smart healthcare: A bibliometric study," *Information Processing and Management*, vol. 60, no. 1, 2023. https://doi.org/10.1016/j.ipm.2022.103113

[32] H. Sikandar, A. F. Abbas, N. Khan, and M. I. Qureshi, "Digital technologies in healthcare: A systematic review and bibliometric analysis," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 8, pp. 34–48, 2022. https://doi.org/10.3991/ijoe.v18i08.31961

[33] B. L. Jimma, "Artificial intelligence in healthcare: A bibliometric analysis," *Telematics and Informatics Reports*, vol. 9, 2023. https://doi.org/10.1016/j.teler.2023.100041

[34] E. Ahmed, A. Ullah, H. Sikandar, N. Khan, and M. Hassan, "Mapping the literature of social network ties in business and management field: A bibliometric study," *Journal of Management Info*, vol. 9, no. 3, pp. 330–346, 2022. https://doi.org/10.31580/jmi.v9i3.2688

[35] H. Sikandar and U. H. Abdul Kohar, "A bibliometric analysis of green innovation research," *Systematic Literature Review and Meta-Analysis Journal*, 2022. https://doi.org/10.54480/slrm.v3i1.32

[36] F. F. Silveira and S. N. Zilber, "Is social innovation about innovation? A bibliometric study identifying the main authors, citations and co-citations over 20 years," *International Journal of Entrepreneurship and Innovation Management*, vol. 21, no. 6, pp. 459–484, 2017. https://doi.org/10.1504/IJEIM.2017.086936

[37] H. Sikandar et al., "Eco-innovation in Small and Medium Enterprises (SMEs): A systematic literature review," *Journal of the Knowledge Economy*, vol. 8, no. 1, p. 129, 2023. https://doi.org/10.1007/s13132-023-01367-w

[38] N. J. Van Eck and L. Waltman, "Manual for VOSviwer version 1.6.10," *CWTS Meaningful metrics*, January, pp. 1–53, 2019.

[39] H. Sikandar, U. H. A. Kohar, and S. Salam, "The evolution of social innovation and its global research trends: A bibliometric analysis," *Systematic Literature Review and Meta-Analysis Journal*, vol. 1, no. 2, pp. 117–133, 2021. https://doi.org/10.54480/slrm.v1i2.9

[40] N. J. Van Eck and L. Waltman, "Manual_VOSviewer_1.6.16.pdf," www.vosviewer.com, 2020.

[41] N. J. Van Eck and L. Waltman, "VOSviewer manual: Manual for VOSviewer version 1.6.7," *Univeristeit Leiden*, no. February, p. 51, 2018.

[42] I. Mustapha, M. Ali, N. Khan, and H. Sikandar, "The impact of industry 4.0 on innovative organisations, a thematic review using the PRISMA statement 2020," *International Journal of Interactive Mobile Technologies*, 2023. https://doi.org/10.3991/ijim.v17i09.39465

[43] H. Sikandar et al., "Mapping the development of open innovation research in business and management field: A bibliometric analysis," *Journal of the Knowledge Economy*, 2023. https://doi.org/10.1007/s13132-023-01280-2

9 AUTHORS

Mohd Amran Mohd Daril, Quality Engineering Research Cluster, Quality Engineering Section, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Johor, Malaysia (E-mail: mamran@unikl.edu.my).

Fozia Fatima, Department of Health Professions Education, National University of Medical Sciences, PWD Campus Islamabad, Rawalpindi, Pakistan.

Samar Raza Talpur, Director ICT, Sukkur IBA University, Sukkur, Pakistan.

Alhamzah F. Abbas, Senior Lecturer in Marketing and Entrepreneurship, Universiti Teknologi Malaysia (UTM), Johor, Malaysia.