OPEN ACCESS

Manuscript ID:
ASH-2024-12017762

Volume: 12

Issue: 1

Month: July

Year: 2024

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Received: 20.04.2024

Accepted: 26.06.2024

Published: 01.07.2024

Citation:
Mallikarjuna, B., et al. "Bibliometric Analysis of the Journal of Big Data: Trends, Impact, and Future Directions." *Shanlax International Journal of Arts, Science and Humanities*, vol. 12, no. 1, 2024, pp. 100–14.

DOI:
https://doi.org/10.34293/sijash.v12i1.7762

Bibliometric Analysis of the Journal of Big Data: Trends, Impact, and Future Directions

B. Mallikarjuna
Head, Library and Information Centre
Government First Grade College for Women, Koppal, Karnataka, India
https://orcid.org/0000-0002-1113-1386

Mallikarjun Kappi
Head, Library and Information Centre
Government First Grade College, Hospet, Karnataka, India
https://orcid.org/0000-0003-1964-3498

Salma Taj
Assistant Professor, Department of Commerce
Government First Grade College, Jagalur, Davanagere, Karnataka, India

Abstract
The Journal of Big Data, a peer-reviewed publication since 2014, has played a pivotal role in the field of big data research by disseminating a total of 788 papers over its ten years of activity. Of these, 649 papers have garnered citations, with 46 receiving over 100 citations. The journal maintains an average of 37.13 citations per paper, with a higher average of 45.09 citations per cited paper, reflecting its significant impact. An h-index of 70 and a g-index of 158 further underscore its academic influence. The journal's active year growth rate of 39.88% signifies its consistent expansion, while the average document age of 2.83 years indicates a focus on current developments. Collaborative research is a hallmark of the journal, evident in its 3.47 average authors per paper and 19.67% of papers with international co-authorships. The analysis of year-wise performance and citation trends offers insights into scholarly output and engagement. Visualizing collaborative relationships among authors highlights prominent nodes of interaction, revealing key contributors. Noteworthy institutions such as Florida Atlantic University, USA, and Bina Nusantara University, Indonesia, emerge in the landscape of research productivity. An exploration of keywords and their co-occurrence uncovers thematic clusters, ranging from big data technologies and deep learning to natural language processing and machine learning. In conclusion, the Journal of Big Data stands as a vital platform for disseminating impactful research in the dynamic realm of big data. Its comprehensive analysis emphasizes its academic influence through citation metrics, collaborative research, and thematic explorations. Through its decade-long journey, the journal has solidified its role in advancing the field and fostering global collaboration among researchers.

Keywords: Bibliometrics Analysis, Journal of Big Data, Citation Analysis, Highly Cited Papers, Topical Trends

Introduction

The Journal of Big Data (JBD) is a prominent open-access publication that serves as a significant platform for researchers and practitioners in data science and analytics. Covering diverse topics, including deep learning algorithms and applications of big data, the journal is a seminal source of innovative material for the academic community, facilitating global collaboration and knowledge dissemination. Dedicated to exploring the challenges facing big data today and, in the future, the journal delves into various aspects, such as data capture and storage, search, sharing, analytics, big data technologies, visualization, and architectures for massively parallel processing. Additionally, it addresses

crucial topics like data mining tools, machine learning algorithms for big data, cloud computing platforms, distributed file systems, databases, and scalable storage systems. This comprehensive approach offers a multifaceted view of big data research.

The impressive citation impact of the journal, evidenced by its 2-Year Impact Factor of 8.1 (2022), Source Normalized Impact per Paper (SNIP) of 5.096, SCImago Journal Rank (SJR) of 2.714, and Cite Score of 23.9, signifies its recognition and influence within the academic community. These metrics underscore the quality and relevance of its published research, establishing its significance in shaping the landscape of big data analytics. Moreover, the JBD prides itself on its efficient and timely handling of submissions, with a median time of 54 days to the first decision for all manuscripts and 106 days for reviewed manuscripts only. The median time from submission to acceptance (181 days) and publication (196 days) further exemplifies the journal's commitment to the prompt dissemination of research findings.

The journal's wide readership and high download statistics, with 2,250,289 downloads in 2022, and substantial online visibility, evidenced by 982 Altmetric mentions in 2021, confirm its strong impact on the research community. This broad outreach ensures that research published in the journal reaches a diverse audience, contributing to knowledge diffusion and fostering global collaborations. Led by distinguished Editors-in-Chief, Borko Furht and Taghi Khoshgoftaar, both affiliated with Florida Atlantic University, USA, the journal benefits from their expertise and guidance, ensuring the quality and integrity of its content. The crucial role played by the Managing Editor, Muhammad Tanveer Jan, in overseeing the editorial process and upholding journal standards further enhances its reputation. This bibliometric study provides a valuable overview of the Journal, a prominent journal in the field of big data analytics (Kappi and Biradar; Prakash et al.). The study utilizes bibliometric analysis to explore the journal's trajectory over its first 10 years. The findings of the study provide insights into the journal's development and identify areas for future improvement.

The JBD has emerged as a leading source of research on deep learning algorithms and applications of big data. This study provides a retrospective analysis of JBD since its inception, to understand its productivity and impact. This article provides an in-depth analysis of the publication, citation, collaboration, and methodological choices of authors in the JBD. The article uses a variety of methods to examine the publication and citation patterns of JBD articles, as well as the trends of contributions among research constituents at the author, institutional, and country levels. The study also examines the thematic structure of JBD articles. The study finds that a variety of themes are covered in JBD articles, with the most common themes being deep learning, big data analytics, and machine learning. The study concludes by discussing five research questions:

- What is the publications growth and citation trend of JBD?
- What is the methodological distribution among the JBD publications?
- What is the level of collaboration among authors, institutions, and countries in the (JBD)?
- What are the most cited articles and the citation distribution of the JBD?
- What are the most prominent keywords and knowledge themes in JBD articles?

The study's findings are valuable for identifying research priorities and guiding future investigations, as well as providing practitioners with a comprehensive understanding of the field. The journal is committed to open access, comprehensive coverage of big data topics, impressive citation impact, an efficient review process, and a global readership. These factors make the journal a vital source of innovative material and a catalyst for advancements in the realm of big data analytics.

The thematic analysis of JBD will provide scholars with a comprehensive overview of the field of deep learning algorithms and applications of big data. The insight into the drivers of citation will guide researchers toward the improved impact of their work. In addition to the above, the study's findings will help scholars to better understand the productivity and impact of JBD. The study's findings will also help practitioners to better understand the latest trends in big data research and development.

Related Literature

The field of bibliometrics has become increasingly important in recent years, as it provides a valuable tool for assessing the impact of journals and identifying trends in research. Many journals have developed bibliometric overviews of their journals, especially for the celebration of a special event. Noteworthy among these scholarly endeavors are research works that delve into the realm of computer science and data science, employing bibliometric methodologies to explore pertinent facets include (Akturk; Jalali; Mallikarjun Kappi et al.; Merigó et al.; Ravichandra Rao and Raghavan; Sajovic et al.; Vaishya et al.; Raju Vaishya et al.; Zurita et al.). This is a useful way to provide a comprehensive bibliometric overview of the Journal of Big Data, focusing on its publications spanning the period from 2014 to 2023.

Data and Methodology

The bibliographic information for the JBD was extracted from the comprehensive Scopus database, renowned as a vast repository of peer-reviewed scholarly literature. This database is widely utilized for similar research studies, as evidenced by previous studies (Chaman Sab et al.; Naveen Donthu et al.; Goyal and Kumar; Kappi et al.; Modak et al.; Raju Vaishya et al.; Zurita et al.). The search for the journal name was conducted within the 'source title' category of Scopus in early August 2023, resulting in the identification of 788 documents relevant to JBD. While bibliometrics is rooted in library sciences, its application extends across diverse fields (Ellegaard and Wallin) leveraging quantitative techniques to extract meaningful insights.

Bibliometric analysis is the most appropriate choice for this study, as it is superior to other review techniques (Spurk). Firstly, it enables the analysis of an extensive corpus of data, thereby providing a broader analytical scope. Secondly, the methodology employs a range of bibliographic tools to delve into bibliometric metrics, facilitating a more nuanced examination. Thirdly, the quantitative nature of this approach yields impartial results. Lastly, the use of networks and graphs enhances data visualization for readers (Donthu et al.). Nonetheless, it is important to acknowledge the presence of limitations within this methodology.

One notable limitation pertains to the challenge of making qualitative assertions based on quantitative outcomes, potentially subjecting qualitative findings to debate. Given the dependence on bibliographic data, any inaccuracies within the data could potentially impact the study's conclusions. Considering these advantages and limitations, the present study employs a combination of bibliometric analysis of the JBD corpus and content analysis to bolster qualitative interpretations. This analysis encompasses the examination of JBD's performance with respect to productivity (publications), influence (citations), and authorship composition. To achieve this, various matrices were employed, including total publications (TP), total citations (TC), number of cited publications (NCP), total citations per publication (CPP), number of active years (NAY), productivity per active year (PAY), number of contributing authors (NCA), and h and g indexes. These metrics were computed using MS Excel, enabling a comprehensive evaluation of JBD's trends, impact, and future directions.

Overall Summary of the Journal of Big Data

The JBD is a peer-reviewed journal that publishes research papers on big data. It has been in publication since 2014 and has published a total of 788 papers in that time. Of these papers, 649 have been cited at least once, and 46 have been cited more than 100 times. The journal has an average of 37.13 citations per paper, and an average of 45.09 citations per cited paper. The journal's h-index is 70, and its g-index is 158. The journal has an active year growth rate of 39.88%, which means that the number of papers published in the journal has been growing by an average of 39.88% each year. The average document age in the journal is 2.83 years, which means that the papers in the journal are relatively recent. A total of 2113 authors contributed to the journal, of which 58 have written single-authored papers. The remaining 2055 authors have written collaborative papers. The average number of authors per paper in the journal is 3.47, and 19.67% of the papers in the journal have international co-authorships.

Table 1 provides a comprehensive summary of the metrics associated with the JBD. Over a span of ten active years, from 2014 to 2023, the journal has

published a total of 788 articles. Out of these, 772 are regular articles, nine are reviews, six are errata, and one has been retracted. The impact has garnered significant attention within the academic community, with a total of 649 publications cited. Notably, there are 46 highly cited papers, defined as those with 100 or more citations. This indicates the journal's contributions to the field and the value it brings to the scientific community. The JBD has accumulated a remarkable total of 29,262 citations, translating to an average of 37.13 citations per published paper. Further emphasizing its influence, the average number of citations per cited paper stands at 45.09, demonstrating the high quality and relevance of the research published. Another indicator of the journal's growth is its active year growth rate, which stands at an impressive 39.88%. This shows a steady expansion in the number of publications and recognition over the ten years. Considering the age of the documents, the journal maintains a relatively low document average age of 2.83, indicating that the research published is quite up-to-date and relevant to current developments in the field of big data.

In terms of authorship metrics, the JBD has seen contributions from 2,113 authors, with 58 single-authored papers and 725 collaborative papers. This shows that a substantial portion of the research is the result of collaboration between multiple authors. On average, there are 3.47 co-authors per paper, highlighting the collaborative nature of research within this domain. Additionally, the journal has a relatively international outlook, with 19.67% of papers featuring co-authors from different countries. The JBD demonstrates significant influence and recognition within the scientific community, as evidenced by its high citation counts, h-index of 70, and g-index of 158. With a consistent and robust growth rate, the journal continues to play a vital role in advancing research in the field of big data and fostering international collaboration among researchers.

Table 1 Summary of the Journal

Description	Metrics
Journal of Big Data	
Total Publications (TP)	788
Total Cited Publications	649
Total Highly Cited Papers (HCP=100 or above Citations)	46
Number of active years (NAY) 2014-2023	10
Active Year Growth Rate %	39.88
Document Average Age	2.83
Article	772
Review	9
Erratum	6
Retracted	1
Journal Citations Metrics	
Total Citations (TC)	29262
Average citations per Paper (TC/TP)	37.13
Average citations per Cited Paper	45.09
Average citations per Active Year (TC/Y)	2926.20
h-index	70
g-index	158
Authorship Metrics	
Authors	2113
Authors of single-authored papers	58
Collaborative Papers	725
Single-authored Papers	63
Co-Authors per Paper	3.47
International Co-authorships %	19.67

Year-wise Performance and Citation Trend

Table 2 provides a comprehensive analysis of the year-wise performance and citation trends from 2014 to 2023. The findings indicate a steady level of scholarly output, with an increasing trend in the number of papers published from 2014 to 2021, peaking at 159 papers in 2021. However, a decline was observed in 2022 and 2023, with 124 and 123 papers, respectively. The total citations also showed an increasing trend from 2014 to 2019, with a peak of 10,261 citations in 2019. However, the number of citations declined significantly in 2022 and 2023, with 549 and 40 citations, respectively. The average citations per paper (CPP) also exhibited fluctuations, with a decline since 2017, reaching its lowest value of 0.33 in 2023. The number of highly cited papers (HCP) displayed variations over the years, with a peak of 14 in 2019 and no HCP in 2022 and 2023. The presence of funded papers (FP) showcased a steady increase until 2021, followed by a decline in 2022 and 2023. In contrast, the number of authors

(TA) contributing to the journal publications showed a progressive rise, reaching its peak at 541 in 2021.

The findings hold significant implications for the scholarly community, offering valuable insights into the impact and influence of this journal over the examined period. The steady increase in the number of papers published indicates a continued dedication to knowledge dissemination and scholarly contributions. However, the decline in citations and CPP raises questions about the potential shifts in research interest and the relative impact of individual papers. The presence of highly cited papers in certain years highlights the recognition and influence of specific research work within the academic domain. Notably, fluctuations in the number of funded papers may signify changes in research funding opportunities and the dynamics of collaborative research endeavors. The progressive rise in the number of authors involved in journal publications underscores the growing trend of interdisciplinary collaboration and knowledge sharing as displayed in Figure 1.

Table 2 Journal Year-Wise Performance with Various Parameters

Year	TP	TC	CPP	HCP	FP	TA
2014	6	304	50.67	1	3	16
2015	27	4564	169.04	10	15	86
2016	26	3337	128.35	2	11	71
2017	49	1603	32.71	3	20	148
2018	53	2211	41.72	6	17	157
2019	112	10261	91.62	14	59	366
2020	109	3062	28.09	6	63	364
2021	159	3331	20.95	4	95	541
2022	124	549	4.43	0	72	454
2023	123	40	0.33	0	79	533
Total	788	29262	37.13	46	434	2736

TP=Total Papers; TC=Total Citations; CPP=Citations per Paper; HCP=Highly Cited Papers; FP=Funded Papers; TA=Total Authors

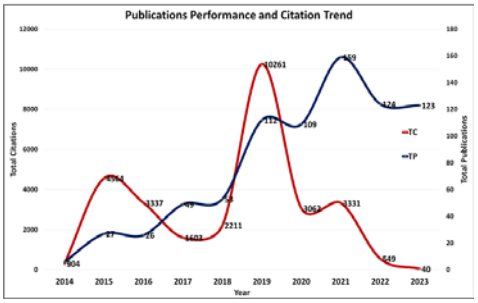

Figure 1 Publications Performance and Citation Trend

Most Productive Authors in JBD

Table 3 presents a ranking of the top 15 most productive authors in the JBD, based on their total number of publications followed by their SAP, MAP, TC, CPP, HCP, h-index, g-index, and NAY. The most productive author in JBD is Khosgoftaar TM, with 44 papers published, 15 HCP, and an h-index of 24. He has been actively contributing to JBD for 9 years. The second most productive author is Leevy JL, with 14 papers published, 1 HCP, and an h-index of 9. She has been actively contributing to JBD for 5 years. The third most productive author is Zhan J, with 12 papers published, 2 HCP, and an h-index of 11. He has been actively contributing to JBD for 8 years. The analysis of the most productive authors in JBD offers a comprehensive view of the scholarly contributions, affiliations, and citation metrics that shape the domain of big data research. It emphasizes the global nature of big data research and underscores the significance of multi-author collaborations in advancing knowledge in this rapidly evolving field. The citation metrics provide a quantitative measure of the impact of each author's work, showcasing their influence within the academic community. This analysis reaffirms the JBD's role as a leading platform for disseminating cutting-edge research and fostering collaboration among researchers worldwide.

Table 3 Most Contributing Authors

Author	Affiliation	TP	SAP	MAP	TC	CPP	HCP	h_index	g_index	NAY
Khoshgoftaar TM	Florida Atlantic University, USA	44	0	44	12460	283.18	15	24	44	9
Leevy JL	Florida Atlantic University, USA	14	0	14	580	41.43	1	9	14	5

Zhan J	University of Arkansas, USA	12	0	12	646	53.83	2	11	12	8
Furht B	Florida Atlantic University, USA	9	0	9	321	35.67	2	6	9	4
Douzi S	Mohammed V University, Morocco	8	0	8	134	16.75	0	6	8	2
Kumar S	South Ural State University, Russia	7	0	7	634	90.57	1	7	7	8
Villanustre F	LexisNexis Business Information Solutions, USA	7	0	7	1549	221.29	1	6	7	8
Hancock J	Florida Atlantic University, USA	7	0	7	61	8.71	0	4	7	2
Bauder RA	Florida Atlantic University, USA	6	0	6	533	88.83	0	6	6	5
Budiharto W	Bina Nusantara University, Indonesia	6	1	5	201	33.50	1	5	6	5
Douzi K	University Hassan II, Morocco	6	0	6	98	16.33	0	5	6	2
Khare N	MANIT, Bhopal, India	6	0	6	258	43.00	1	4	6	7
Wibisono A	Universitas Indonesia, Indonesia,	6	0	6	35	5.83	0	4	5	4
Asemi A	Corvinus University of Budapest, Hungary	6	0	6	25	4.17	0	3	5	2
Sarker IH	Swinburne University of Technology, Australia	5	1	4	426	85.20	2	5	5	4
Seliya N	Ohio Northern University, USA	5	0	5	1880	376.00	2	5	5	8
Agarwal A	Florida Atlantic University, USA	5	0	5	53	10.60	0	4	5	7
Chowanda A	Bina Nusantara University, Indonesia	5	1	4	80	16.00	0	4	5	4
Hssina B	University Hassan II, Morocco	5	0	5	92	18.40	0	4	5	2
Jafar A	Higher Institute for Applied Sciences and Technology, Syrian Arab Republic	5	0	5	214	42.80	1	4	5	4
Jain P	MANIT, Bhopal, India	5	0	5	256	51.20	1	4	5	7
Surendro K	Bandung Institute of Technology, Indonesia	5	0	5	41	8.20	0	4	5	2
Zuech R	Florida Atlantic University, USA	5	0	5	271	54.20	1	4	5	8

Mursanto P	Universitas Indonesia, Indonesia	5	0	5	23	4.60	0	3	4	4
Suhartono D	Bina Nusantara University, Indonesia	5	0	5	41	8.20	0	3	5	3

TP=Total Publications; SAP=Single Author Papers; MAP=Multi-authors Paper; TC=Total Citations; CPP=Citations per Paper; HCP=Highly Cited Papers; NAY=Number of Active Years

Figure 2 shows the top 20 collaborative relationships between authors. The size of the nodes represents the number of publications by the corresponding author, and the thickness of the lines represents the strength of the collaboration between the two authors. The figure shows that Khoshgoftaar TM (TLS=46) is the most central node in the network, indicating that he has the highest number of co-authorship links with other authors in the journal. Other prominent coauthorship groups include authors such as Seliya N (TLS=11), Zuech R, and Villanovastre F (TLS=10 each). These 20 authors were grouped into six clusters, suggesting that there are a number of different research communities within the field. The high level of collaboration among these authors suggests that there is a strong community of researchers working in this area. This is likely because the field is relatively new and there is still a lot of research that needs to be done.

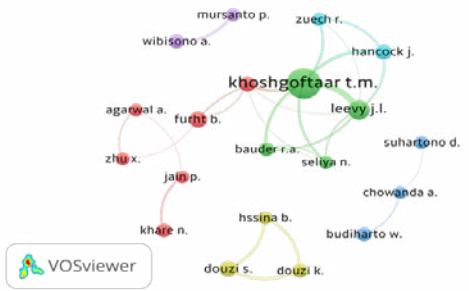

Figure 2 Top 20 Authors Collaboration Network

Most Contributed Institutions

Table 4 provides a comprehensive analysis of the most contributed institutions to JBD. These institutions have made significant contributions to the advancement of big data knowledge and have had a major impact on the academic community. Florida Atlantic University, USA, is the most prolific institution, with 57 total publications. The institution's research output has garnered significant attention and recognition within the scholarly community, with a remarkable total citation count of 12,615 and a high citation per paper (CPP) of 221.32. The presence of 15 highly cited papers further signifies the institution's ability to produce influential and impactful research. Bina Nusantara University in Indonesia is the second most contributed institution. Despite a lower total publication count, the institution's research has made a considerable impact, as indicated by a substantial total citation count of 307. Though the CPP is comparatively lower, the institution's research has contributed significantly to the knowledge base in big data, making it an essential player in the field. Hassan II University of Casablanca in Morocco holds a prominent position among the most contributed institutions. Despite not having any highly cited papers, the institution's research output has garnered 216 total citations, suggesting a notable level of recognition within the academic community. The institution's dedication to research over 10 active years signifies its continued relevance and commitment to advancing big data knowledge.

The other institutions listed in the table have also made significant contributions to the field of big data research. Their contributions reflect the diversity and global reach of big data studies, contributing to a rich and varied landscape within the Journal. The institutions listed in the table play a crucial role in shaping the field of big data research within the JBD. Their substantial contributions in terms of research productivity, citation impact, and longevity of activity highlight their significance in advancing knowledge and understanding in this rapidly evolving field. These institutions serve as pillars of academic excellence and innovation, fostering collaboration and knowledge exchange among researchers worldwide, and solidifying the JBD position as a leading platform for disseminating cutting-edge research in the domain of big data. Figure 3 shows the top 20 collaborative relationships between institutions.

Table 4 Most Contributed Institutions

Institution	TP	CP	TC	CPP	HCP	NAY	PAY	h_index
Florida Atlantic University, USA	57	49	12615	221.32	15	10	5.70	24
Bina Nusantara University, Indonesia	20	17	307	15.35	1	6	3.33	8
Hassan II University of Casablanca, Morocco	19	19	216	11.37	0	6	3.17	10
Universitas Indonesia, Indonesia	17	16	144	8.47	0	5	3.40	7
Higher Institute for Applied Sciences and Technology, Syrian Arab Republic	11	10	449	40.82	2	3	3.67	9
Sidi Mohamed Ben Abdellah University, Morocco	11	10	277	25.18	0	5	2.20	6
Mohammed V University in Rabat, Morocco	11	10	140	12.73	0	3	3.67	10
University of Nevada, USA	11	10	212	19.27	0	4	2.75	10
Sharif University of Technology, Iran	10	8	124	12.40	0	5	2.00	6
University of Belgrade, Serbia	10	2	29	2.90	0	3	3.33	2
Faculty of Sciences Rabat, Morocco	8	7	77	9.63	0	4	2.00	6
Queensland University of Technology, Australia	8	8	1331	166.38	1	5	1.60	4
Islamic Azad University, Iran	8	7	110	13.75	0	4	2.00	5
Bahar Dar University, Ethiopia	8	7	51	6.38	0	3	2.67	4
K L Deemed to be University, India	8	7	101	12.63	0	4	2.00	5

TP=Total Publications; CP=Cited Papers; TC=Total Citations; CPP=Citations per Paper; HCP=Highly Cited Papers; NAY=Number of Active Years; PAY=Productivity per Active Year

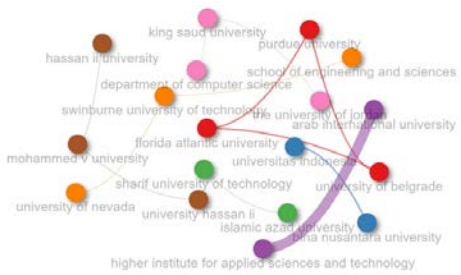

Figure 3 Top 20 Institutions Collaboration Network

Most Productive Countries

Table 5 presents a comprehensive and comparative analysis of the most productive countries. The USA is the most productive country, with 164 publications, total citations of 16,890, and an h-index of 36. These metrics indicate that the USA has made significant contributions to the field of big data research and has had a major impact on the academic community. India is in second place with 71 TP. While India's TC count is lower than that of the USA, the country's CPP score of 31.21 indicate that its research output has had a significant impact. India also has 5 HCPs, which further demonstrates the quality of its research contributions. Indonesia ranks third with 60 publications. Indonesia's CPP score of 15.83 is also noteworthy, suggesting that its research has had a significant impact. The country also has 6 HCPs, which further demonstrates its research contributions.

Morocco, Iran, and China follow with 53, 47, and 40 TP, respectively. These countries all have respectable CPP scores, which indicates the quality of their research. Morocco and Iran also have 2 HCPs each, while China has 1. The table also includes countries with lower but notable research contributions, such as Australia, Italy, and the UK. Although their TP counts are lower, their research has resulted in substantial citation counts, and they have produced highly cited papers. Table 5 reveals that the United States, India, and Indonesia are among the most productive countries in the field of big data research. These countries have made significant contributions to the field and have had a major impact on the academic community. Morocco, Iran, and China also demonstrate noteworthy research engagement and impactful contributions to the field. The metrics presented offer valuable insights into the scholarly productivity and impact of each country, highlighting the global nature of big data research

and the importance of international collaboration in advancing knowledge in this domain. These findings underscore the significance of the JBD as a leading platform for disseminating influential research and fostering collaboration among researchers worldwide.

Table 5 Most Productive Countries

Country	TP	CP	TC	CPP	HCP	NAY	PAY	h_index
USA	164	140	16890.00	102.99	22	10	16.40	36
India	71	60	2216.00	31.21	5	9	7.89	21
Indonesia	60	56	950.00	15.83	2	6	10.00	15
Morocco	53	49	1276.00	24.08	2	7	7.57	16
Iran	47	42	680.00	14.47	0	7	6.71	15
China	40	22	824.00	20.60	1	9	4.44	8
Australia	37	32	2140.00	57.84	4	9	4.11	15
Italy	35	28	850.00	24.29	1	8	4.38	14
UK	35	28	1493.00	42.66	1	8	4.38	11
Malaysia	24	21	500.00	20.83	2	8	3.00	10
Ethiopia	23	20	164.00	7.13	0	7	3.29	8
Canada	21	16	413.00	19.67	0	7	3.00	9
Saudi Arabia	21	16	177.00	8.43	0	5	4.20	9
Syrian Arab Republic	17	15	540.00	31.76	2	4	4.25	13
Germany	16	11	179.00	11.19	0	8	2.00	5
South Korea	16	5	72.00	4.50	0	3	5.33	4
Russian Federation	15	12	694.00	46.27	1	5	3.00	9
UAE	15	12	111.00	7.40	0	5	3.00	6
Japan	14	13	53.00	3.79	0	6	2.33	4
Nigeria	14	11	202.00	14.43	0	6	2.33	7

TP=Total Publications; CP=Cited Papers; TC=Total Citations; CPP=Citations per Paper; HCP=Highly Cited Papers; NAY=Number of Active Years; PAY=Productivity per Active Year

Figure 4 presents the top 20 collaborative countries contributing to the JBD in distinct clusters. In the Blue cluster, the United States (US) leads with 164 documents and a total link strength of 46, indicating substantial research output and collaboration. The United Kingdom (UK) and United Arab Emirates (UAE) also contribute notably with 35 and 12 documents, respectively. China leads the green cluster with 40 documents and a total link strength of 10, followed by India with 71 documents and a link strength of 11. Saudi Arabia and Morocco represent the Lavender cluster with moderate collaboration. Australia leads the red cluster with 37 documents and a link strength of 10. Italy dominates the yellow cluster with 35 documents and a link strength of 10, while the Russian Federation and the Syrian Arab Republic also show involvement. The US, China, and Australia emerge as influential contributors, while other countries also demonstrate valuable research output and collaboration in JBD.

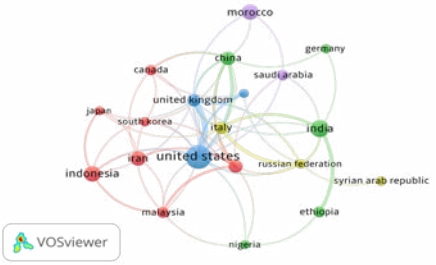

Figure 4 Top 20 Countries Collaboration Network

Distribution of Citations and Highly Cited Papers

Table 6 illustrates the distribution of citations across 788 papers, revealing valuable insights

about the impact and influence of research in the field. The analysis showcases a common citation distribution pattern, where a majority of papers (139) remain uncited, indicating their early stage of dissemination and recognition. This is typical, especially for recently published work. The citation distribution follows a power-law or long-tail pattern, with a small number of highly cited papers and a large number of papers with lower citation counts. The most cited paper receives an impressive 10,633 citations, highlighting its significant impact on the field. However, the majority of papers fall into the 11-50 citation slab (229 papers), indicating moderate attention and citations.

As the number of citations increases, the number of papers in each slab declines, demonstrating the gradual drop-off in citation frequency. This pattern is consistent with the typical citation distribution observed in various research domains. Moreover, the cumulative citations highlight that a relatively small subset of highly cited papers (79 papers, approximately 10% of the total) contributes to more than half of the total citations (15,641). This skewness in the distribution signifies the presence of influential papers that drive a substantial portion of the overall impact. The citation distribution pattern observed in this analysis is common in academic research, and it signifies a healthy and dynamic research field. It suggests that a few highly influential papers have a significant impact on the field's progress, while the majority of papers contribute to the broader body of knowledge with relatively fewer citations.

Table 7 shows the top 10 most cited papers in the JBD. The most highly cited paper is 'A Survey on Big Data Analytics' by Shorten et al. (2019), with a total of 4300 citations. This paper has been cited an average of 860 times per year and has a normalized citation score of 46.93. The second most HCP is 'A Survey of Big Data Quality Issues' by Weiss et al. (2016), with a total of 2635 citations. This paper has been cited an average of 329.38 times per year and has a normalized citation score of 20.53. The third most HCP is "A Survey on Big Data Management" by Najafabadi et al. (2015), with a total of 1452 citations. This paper has been cited an average of 161.33 times per year and has a normalized citation score of 8.59.

The other papers in the top 10 are all highly cited, with total citations ranging from 1182 to 514. These papers cover a wide range of topics related to big data, including big data analytics, big data management, big data quality, big data security, and big data applications. The significance of these papers is that they have had a major impact on the field of big data. They have been cited by other researchers, and they have been used to teach students about big data. These papers have helped to shape the way that we think about big data, and they have helped to advance the field of big data research.

Table 6 Citation Distribution

Citation Slab	Papers	Citations	Cum Citations
Uncited	139	0	0
01-10	324	1407	1407
11-50	229	4934	6341
51-99	50	3583	9924
100-499	40	8211	18135
500-1000	1	514	18649
1001-4300	5	10633	29282
0-4300	788	29282	

Table 7 Top 10 Highly Cited Papers

Paper	DOI	TC	TCpY	Normalized TC
Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning.	10.1186/s40537-019-0197-0	4300	860.00	46.93
Weiss, K. et al. (2016). A survey of transfer learning.	10.1186/s40537-016-0043-6	2635	329.38	20.53
Najafabadi, M. M. et al. (2015). Deep learning applications and challenges in big data analytics.	10.1186/s40537-014-0007-7	1452	161.33	8.59

Alzubaidi, L. et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.	10.1186/s40537-021-00444-8	1182	394.00	56.42
Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance.	10.1186/s40537-019-0192-5	1064	212.80	11.61
Tsai, C. W. et al. (2015). Big data analytics: a survey.	10.1186/s40537-015-0030-3	514	57.11	3.04
Dash, S. et al. (2019). Big data in healthcare: management, analysis and future prospects.	10.1186/s40537-019-0217-0	498	99.60	5.44
Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data.	10.1186/s40537-015-0015-2	434	48.22	2.57
Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis.	10.1186/s40537-019-0276-2	411	82.20	4.49
Leevy, J. L. et al. (2018). A survey on addressing high-class imbalance in big data.	10.1186/s40537-018-0151-6	348	58.00	8.34

Topical Trends in the JBD

The top 50 keywords that co-occur most frequently in Table 8 are organized into clusters based on their thematic relationships. These clusters represent the key topics studied by scholars in the field of data science.

Keywords co-occurrence analysis is a powerful technique for identifying thematic relationships and prevailing trends within a body of research. By analysing the frequency and link strength of keywords that co-occur in a corpus of text, researchers uncover significant patterns that reflect the interconnected nature of various concepts and the overarching themes driving scholarly discourse. For example, if the keywords “big data” and “machine learning” co-occur frequently in a corpus of text, this suggests that there is a strong relationship between these two concepts. This information could be used to identify research papers that are relevant to both big data and machine learning or to develop new machine learning algorithms that are specifically designed for big data applications.

Table 8 Most Occurred Keywords in JBD Publications

Cluster	Keywords	Occurrence
Cluster 1: Big Data (occurrences: 449, total link strength: 418)	Big Data; Hadoop; Mapreduce; Big Data Analytics; Apache Spark; Internet of Things; Cloud Computing; Stream Processing; Unsupervised Learning; Big Data Applications; K-Means Clustering	200; 26; 30; 33; 12; 11; 9; 6; 6; 5; 5
Cluster 2: Deep Learning (occurrences: 117, total link strength: 144)	Deep Learning; Convolutional Neural Network; Transfer Learning; Image Classification; Lstm; Neural Networks; Object Detection; Ensemble Learning; Dimensionality Reduction; Imbalanced Data	86; 27; 16; 9; 10; 10; 5; 7; 5; 6
Cluster 3: Natural Language Processing (occurrences: 60, total link strength: 73)	Sentiment Analysis; Natural Language Processing; Twitter; Social Media; Text Mining; Social Networks; Topic Modelling; Text Classification; Covid-19; Information Retrieval	25; 20; 18; 17; 8; 12; 6; 5; 6; 5
Cluster 4: Clustering (occurrences: 40, total link strength: 40)	Clustering; Data Science; Predictive Analytics; Intelligent Systems; K-Means; Principal Component Analysis	20; 11; 7; 6; 6; 5
Cluster 5: Machine Learning (occurrences: 134, total link strength: 144)	Deep Neural Networks; Anomaly Detection; Google Trends; Time Series	14; 7; 6; 5

Cluster 6	Machine Learning; Genetic Algorithm; Bayesian Approach	117; 8; 5
Cluster 7	Data Mining; Data Analytics; Distributed Computing	33; 14; 5
Cluster 8	Class Imbalance; Intrusion Detection	15; 11
Cluster 9	Artificial Intelligence; Remote Sensing	21; 5

Figure 5 shows a co-occurrence network map of the top 50 most frequently occurring keywords, with a threshold of five co-occurrences and 246 of the most representative connections. The map shows 604 total link strengths, and the keywords are scattered into nine distinctive clusters:

Cluster 1: Big Data (occurrences: 449, total link strength: 418). This cluster is focused on the use of big data technologies, such as Hadoop, MapReduce, and Apache Spark, to analyse large datasets. The keywords in this cluster are all related to the processing and analysis of big data, and they reflect the growing importance of big data in the field of data science. Cluster 2: Deep Learning (occurrences: 117, total link strength: 144). This cluster is focused on the use of deep learning algorithms, such as convolutional neural networks and LSTMs, to solve a variety of problems, including image classification, object detection, and natural language processing. The keywords in this cluster are all related to deep learning, and they reflect the rapid growth of this field in recent years. Cluster 3: Natural Language Processing (occurrences: 60, total link strength: 73). This cluster is focused on the use of natural language processing techniques to extract meaning from text data. The keywords in this cluster are all related to natural language processing, and they reflect the increasing importance of this field in the age of big data. Cluster 4: Clustering (occurrences: 40, total link strength: 40). This cluster is focused on the use of clustering algorithms to group data points together based on their similarity. The keywords in this cluster are all related to clustering, and they reflect the importance of this technique in a variety of fields, such as data mining, machine learning, and artificial intelligence. Cluster 5: Machine Learning (occurrences: 134, total link strength: 144). This cluster is focused on the use of machine learning algorithms to learn from data and make predictions. The keywords in this cluster are all related to machine learning, and they reflect the growing importance of this field in a variety of fields, such as healthcare, finance, and transportation. Cluster 6 (occurrences: 21, total link strength: 9): This cluster is focused on machine learning algorithms that use genetic algorithms or Bayesian approaches. Cluster 7 (occurrences: 48, total link strength: 43): This cluster is focused on data mining and data analytics. Cluster 8 (occurrences: 26, total link strength: 30): This cluster is focused on class imbalance and intrusion detection. Cluster 9 (occurrences: 26, total link strength: 33): This cluster is focused on AI's diverse applications, including environmental monitoring through remote sensing. The significance of this cluster lies in the expanding role of AI in various domains beyond traditional computer science applications.

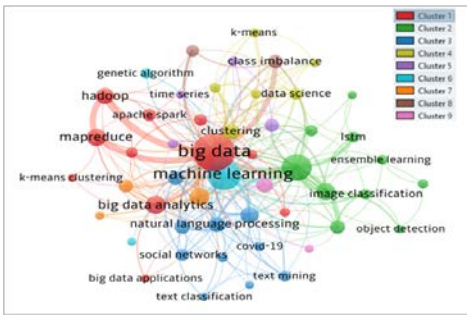

Figure 5 Top 50 Most Keywords Co-Occurred Network

Discussion and Conclusion

The comprehensive analysis of the JBD's research landscape reveals substantial insights into its developmental trends, impact, authorship dynamics, institutional contributions, and thematic trends. The journal has demonstrated remarkable growth since its establishment in 2014, publishing 788 papers over its ten active years. The increasing citation count and the presence of 46 highly cited papers, each with over 100 citations, affirm the journal's influence and contributions to the field of big data. The citation metrics, including an average of 37.13 citations per paper and 45.09 citations per cited paper, underscore the high quality and relevance of its published research.

The active year growth rate of 39.88% reflects the sustained expansion of the journal's scholarly output and recognition. The average document age of 2.83 years indicates the timely and relevant nature of the research published, aligning with the rapidly evolving landscape of big data. Collaboration emerges as a hallmark of JBD, with an average of 3.47 co-authors per paper, highlighting the interdisciplinary nature of research and the journal's role in fostering international co-authorships. Furthermore, the substantial presence of international co-authorships in 19.67% of papers underscores the global reach of JBD. The institutional analysis highlights Florida Atlantic University, USA, as a prolific contributor, exemplified by its 57 total publications, high citation count, and influential research. The collaboration network illustrates the centrality of certain authors, with Khosgoftaar TM being the most central node, reinforcing the collaborative nature of research in the field. The country-wise analysis places the USA, India, and Indonesia at the forefront, reflecting their substantial contributions and impact.

The keyword co-occurrence analysis identifies nine distinct thematic clusters, emphasizing the significance of big data technologies, deep learning, natural language processing, clustering, and machine learning. These clusters reflect the evolving trends and forefront of research within the domain of big data. The power-law distribution of citations signifies the presence of a few highly influential papers while highlighting the contribution of the broader body of work.

In conclusion, the JBD stands as a pivotal platform in the realm of big data research, characterized by its remarkable growth, substantial citations, and interdisciplinary collaborations. Its impact is evident through highly cited papers and its influential role in advancing research themes such as big data technologies, deep learning, and natural language processing. The journal continues to facilitate knowledge dissemination, foster international collaborations, and contribute significantly to the evolving landscape of big data research. As evidenced by its multifaceted analyses, this study provides valuable insights that guide future research directions and reinforce the journal's position as a cornerstone of big data analytics.

Declarations

Ethics Approval and Consent to Participate: As this work is a bibliometric analysis and does not involve human participants or any form of data collection that requires ethical approval or consent, ethics approval and consent to participate are not applicable to this study.

Consent for Publication: Consent for publication is not applicable to this study as it involves the analysis of publicly available data and does not involve any individual participants or their personal information. The study focuses solely on bibliometric analysis of existing research publications and does not require consent from authors or participants for publication.

Availability of Data and Materials: The data analysed in this study are drawn from publicly available sources, primarily the Journal of Big Data (JBD) publications and their associated citation data. These materials are accessible through the JBD's official website, databases, and repositories. Specific references to the publications, authors, and institutions are provided within the study for transparency and reproducibility.

Competing Interests: The authors declare no competing interests in conducting this bibliometric analysis of the Journal of Big Data. The study aims solely to provide an objective assessment of research trends, impact metrics, and thematic patterns within the field of big data, without any personal or financial conflicts that could influence the interpretation or reporting of the findings.

Funding: This study did not receive any external funding or specific grants. It was conducted as part of academic research to analyse and assess the research trends and impact metrics within the Journal of Big Data. The authors independently conducted the bibliometric analysis without any financial support or sponsorship.

Authors' Contributions: The author's contributions to this manuscript were collaborative and complementary. All authors were actively engaged in conceiving and designing the study, collecting and analysing the data, and interpreting the results. Additionally, each author played a crucial role in drafting and revising the manuscript for intellectual content. Their collective expertise in big data analytics, data science, machine

learning, and deep learning algorithms contributed to the thoroughness and depth of the research. The manuscript's final version reflects the input, insights, and collaborative efforts of all authors.

Acknowledgments: The authors would like to express their appreciation for the invaluable contributions of the Journal of Big Data (JBD) and its dedicated editorial team in providing the extensive dataset essential for conducting this comprehensive bibliometric analysis. Furthermore, sincere gratitude is extended to Prof. Vidyashree T, Assistant Professor, for her valuable guidance and support throughout the study. Her assistance has been instrumental in conducting thorough research and gaining insights into various novel aspects. The authors are genuinely thankful for her mentorship and expertise.

References

Akturk, Ahmet Oguz. "Thirty-five Years of the Journal of Computer Assisted Learning: A Bibliometric Overview." *Journal of Computer Assisted Learning*, vol. 38, no. 5, 2022.

Chaman Sab, M., et al. "Indian Journal of Marketing: A Bibliometric Analysis." *Indian Journal of Marketing*, vol. 50, no. 4, 2020, pp. 55-64.

Donthu, Naveen, et al. "Journal of Service Theory and Practice at age 30: Past, Present and Future Contributions to Service Research." *Journal of Service Theory and Practice*, vol. 31, no. 3, 2021, pp. 265-95.

Ellegaard, Ole, and Johan A. Wallin. "The Bibliometric Analysis of Scholarly Production: How Great is the Impact?." *Scientometrics*, vol. 105, 2015.

Goyal, Kirti, and Satish Kumar. "A Bibliometric Review of Journal of Financial Counseling and Planning between 1990 and 2022." *Journal of Financial Counseling and Planning*, vol. 34, no. 2, 2023, pp. 138-68.

Jalali, Mohsen. "A Bibiliometric Analysis of the Journal of Computers in Human Behavior: 1985-2019." *Library Philosophy and Practice*, 2020.

Kappi, Malliarjun, and B. S. Biradar. "Twenty-Six Years of Research Performance of the Journal of Optics: A Bibliometric Analysis and Future Path." *Journal of Optics*, vol. 52, 2022, pp. 77-89.

Kappi, Mallikarjun, et al. "Prominence and Impact of the 'Indian Journal of Finance' during 2013-2019 Using Scientometric Methods." *Indian Journal of Finance*, vol. 15, no. 9, 2021.

Mallikarjun Kappi, et al. "Current Science: Bibliometric Visualization based on SCOPUS Database." *Kelpro Bulletin*, vol. 24, no. 2, 2020.

Merigó, J. M., et al. "A Bibliometric Analysis of the First Twenty Years of Soft Computing." *Advances in Intelligent Systems and Computing*, vol. 642, 2017.

Modak, Nikunja, et al. "Forty Years of Computers & Chemical Engineering: A Bibliometric Analysis." *Computers & Chemical Engineering*, vol. 141, 2020.

Naveen Donthu, et al. "Forty Years of the International Journal of Information Management: A Bibliometric Analysis." *International Journal of Information Management*, vol. 57, 2021.

Prakash, Surya, et al. "A Decade of the International Journal of Lean Six Sigma: Bibliometric Overview." *International Journal of Lean Six Sigma*, vol. 13, no. 2, 2022, pp. 295-341.

Raju Vaishya, et al. "Scientometric Analysis of Indian Orthopaedic Research in the Last Two Decades." *International Orthopaedics*, vol. 46, 2022.

Ravichandra Rao, I. K., and K. S. Raghavan. "COLLNET Journal of Scientometrics and Information Management (2007-2013): An Analysis." *Collnet Journal of Scientometrics and Information Management*, vol. 9, no. 1, 2015, pp. 15-23.

Sajovic, Irena, et al. "Bibliometric Study and Mapping of a Journal in the Field of Visualization and Computer Graphics." *COLLNET Journal of Scientometrics and Information Management*, vol. 12, no. 2, 2018, pp. 263-87.

Spurk, Daniel. "Vocational Behavior Research: Past Topics and Future Trends and Challenges." *Journal of Vocational Behavior*, vol. 126, 2021.

Vaishya, Raju, et al. "International Orthopaedics Journal: A Bibliometric Analysis during 1977-2022." *Iberoamerican Journal of Science Measurement and Communication*, vol. 3, no. 1, 2023.

Zurita, Gustavo, et al. "A Bibliometric Overview of the Journal of Network and Computer Applications between 1997 and 2019." *Journal of Network and Computer Applications*, vol. 165, 2020.

Author Details

B. Mallikarjuna, *Head, Library and Information Centre, Government First Grade College for Women, Koppal, Karnataka, India,* ***Email ID****: maskimalli@gmail.com*

Mallikarjun Kappi, *Head, Library and Information Centre, Government First Grade College, Hospet, Karnataka, India,* ***Email ID****: mkmallikarjun@gmail.com*

Salma Taj, *Assistant Professor, Department of Commerce, Government First Grade College, Jagalur, Davanagere, Karnataka, India,* ***Email ID****: salmat912@gmail.com*