
Research Output on Big Data: A Scientometric Analysis Using SCOPUS 2001 – 2025

S. Valampuri Moorthy

Research Scholar
SCSVMV University, Kanchipuram.
moorthys@gmail.com.

S.S. Devika

Assistant Librarian / Research Guide,
SCSVMV University, Kanchipuram.
ss.devika@gmail.com.

Abstract

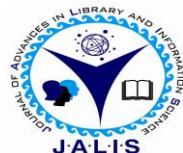
The present paper presents a scientometric analysis of research output on Big Data using Scopus 2001-2025. The data collected from the Scopus Database during 2001-2025. The search keywords "Big Data" were found in 1,15,4142 records in the study. The study focuses on the year's publications, research output, RGR, DT, document types, top authors, source titles, keywords, subject areas, source types, affiliations, languages, funding sponsors, and country-wise distribution. The data analysis uses Microsoft Excel for calculations, percentages, tables, and bar diagrams, as well as VOSviewer. The Big Data analysis's year-to-year growth trend is beginning to slow after a period of accelerating growth. The Relative Growth Rate is starting years are decreasing in order from year 2001 to 2008, it has no growth. In the coming years, the increase in 2025 is 18.66 per cent. The doubling time growth in 2003 is 0.88 and the increase in the year is 3.36.

Keywords

Big Data; Scopus Database; Relative Growth Rate; Doubling Time and Scientometric Study

Electronic access

The journal is available at www.jalis.in
DOI: 10.5281/zenodo.23210658



Journal of Advances in Library and Information Science
ISSN: 2277-2219 Vol. 15. No.4. 2026. pp.413-420

1. INTRODUCTION

Scientometrics is the study of measuring, tracking, and analyzing the quantity and quality of science, technology, and innovation using mathematical techniques. It is the "quantitative study of science, communication in science, and science policy," according to Hess's 1997 definition.

Big data is a relatively new digital phenomenon used in computer science. It is a multiple-source application used in many fields of computer systems-based software. Additionally, it improves business process efficiency, personal data tracking, performance metrics, healthcare and public health outcomes, athletic results, scientific research, security protocols, and device functionality.

2. SCOPUS DATABASE

Scopus indicates that the title, abstract, and keywords are based on the scientific abstract and one of the citation databases; the academic publisher Elsevier launched it as a competitor to Web of Science in 2004. It contains articles from 1966 to 2019, and major subjects, namely life sciences, social sciences, etc. It indexes more than 28,000 journals, 20 million author profiles, 100 million documents, more than 90,000 organization profiles, and more than 7,000 publishers.

3. REVIEW OF LITERATURE

Samuvel Raja and Sivaraman (2024) explore Big Data Analytics through the lens of scientometrics, highlighting its applications across various fields that leverage large datasets and technologies such as machine learning. Their analysis employs Bradford's and Lotka's laws to assess prolific authors and influential papers. Selvantony and Murugan (2024) examine global productivity in Polymer Chemistry research from 2013 to 2022, utilizing data from the Scopus database. Their investigation incorporates keyword analysis, annual growth rates (RGR), document types, and author rankings, revealing that China leads research output, with significant findings indicating a decreasing RGR over time and a peak in publication types in 2022.

Safiqur Rahaman and Yougal Joshi (2022) conduct a comprehensive study of Big Data publications in Library and Information Science, revealing trends from 2005 to 2022 based on 1,780 papers from the

Web of Science. Their findings highlight leading countries such as the USA and China, collaboration tendencies, and the distribution of citations. Dhawan and Gupta (2018) analyze global Big Data research output from 2007 to 2016 based on Scopus data, noting an average citation impact of 3.75 per paper, with a small fraction of highly cited works.

Lastly, Murugan and Saravanan (2017) review the "International Journal of Remote Sensing" from 2006 to 2015, focusing on publication metrics such as document types and institutional contributions. This collective research underscores significant trends and indicators across the fields of Big Data and Polymer Chemistry.

4. METHODOLOGY

The present study used the keyword "Big Data" for a search from the Scopus database from 2001 to 2025, covering 61,016 records. Data analysis utilized Microsoft Excel, bar charts, percentage analysis with tables, and VOSviewer software.

5. SCIENTOMETRIC INDICATOR

5.1. Relative Growth Rate (RGR)

According to Mahapatra, it is described as the term relative growth rate and doubling time. The growth rate increase in the number of times for publications and equation (Mahapatra, 1985) mentioned below:

$$R(1-2) \frac{W2 - W1}{T2 - T1}$$

Where,

R = Mean relative growth rate over the specific period of the interval;

W1 = log w1 (Natural log of initial number of publications);

W2 = log w2 (Natural log of initial number of publications);

T2-T1 = Unit difference between the initial time and final time.

Therefore,

R (a) = Relative growth rate per unit of publications per unit of time (year)

R (p) = Relative growth rate per unit of pages per unit of time (year)

6. OBJECTIVES

The objectives given below:

- Examine the big data research output by year.
- The Relative Growth Rate and Doubling Time
- To explain different kinds of documents
- The top ten authors publications
- Measuring the source titles
- Most preferable keywords
- Explain the top ten subjects
- Find out in various types of sources
- To depict the top affiliations
- To reveal the languages and
- Describe the funding sponsors and
- Identify the country -wise distribution

7. RESULT ANALYSIS

7.1 Growth by year on big data

Table 1 indicates the big data analysis shows growth by year in big data during the period from 2001 to 2025. Finally, 1,154,142 records are available in the data in 25 years. The initial years 2001–2008 did not show any growth or percentage. Between 2009 and 2015, there were 17,446 records (0.01–1.51%); in 2016, 28,038 (2.43%); in 2017, 40,448 (3.50%); in 2018, 54,784 (4.75%); and in 2019, 76,100 (6.59%). In the year 2020, there were 95,178 (8.25%); in 2021, 119,680 (10.37%); in 2022, 143,519 (12.44%); in 2023, 158,867 (13.76%); in 2024, 188,145 (16.30%); and in the year 2025, 215,368 (18.66%). Therefore, the maximum publication quantity is in 2025 (18.66%), while the minimum is in 2009.

Table 1. Year-wise Research Production on Big Data

Year	Records	Percentage	Year	Records	Percentage
2001	10	0.00	2014	9869	0.86
2003	12	0.00	2015	17446	1.51
2004	19	0.00	2016	28038	2.43
2005	16	0.00	2017	40448	3.50
2006	16	0.00	2018	54784	4.75
2007	30	0.00	2019	76100	6.59
2008	41	0.00	2020	95178	8.25
2009	98	0.01	2021	119680	10.37
2010	115	0.01	2022	143519	12.44
2011	273	0.02	2023	158867	13.76
2012	1298	0.11	2024	188145	16.30
2013	4772	0.41	2025	215368	18.66
Total				1,15,4142	100.00

7.2. RELATIVE GROWTH RATE AND DOUBLING TIME

Table 7.2 and Figure 1 show the Relative Growth Rate and the Doubling Time. RGR figures from 2003 to 2011 range from 0.79 to 0.57; from 2014 to 2025, RGR ranges from 0.91 to 0.21, with 2014 showing a negative growth rate of 0.91. The RGR for 2012 is 1.12, with the highest growth rate in 2013 reaching 1.25. The doubling time from 2003 to 2011, starting with 0.88 in 2004, is 1.11 for 2004 and 2005. Between 2009 and 2011, and again in 2016 and 2019, the values were below 2. In 2025, the maximum doubling time value is 3.36, while in 2013, the minimum value is 0.56.

Table2. Relative Rate of Growth and Doubling Time

Year	No. of Records	Cumulative	log W1	log W2	RGR	Doubling Time
2001	10	10	0.00	2.30	-	-
2003	12	22	2.30	3.09	0.79	0.88
2004	19	41	3.09	3.71	0.62	1.11
2005	16	57	3.71	4.04	0.33	2.10
2006	16	73	4.04	4.29	0.25	2.80
2007	30	103	4.29	4.63	0.34	2.01
2008	41	144	4.63	4.97	0.34	2.07
2009	98	242	4.97	5.49	0.52	1.33
2010	115	357	5.49	5.88	0.39	1.78
2011	273	630	5.88	6.45	0.57	1.22
2012	1298	1928	6.45	7.56	1.12	0.62
2013	4772	6700	7.56	8.81	1.25	0.56
2014	9869	16569	8.81	9.72	0.91	0.77
2015	17446	34015	9.72	10.43	0.72	0.96
2016	28038	62053	10.43	11.04	0.60	1.15
2017	40448	102501	11.04	11.54	0.50	1.38
2018	54784	157285	11.54	11.97	0.43	1.62
2019	76100	233385	11.97	12.36	0.39	1.76
2020	95178	328563	12.36	12.70	0.34	2.03
2021	119680	448243	12.70	13.01	0.31	2.23
2022	143519	591762	13.01	13.29	0.28	2.49
2023	158867	750629	13.29	13.53	0.24	2.91
2024	188145	938774	13.53	13.75	0.22	3.10
2025	215368	1154142	13.75	13.96	0.21	3.36

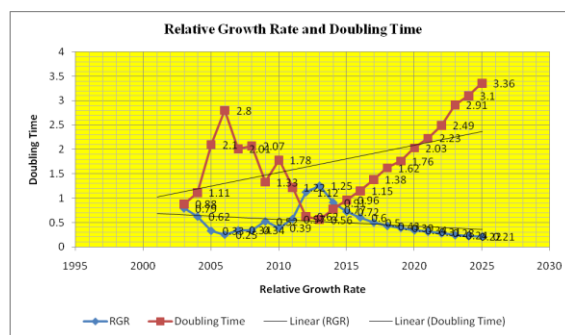


Figure.1

7.3. DOCUMENT TYPE

The published articles are of different types, namely articles, conference papers, book chapters, reviews, books, etc., described in Table 2. Out of 1,154,142 records, the maximum number of publications are from "Articles", 609,320 (52.79%), followed by "Conference Papers", 376,188 (32.59%), "Book Chapters", 72,117 (6.25%), "Reviews", 59,088 (5.12%), "Books", 10,310 (0.89%), "Editorials", 8,606 (0.75%), "Retracted", 5,150 (0.45%), "Conference Reviews", 3,915 (0.34%), "Letters", 1,839 (0.16%), "Errata", 1,498 (0.13%), "Short Surveys", 1,489 (0.13%), "Data Papers", 729 (0.06%), and "Articles in Press", 8, which is a negligible percentage.

Table 3: Document Type

S. No	Document Type	Records	Percentage
1	Article	609320	52.79
2	Conference Paper	376188	32.59
3	Book Chapter	72117	6.25
4	Review	59088	5.12
5	Book	10310	0.89
6	Editorial	8606	0.75
7	Retracted	5150	0.45
8	Conference Review	3915	0.34
9	Note	3885	0.34
10	Letter	1839	0.16
11	Erratum	1498	0.13
12	Short Survey	1489	0.13
13	Data Paper	729	0.06
14	Article in Press	8	0.00
Total		1,15,4142	100.00

7.4. TOP TEN AUTHORS' PUBLICATIONS

The top 10 authors in Big Data research are listed in Table 7.4. Out of 35,843 articles with 160 authors selected, only the top 10 authors are analyzed in the data. The majority of the articles are by “Jin, H”, 829 (18.73%), “Tian, J”, 489 (11.05%), “Choo, K.K.R”, 459 (10.37%), “Cuzzocrea, A”, 440 (9.94%), “Yu, P.S”, 409 (9.24%), “Guizani, M”, 404 (9.13%), “Kumar, N”, 361 (8.16%), “Pedrycz, W”, 351 (7.93%), “Li, T”, 345 (7.80%), and “Yang, L.T”, 338 (7.64%).

Table.4. Top ten Authors

S.No	Author Name	Records	Percentage
1	Jin, H.	829	18.73
2	Tian, J.	489	11.05
3	Choo, K.K.R.	459	10.37
4	Cuzzocrea, A.	440	9.94
5	Yu, P.S.	409	9.24
6	Guizani, M.	404	9.13
7	Kumar, N.	361	8.16
8	Pedrycz, W.	351	7.93
9	Li, T.	345	7.80
10	Yang, L.T.	338	7.64
Total		4425	100.00

7.5. TOP TEN SOURCE TITLES

The various available source titles are presented in Table 7.5 and Figure 7.2. Out of 3,25,324 source titles, the top 10 were analysed. The largest number of sources comes from the Lecture Notes in Computer Science, which ranks 1st with 18,069 records, followed by IEEE Access in 2nd with 17,595 records. The International Conference Proceeding Series (ACM) contains 15,548 records and ranks 3rd; Networks and Systems contains 10,073 records and ranks 4th; and Sustainability Switzerland contains 9,289 records and ranks 5th. Computer and Information Science ranks 6th with 8,879 records, and Applied Sciences Switzerland ranks 7th with 7,068 records. The 8th to 10th ranks are held by the Conference Series Reports.

Table 5. Top ten Source Titles

S.No	Source Titles	Record s	Ran k
1	Lecture Notes in Computer Science	18069	1

2	IEEE Access	17595	2
3	International Conference Proceeding Series of ACM	15548	3
4	Networks and Systems	10073	4
5	Sustainability Switzerland	9289	5
6	Computer and Information Science	8879	6
7	Applied Sciences Switzerland	7068	7
8	Journal of Physics Conference Series	6677	8
9	Scientific Reports	6511	9
10	Conference Series Reports	6424	10

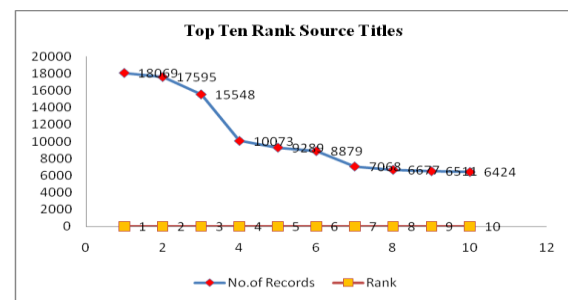


Figure.7.2

7.6. MOST PREFERABLE KEYWORDS

Table 7.2 and Figure 1 show the Relative Growth Rate and Doubling Time. The Relative Growth Rate values between 2003 and 2011 vary from 0.79 to 0.57; in the period from 2014 to 2025, the RGR spans from 0.91 to 0.21, with 2014 experiencing a negative growth rate of -0.91, followed by "Big Data" with 119,031 (13.44%), "Human" with 111,967 (12.64%), "Deep Learning" with 101,026 (11.41%), "Article" with 84,647 (9.56%), "Learning Systems" with 81,858 (9.24%), "Artificial Intelligence" with 80,401 (9.08%), "Humans" with 76,759 (8.67%), "Machine Learning" with 57,176 (6.46%), and "Internet of Things" with 50,556 (5.71%).

Table 6. Most Preferred Keywords

S.No	Keywords	No. of Records	Percentage
1	Machine Learning	122085	13.79
2	Big Data	119031	13.44
3	Human	111967	12.64
4	Deep Learning	101026	11.41
5	Article	84647	9.56
6	Learning Systems	81858	9.24
7	Artificial Intelligence	80401	9.08
8	Humans	76759	8.67

9	Machine-learning	57176	6.46
10	Internet of Things	50556	5.71
Total		885506	100.00

7.7. TOP TEN SUBJECTS

The subjects covered in Table 7.7 and Figure 7.3. Out of 23, 99, 372, twenty-seven subjects are found in the data. The top ten selected in the study is maximum of the subjects are from “Computer Science”, 6,57,236 (32.97%), followed by Engineering, 3,95,164 (19.83%), Mathematics, 1,89,162 (9.49%), Social Sciences, 1,71,241 (8.59%), Decision Sciences, 1,36,472 (6.85%), Medicine, 1,14,974 (5.77%), Business, Management and Accounting, 1,02,267 (5.13%), Physics and Astronomy, 82,871 (4.16%), Environmental Science, 73,977 (3.71%) and Energy, 69,805 (3.50%).

Table 7. Top Ten Subject Areas

S. No	Subjects	No.of records	Percentage
1	Computer Science	657236	32.97
2	Engineering	395164	19.83
3	Mathematics	189162	9.49
4	Social Sciences	171241	8.59
5	Decision Sciences	136472	6.85
6	Medicine	114974	5.77
7	Business, Management and Accounting	102267	5.13
8	Physics and Astronomy	82871	4.16
9	Environmental Science	73977	3.71
10	Energy	69805	3.50
Total		1993169	100.00

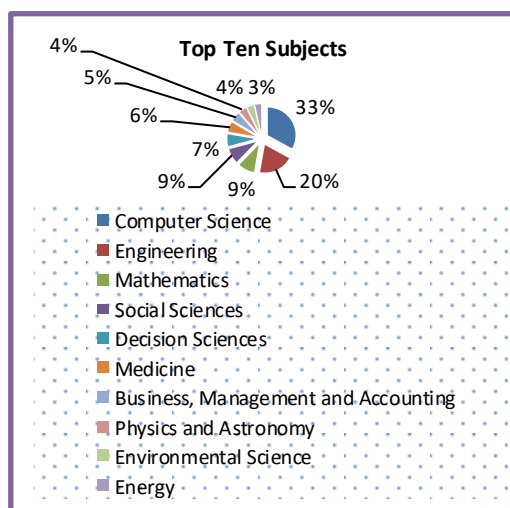


Figure.7.3

7.8. AFFILIATIONS

Table 7.8 presents an explanation of the top ten affiliations. The top ten affiliations include the “Chinese Academy of Sciences” at the top with 22,900 records (21.19%), followed by the Education, China with 20,988 (19.42%), the University of Chinese Academy of Sciences with 12,064 (11.17%), Tsinghua University with 9,662 (8.94%), Zhejiang University with 8,230 (7.62%), Beihang University with 7,776 (7.20%), Peking University with 6,894 (6.38%), Wuhan University with 6,666 (6.17%), the UEST, China with 6,534 (6.05%), and Shanghai Jiao Tong University with 6,337 (5.86%).

Table 8. Affiliations

S. No	Affiliation	No.of Records	Percentage
1	Chinese Academy of Sciences	22900	21.19
2	Education, China	20988	19.42
3	University of Chinese Academy of Sciences	12064	11.17
4	Tsinghua University	9662	8.94
5	Zhejiang University	8230	7.62
6	Beihang University	7776	7.20
7	Peking University	6894	6.38
8	Wuhan University	6666	6.17
9	UEST of China	6534	6.05
10	Shanghai Jiao Tong University	6337	5.86
Total		108051	100.00

7.9. TYPES OF SOURCES

In Table 7.9, analyse the types of sources. Out of 1,15,4142 records, 6, 93,355 (60.08%) are "Journal", which is the highest number of publications, followed by Conference Proceeding, 3,11,216 (26.97%), Book Series, 83,975 (7.28%), Book, 64,165 (5.56%), and Trade Publication, 1,431 (0.12%).

Table 9. Source Type

S.No	Types of Sources	No.of Records	Percentage
1	Journal	693355	60.08
2	Conference Proceeding	311216	26.97
3	Book Series	83975	7.28

4	Book	64165	5.56
5	Trade Publication	1431	0.12
Total		1154142	100.00

7.10. TOP TEN RANK OF LANGUAGES

Table 9 shows various articles published in different languages. The highest quantity of publications is in "English" with 1, 10, 8,345 (96.05%), followed by Chinese, 32,168 (2.79%), Spanish, 4,579 (0.40%), German, 2,003 (0.17%), Russian, 1,969 (0.17%), Portuguese, 1,557 (0.13%), French, 1,129 (0.10%), Korean, 822 (0.07%), Italian, 736 (0.06%), and Turkish, 569 (0.05%).

Table 10. Top Ten Rank of Language

S.No	Languages	No. of Records	Percentage
1	English	1108345	96.05
2	Chinese	32168	2.79
3	Spanish	4579	0.40
4	German	2003	0.17
5	Russian	1969	0.17
6	Portuguese	1557	0.13
7	French	1129	0.10
8	Korean	822	0.07
9	Italian	736	0.06
10	Turkish	569	0.05
Total		1153877	100.00

7.11. MOST PRODUCTIVE FUNDING ORGANIZATIONS

Table 11 shows the most productive funding organizations. The highest count in "Natural Science Foundation of China" 1,46,436 (48.91%) followed by R&D Program of China, 35,204(11.76%), the National Science Foundation (28279, 9.44%), Central Universities of Research Fund, 20,388 (6.81%), Horizon 2020 Framework Program (16698, 5.58%), the European Commission (15020, 5.02%), the National Institutes of Health (11612, 3.53%), the National Research Foundation of Korea (10569, 3.53%), the China Postdoctoral Science Foundation (8379, 2.80%), and the Education, China (6829, 2.28%).

Table 11.Funding Organizations

S. No	Funding Organizations	No.of Records	Percentage
1	NSF, China	146436	48.91
2	Research and Development Program of China	35204	11.76
3	National Science Foundation	28279	9.44
4	Central Universities of Research Fund	20388	6.81
5	Horizon 2020 Framework Programme	16698	5.58
6	European Commission	15020	5.02
7	National Institutes of Health	11612	3.88
8	National Research Foundation of Korea	10569	3.53
9	China Postdoctoral Science Foundation	8379	2.80
10	Education, China	6829	2.28
Total		299414	100.00

7.12. COUNTRY-WISE PUBLICATIONS

There are several published big data research tables by country (Table 7.12 for the top 10 selected). The VOSviewer software was used for the country-wise publication analysis of Big Data research. The results are as follows: the United States, 1, 85,583 (18.50%); India, 1, 37,649 (13.72%); the United Kingdom, 71,593 (7.14%); Germany, 46,715 (4.66%); Australia, 41,405 (4.13%); Italy, 38,688 (3.86%); Canada, 34,700 (3.46%); South Korea, 33,057 (3.30%); and Spain, 29,903 (2.98%).

Table 12. Country

S.No	Country	Records	Percentage
1	China	383927	38.27
2	United States	185583	18.50
3	India	137649	13.72
4	United Kingdom	71593	7.14
5	Germany	46715	4.66
6	Australia	41405	4.13
7	Italy	38688	3.86
8	Canada	34700	3.46
9	South Korea	33057	3.30
10	Spain	29903	2.98

Total	1003220	100.00
--------------	----------------	---------------

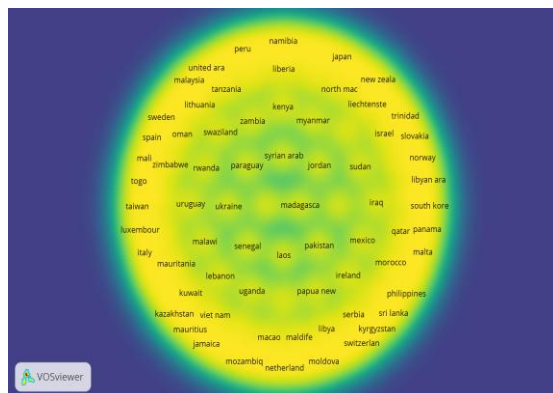


Figure.4. Country-wise Big Data Publications on VOSviewer Software

8. CONCLUSION

The study concludes the Big Data research output over the past 25 years. The year 2009 saw minimal research on Big Data, far outpaced by 18.66% in 2025. The Relative Growth Rate in 2013 was 1.25; it decreased to 0.21 in 2025. The Doubling Time in 2025 was 3.36; the growth value decreased from 0.56 in 2013. Compared to the RGR and DT, there is zigzag growths in Big Data. The most common publication type is "Article" (52.79%), followed by "Conference Paper" and "Book Chapter" (6.25%) in the top three positions. Among the top ten most productive authors, Jin, H. (7.94%) is top level and Li, T. (3.30%) is in the tenth position. Regarding the source titles selected at the top level, the first three positions are secured by Computer Science, followed by IEEE Access in 2nd rank, and the ACM Series.

Keyword search shows the maximum numbers from "Machine Learning" (13.79%), followed by "Big Data", "Human", and "Internet of Things" (5.71%). Regarding subject areas, the top ten majority are from "Computer Science" (32.97%) and "Energy" (3.50%). The majority of the affiliations are "Chinese Academy of Sciences" (21.19%), followed by the Education, China (19.42%), University of Chinese Academy of Sciences (11.17%), and Shanghai Jiao Tong University (5.86%). For source types, the majority of the published articles are in "Journals" (60.08%), followed by Book series (26.97%), Book Series (7.28%), Books (5.56%), and Trade Publications (0.12%). The highest number of publications are "English" language (96.05%), followed by Chinese (2.79%), Spanish (0.40%), and Turkish (0.05%). The majority of the funding

sponsors are the "National Natural Science Foundation of China" (11.76%) and the Education, China (2.28%). Country-wise contributions are by the United States (18.50%), India (13.72%), and Spain (2.98%).

REFERENCES

- 1) Dhawan , S.M., Gupta B.M & Ritu Gupta.(2018). Mapping Global Research Output in Big Data during 2007-16. *Emerging Science Journal*, 2(3),117-129.
- 2) Gnanasekaran, D. & Balamurugan, S. (2016). Scientometrics Profile of Global Intellectual Property Rights Research. *Journal of Information Science Theory and Practice*, 4(2), 53-65, <https://doi.org/10.1633/JISTaP.2016.4.2.4>
- 3) Mahapatra, G. (1985). Correlation between growth of publications and citations: A study based on growth curves: *Annals of Libraries and Information Studies*, 41(1), 8-12.
- 4) Murugan K. & Saravanan. G.(2017). A Scientometric Study on International Journal of Remote Sensing: 2006-2015. *Journal of Advances in Library and Information Science*, 6 (2),139-145
- 5) Ouariach, S., Ouariach, F.Z., Khan, I.U., Latrache, F., Jbira, N., & Meena, V. (Eds.). (2026). Instructional Design in E-Learning: Principles, Trends, and Opportunities (1st ed.). CRC Press. <https://doi.org/10.1201/9781003683643>
- 6) Ramesh,S and Murugan.K & Anjaline.J. (2022). Research Productivity on Big Data during 2017 – 2021: A Bibliometric Study, Knowledge Librarianship: A Data-Driven Approach, Ess Ess Publications New Delhi, pp.87-97. <http://jalis.in/pdf/6-2/Murugan.pdf>
- 7) Safiqur Rahaman MD & Yougal Joshi.(2022).Big Data Research Publications in Library and Information Science: A Scientometric Analysis. *Kelpro Bulletin*, 26 (1), 62-79.
- 8) Samuvel Raja, D & Sivaraman. P. (2024) Big Data Analytics Research – A Scientometric View, *Library Progress International*, 44 No. 3, 867-876.
- 9) Santha Kumar, R & Kaliyaperumal, K. (2015).A scientometric analysis of mobile technology Publications. *Springer Nature Link*, 105, 921 - 939.
- 10) Selvantony. A, Murugan.K.(2024). Global Level Scientific Productivity on Polymer Chemistry

- Research During 2013 -2022 .*Dharmadoot*, 101(1),144-152.
- 11) Shanmugam Sachithanantham, & Selvaraju Raja (2015). "Scientometric analysis of rabies research Literature in India: 1950 – 2014, *Springer Nature Link*, 105, 567-575.
 - 12) Shukla, R. (2019). Indian research output on genetic disorder publication using the Scopus database : A scientometric study. *COLLNET Journal of Scientometrics and Information Management*, 13(1),91–102.
<https://doi.org/10.1080/09737766.2018.1550044>.