

# Evolution of Big Data Research in different Disciplines: A Scientometric Study

S. Valampuri Moorthy<sup>1</sup>; Dr. S. S. Devika<sup>2</sup>

Research Scholar, SCSVMV University, Kanchipuram<sup>1</sup>; Assistant Librarian/Research Guide,  
SCSVMV University, Kanchipuram, India<sup>2</sup>

moorthys@gmail.com; ss.devika@gmail.com

## ABSTRACT

*Big Data has become one of the most powerful research areas of the twenty-first century, revolutionising scientific research, decision-making and innovation in many academic environments. The explosion of data-intensive technologies has resulted in a sharp increase in scholarly publications in fields like computer science, health care, engineering, business, social sciences, environmental studies, education, and library and information science. This study offers a scientometric review of the evolution of Big Data research, synthesising the existing scholarly literature and examining the intellectual development of the field across disciplines. We took a systematic review approach to identify and analyse published studies on publication growth, thematic evolution, research collaboration, citation impact and emerging research trends. The literature review shows that Big Data research has evolved from simple ideas of data storage and distributed computing to sophisticated applications based on artificial intelligence, machine learning, cloud computing, Internet of Things and data-driven decision support. The synthesis further highlights increasing interdisciplinary collaboration, the growing importance of visualization techniques in scientometric analysis, and the expanding influence of Big Data in addressing complex societal and scientific challenges. Although there are important advances, the literature points to gaps regarding interdisciplinary integration, standardisation of methodology and comprehensive mapping of emerging research themes. The research contributes by providing an integrated perspective on the development of Big Data research and identifying future directions for researchers, policymakers, and research organisations. The results offer valuable insights into the changing structure of global Big Data research, and reinforce the usefulness of scientometric approaches to understand the dynamics of fast-developing knowledge landscapes.*

**KEYWORDS:** Big Data, Bibliometric Analysis, Scientometrics, Interdisciplinary Research, Knowledge Mapping, Research Collaboration, Thematic Evolution.

## 1. INTRODUCTION

The growth of digital technologies to levels never seen before has changed how data are created, gathered, handled and used in academia, industry and government. The explosion of internet-enabled devices, cloud computing platforms, social media, sensors and digital information systems has generated massive amounts of structured, semi-

structured and unstructured data continuously. This is known as Big Data. With the growing availability of such data, research is moving from traditional data processing methods to sophisticated analytics techniques that can extract meaningful knowledge from complex and high volume datasets. Therefore, the Big Data has become a multidisciplinary research area that facilitates innovation, evidence-based decision-making, and scientific discovery in a variety of disciplines. In terms of research, the early efforts were focused on overcoming the technical challenges of storing, processing, and managing large-scale data sets through distributed computing frameworks and scalable database technologies. With the increase in computational power, the scope of Big Data has expanded broadly including artificial intelligence, machine learning, cloud computing, data mining, deep learning, the Internet of Things (IoT), blockchain, and predictive analytics. These technological advances have given researchers the opportunity to address complex issues in healthcare, business, engineering, environmental science, agriculture, education, finance, transportation, and library and information science. At present, Big Data is not only considered as a technological innovation but also a strategic resource to improve the quality of research, organization performance and social development. The rapid growth of Big Data research has also brought about an important change in the way of scientific communication. Thousands of research articles, conference papers, review studies and book chapters have been published in different disciplines, reflecting the increasing importance of data-intensive research methodologies. This exponential growth in scientific output has created a need for systematic approaches to understand the intellectual structure, research productivity, collaboration networks, influential publications and emerging thematic areas in the field. Scientometric analysis has become an effective quantitative approach to explore such aspects through the application of bibliometric indicators and knowledge-mapping techniques to evaluate the evolution of scientific research.

The reviewed literature demonstrates that previous studies have primarily focused on publication trends, citation performance, collaborative networks, keyword analysis, and thematic mapping within specific disciplines or limited time periods. The synthesis of the studies included in the attached literature review further indicates that Big Data research has experienced substantial interdisciplinary expansion, with increasing collaboration among researchers from computer science, medicine, engineering, management, environmental sciences, social sciences, and information science. Dynamic nature of the field is reflected in prominent research directions such as artificial intelligence, smart healthcare, cybersecurity, sustainable development, digital transformation and data-driven decision making.

Despite these valuable contributions, the existing literature reveals several limitations. Most of the scientometric studies have been limited to a single discipline, a specific geographical region or a bibliographic database, and thus, only reflect a part of the general development of Big Data research. Relatively few studies have addressed the development of Big Data in various academic disciplines in a comprehensive manner, integrating thematic development, interdisciplinary collaboration and emerging research fronts into one framework. These gaps highlight the need for a more comprehensive scientometric perspective that can illustrate how Big Data has grown into a multidisciplinary research ecosystem. Therefore, the current study aims at offering a comprehensive scientometric overview of the evolution of Big Data research in various disciplines through synthesis of existing scientific literature. The aim of this paper is to explore the major publication trends, thematic development, collaborative patterns, influential research areas, and future directions characterizing the growth of Big Data as a global research

domain. By synthesizing the findings from the literature review, this paper contributes to a clearer understanding of the intellectual development of the Big Data research field and serves as a valuable reference for researchers, policy-makers, librarians, and academic institutions seeking to understand the changing landscape of data-driven science.

## **2. LITERATURE REVIEW METHODOLOGY:**

This study adopts a systematic literature review methodology to examine the evolution of Big Data research across academic disciplines through the synthesis of published scholarly literature. The methodology was designed to identify, screen, evaluate, and integrate relevant studies that investigate the growth, intellectual structure, thematic development, and scientometric characteristics of Big Data research. Rather than analyzing a single application domain, the review focuses on understanding the multidisciplinary expansion of Big Data and the research trends reported in previous scientometric investigations.

### **2.1 Research Design**

The study employs a qualitative descriptive research design based on systematic literature synthesis. The review-based approach was chosen for its ability to integrate results of different scientometric studies published in different disciplines. The methodology is based on the identification of common research themes, methodological approaches, emerging technologies, collaboration patterns, and knowledge gaps reported in previous investigations. This approach provides a comprehensive understanding of how Big Data research has evolved over time and across disciplinary boundaries.

### **2.2 Data Sources**

The review is based exclusively on secondary sources of scholarly material. The main foundation for the analysis is the structured literature review data set provided for this study, comprising peer-reviewed journal articles on Big Data, scientometrics, bibliometrics, knowledge mapping and interdisciplinary research. Additional supporting literature was considered to put into context the evolution of Big Data research and its use in different scientific domains.

The examined studies are articles from several disciplines, such as:

- Computer Sciences
- Information Science
- Health care and medicine
- Engineering
- Business & Management
- Environmental Science
- Education
- Social Sciences

Together, these studies provide data on the rise of publications, citation analysis, collaboration networks, topic progression, and rising research trends on Big Data.

### **2.3 Search Strategy**

We employed a structured search method to locate studies relevant to the aims of the review. The literature included in the collection was picked by using combinations of keywords related to Big Data and scientometric study. The average search queries were:

- Big Data
- Big Data Analysis
- Scientometric Study
- Bibliometric Analysis
- Research Trends
- Knowledge Mapping
- Citation Studies
- Analysis of Co-Authorship
- Co-citation analysis

Keywords Co-occurrence Inter-disciplinary Research During literature retrieving, Boolean operators (AND, OR, and NOT) were used to enhance the relevancy and precision of search results. 3.4 Inclusion and exclusion criteria Explicit inclusion and exclusion criteria were set to ensure quality and relevancy of the reviewed literature.

### **2.5 Literature Analysis**

Thematic synthesis approach was used to analyse the selected literature. The findings from the reviewed studies were classified into major analytical categories to identify research patterns and nascent developments. The analysis covered research gaps and future directions The synthesized findings provide an in-depth analysis of the intellectual growth of Big Data research, and its evolution into a multidisciplinary field, driven by technological innovation and collaborative scientific inquiry.

## **3. LITERATURE REVIEW SYNTHESISED**

The literature of Big Data shows an astonishing evolution from a technology-oriented research domain to a multidisciplinary field affecting scientific research, industrial innovation and public policy. The collective review of the studies reveals that Big Data has moved beyond the traditional notions of data storage and processing into a strategic asset for the creation of knowledge, predictive analysis and evidence-based decision making. From the synthesis of the selected literature, five major themes emerge that characterise the evolution of big data research across the disciplines.

### **3.1 Evolution and Development of Big Data Research**

The reviewed studies report a consistent and rapid growth of publications on Big Data in the last two decades. Much of the prior work has mainly concentrated on the computer architectures of distributed computing, the technologies for data storage, database management systems, and the issues with the processing of datasets with high volume. The development of digital technologies has increased the interest of researchers to the applications of Big Data for solving complicated scientific and social problems. The literature reveals that the sudden growth of publications is caused by the development of cloud computing, high-performance computing, artificial intelligence, and digital transformation initiatives all over the world.

### **Evolution of Big Data Research:**

| <b>Research Phase</b> | <b>Major Focus</b>        | <b>Key Technologies</b>                   | <b>Research Characteristics</b>   |
|-----------------------|---------------------------|-------------------------------------------|-----------------------------------|
| Early Stage           | Data Storage & Processing | Hadoop, Distributed Computing             | Infrastructure development        |
| Growth Stage          | Data Analytics            | Cloud Computing, Data Mining              | Data-driven decision making       |
| Current Stage         | Intelligent Systems       | AI, Machine Learning, IoT                 | Interdisciplinary applications    |
| Emerging Stage        | Smart Technologies        | Generative AI, Blockchain, Explainable AI | Sustainable and ethical analytics |

Scientometric studies also indicate that the development of Big Data research has many stages of development. The first phase was technology infrastructure and the second phase moved into data analytics and intelligent systems. More recent studies point to a direction toward transdisciplinary applications, where Big Data can stimulate innovation in healthcare, education, environmental monitoring, finance, agriculture and public administration. This trend reflects the rising maturity of the discipline and its importance to many areas of research and practice.

### **3.2 Interdisciplinary Development of Big Data**

A notable result arising from the studied literature is the interdisciplinary character of Big Data research. Although computer science remains the primary contributor, substantial development has happened in several other areas. Healthcare researchers apply Big Data analytics for disease prediction, personalized medication, medical imaging, and healthcare administration. Business and management studies apply predictive analytics to better customer behavior analysis, strategic planning, and organizational performance. Environmental scientists integrate Big Data with remote sensing and geographic information systems to monitor climate change, biodiversity, and natural resource management.

### **Major Disciplines Applying Big Data:**

| <b>Discipline</b>             | <b>Major Applications</b>                 |
|-------------------------------|-------------------------------------------|
| Computer Science              | Algorithms, Data Mining, AI               |
| Healthcare                    | Disease Prediction, Personalized Medicine |
| Engineering                   | Smart Manufacturing, Automation           |
| Business                      | Customer Analytics, Decision Support      |
| Environmental Science         | Climate Monitoring                        |
| Education                     | Learning Analytics                        |
| Library & Information Science | Digital Libraries, Scientometrics         |

Engineering disciplines utilise Big Data in smart manufacturing, smart infrastructure and industrial automation. Educational scholars explore learning analytics in order to enhance teaching effectiveness and student success.

Within the discipline of library and information science, the use of Big Data in digital libraries, information retrieval, research assessment, scholarly communication, and knowledge management is increasingly studied. The literature studied thus shows that Big Data has become a common analytical framework to support interdisciplinary collaboration and innovation.

### **3.3. Emerging Research Themes and Technology Integration**

The theme development documented throughout the articles evaluated here shows that Big Data research is dynamic. Previous works mainly described topics like distributed computing, hadoop ecosystems, database technologies and data management at large scale. New study themes have evolved throughout time with the incorporation of improved computational tools.

#### **Emerging Research Themes :**

| <b>Rank</b> | <b>Theme</b>            | <b>Status</b> |
|-------------|-------------------------|---------------|
| 1           | Artificial Intelligence | Very High     |
| 2           | Machine Learning        | Very High     |
| 3           | Cloud Computing         | High          |
| 4           | Internet of Things      | High          |
| 5           | Blockchain              | Moderate      |
| 6           | Smart Healthcare        | High          |
| 7           | Cybersecurity           | High          |
| 8           | Industry 4.0            | Moderate      |
| 9           | Sustainable Development | Emerging      |
| 10          | Explainable AI          | Emerging      |

Artificial intelligence and machine learning are the engines of Big Data research, allowing for automatic pattern detection, forecasting, and intelligent decision support. Cloud computing offers scalable infrastructure to handle huge amounts of data. The Internet of Things enables continuous data generation through interconnected devices. Blockchain technology have increased data security and transparency. Algorithms using deep learning have dramatically improved analytical accuracy in recognising images, natural language processing, and healthcare diagnostics. The evaluated papers also show a rising interest in research on cybersecurity, smart cities, precision agriculture, digital healthcare, sustainable development, financial analytics and Industry 4.0. Big Data is an enabling technology that supports innovation in scientific, industrial and social realms, as demonstrated by these rising themes.

### **3.4 Scientometric Methods of Big Data Research**

The literature reflects great methodological progress in scientometric studies of Big Data. Previous research has mostly used descriptive bibliometric indicators, such as publication numbers, citation analysis and productivity measurements. Recent study is employing increasingly complex knowledge-mapping approaches to better comprehend the intellectual structure of the discipline.

### **Scientometric Techniques Used:**

| <b>Technique</b>       | <b>Purpose</b>                   |
|------------------------|----------------------------------|
| Citation Analysis      | Measure research impact          |
| Co-authorship Analysis | Study collaboration              |
| Co-citation Analysis   | Identify influential works       |
| Keyword Co-occurrence  | Discover research themes         |
| Bibliographic Coupling | Identify related studies         |
| Knowledge Mapping      | Visualize intellectual structure |

The most frequent visualisation tools applied to build co-authorship networks, co-citation maps, bibliographic coupling, keyword co-occurrence networks, and theme evolution maps include VOSviewer, CiteSpace, Bibliometrix, ScientoPy, and Gephi. These methods help researchers to identify significant authors, leading institutions, collaborating groups, rising research fronts, and conceptual connections between scientific issues. Visualisation techniques have also greatly enhanced the understanding of intricate patterns of scholarly communication in Big Data study.

### **3.5 Research Gaps and Future Directions**

Big Data research has grown a lot, yet the literature we analysed points to some major limits. Most scientometric investigations are confined to a single discipline, country or a single bibliographic database. Fewer research study the multidisciplinary evolution of Big Data under a common analytical approach, taking into account numerous scientific areas.

Another issue is the absence of consistency of methodological approaches utilised in scientometric studies. However, it is often difficult to compare among research due to differences in database selection, search methodologies and citation metrics and visualisation approaches. Moreover, relatively underexplored scientometric topics are developing study areas such as generative artificial intelligence, explainable AI, responsible data governance, ethical analytics, quantum computing and sustainable data science.

The synthesis thus emphasises the necessity for extensive and up-to-date scientometric research that may investigate publishing growth, thematic evolution, multidisciplinary collaboration and developing technologies through standardised methodology and wider data sources. Such studies will facilitate a more comprehensive understanding of the intellectual growth of Big Data research, which will benefit future scientific planning, policy formulation and multidisciplinary collaboration.

## **4. DISCUSSION AND ANALYSIS OF THE RESULTS**

The review of the selected studies reveals that Big Data research has developed rapidly in the recent two decades. At first the research was mostly about data storage, processing and distributed computing. The focus then switched to sophisticated technologies like artificial intelligence (AI), machine learning, cloud computing and Internet of Things (IoT). This development reflects the fact that Big Data is becoming a significant research field in many disciplines.

Also, literature reveals that Computer Science is the main discipline in Big Data study. But its uses have been extended to health care, business, engineering, education, environmental science, agriculture, finance, and library and information science. Researchers from different fields more and more join together to solve the real world challenges with Big Data technologies.

Another major result is that scientometric methodologies have been extensively applied to research performance evaluation. Many studies employ techniques such as citation analysis, co-authorship analysis, co-citation analysis and keyword co-occurrence analysis to discover significant authors, institutions, nations and rising research topics. Knowledge mapping and visualisation are often performed using software tools such as VOSviewer ,Bibliometrix and CiteSpace .

Also, the literature shows that the fastest research emerging subjects related to Big Data include artificial intelligence and machine learning . Other growing subjects are smart healthcare, cyber security, blockchain, smart cities, Industry 4.0, sustainable development and data-driven decision making. These discoveries suggest that Big Data is always changing and new research opportunities are arising.

However, despite this development the evaluated papers point to a number of problems. Many scientometric studies have focused on a single discipline or a single database, making it impossible to get a broad view of the global research on Big Data. Additionally, there is a need for more multidisciplinary collaboration, standardised research techniques, and ongoing updates of scientometric analyses to catch freshly developing technology and research trends.

In general, the findings indicate that Big Data has evolved from a technological notion into a multidisciplinary study area that fosters innovation, scientific discovery, and evidence-based decision-making across a wide range of academic fields.

### **Summary**

One of the fastest-growing research fields in the modern scientific world has been Big Data. The studied literature reveals that research has moved from computer science to many other areas like health care, business, engineering, education, environmental science and library and information science. The convergence of big data with artificial intelligence, machine learning, cloud computing and the Internet of Things has greatly augmented the academic effect and practical uses of big data.

The literature also suggests the utility of scientometric approaches for assessing publishing trends, collaboration patterns, key researchers and emerging research issues. These strategies allow academics and policy makers to decide where future research should go and encourage evidence based decision making.

However, the assessment points to the necessity for more thorough scientometric studies covering multiple databases, interdisciplinary views and current analytical tools. Emerging topics such as generative artificial intelligence, ethical data management, explainable AI and sustainable Big Data applications should also be the focus of future study.

To summarise, Big Data continues to transform research across disciplines fostering innovation, cooperation and knowledge production. The paper gives an overview of the history of the study and emphasises the significance of the scientometric analysis in understanding the growth and future direction of Big Data research.

## **REFERENCES & BIBLIOGRAPHY**

- [1] P. G. Acuna, A. C. J. Ellasos, & M. Devaraj (2017). Scientometric Mapping of Social Science Research in East Asia and the Philippines. *International Journal of u- & e-Service, Science and Technology*.
- [2] Baas, J., Schotten, M., Plume, A., Cote, G., Karimi, R. (2020). Scopus as a curated, high quality bibliometric data repository for academic research in quantitative science subjects. *Quantitative Science Studies*,
- [3] Bratt, S. E., Hemsley, J. J., Qin, J., and Costa, M. R. (2017). large data, large metadata and the quantitative study of science: A big scientometrics workflow model *Association for Information Science and Technology Proceedings*.
- [4] F. J. Cantu-Ortiz,. (2017). Data analytics and scientometrics: The rise of research analytics. Springer. 7.
- Chen, C., & Song, M. . (2019). A systematic scientometric review methodology: Visualising a topic of research. *PLOS ONE*, 14(10). <https://doi.org/10.1371/journal.pone.0223994>
- [5] Dhawan, S. M., Gupta, B. M., and Gupta, R. (2018). Mapping the global research output of big data over 2007-2016. *Emerging Science Journal*.
- [6] H. J. Esfahani, K. Tavasoli, and A. Jabbarzadeh. (2019). Big data and social media: A scientometric review. *International Journal of Data and Network Science*, 3(3), 145-164.
- [7] Fan, J., Han, F., Liu, H. (2014). Big data analysis challenges. *National Science Review*, 1(2), 293-314. The paper is available at: <https://doi.org/10.1093/nsr/nwt032>
- [8] Fingerma, S. Current Features and Capabilities of Web of Science and Scopus. *Problems in Science and Technology Librarianship*
- [9] Chi, P.-S. and Glänzel, W. (2020). The great problem of Scientometrics 2.0: Measuring greater impact of scientific research in public health. *Scientometrics*
- [10] Guskov, A. E. ,& Shrayberg, Y. L. . (2023). Challenges to the development of scientometric studies. *Scientific and Technical Libraries*
- [11] Heidari, H., Razavi, F., Mousakhani, M., & Gharayagh Zandi, Y. . (2017). A scientometric examination of big data literature. *Proceedings of the 5th Iranian Joint Congress of Fuzzy and Intelligent Systems (CFIS)* <https://doi.org/10.1109/CFIS.2017.8003654>
- [12] Heilig, L. und Voß, S. (2014). Scientometric examination of cloud computing literature. *IEEE Transactions on Cloud Computing*, 2(3).
- [13] Huang, X. & Chelliah, S. (2024). Big data in tourism: A bibliometric analysis (2014–2024). *Journal of Accounting, Business and Management*, 32(1).
- [14] Jensen, M. H., & Persson, J. S. (2023). Big data technologies to big data benefits. *Computer*.
- Kaswan, K. S., Baliyan, A., Dhatteval, J. S. & Prakash, O. (2023). Introduction to Big Data In *Big Data*
- [15] Kenna, R., Mryglod, O., & Berche, B. (2017). Scientometrics from a scientist's point of view: Not all that is counted counts. *Condensed Mat*.
- [16] Lundberg, L. (2023). Bibliometric analysis of research trends and orientations in big data. *Journal of Big Data*, 10(1) <https://doi.org/10.1186/s40537-023-00793-6>

- [17] Mahmood, M. N., & Dhakal, S. P. (2023). Ageing population and society: A scientometric analysis. *Quantity & Quality* 57: 3133–3150
- [18] Pan, C., Yang, W., Liang, X., & Tang, Z. . (2024). Sustainable development, big data and economy: A scientometric analysis. *Proceedings of the 4th International Conference on Economic Management and Big Data Applications*
- [19] Raban, D. & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric examination. *Scientometrics*.
- [20] Raubenheimer, J. (2021). Big data in academic research: Challenges, problems, and potential. In *Big Data in Education - Pedagogy and Research* Springer.
- [21] K. S. Rawat, S. K. Sood (2021). Emerging trends and global reach of big data analytics: A scientometric analysis. *Quality & Quantity*, 55, 1371–96.
- [22] Ruiz-Rosero, J., Ramírez-González, G. A., & Pisharody, G. . (2017). Internet of Things: A scientometric review *Symmetry*:
- [23] Raja, D. S. & Sivaraman, P. (2024). A scientometric perspective on big data analysis research. *Library Progress International*, 44(3), 867-876.
- [24] Sánchez Castillo V, Pérez Gamboa AJ, Gómez Cano CA. (2024). Trends and evolution of scientometric and bibliometric research in SCOPUS database Dialnet, 20(1).
- [25] T. Sharma. (2021). Concept, problems and research topic in big data analysis *International Journal of Research in Science & Technology*, 11(3):.
- [26] Singh, V. K., Banshal, S. K., Singhal, K. and Uddin, A. (2015). Scientometric mapping of research on large data. *Scientometrics*, 105(1), 727–741. <https://doi.org/10.1007/s11192-015-1729-9>
- [27] Srivastava, D. K. Big problems in big data research. *Data Mining and Knowledge Engineering*, 6(7), 282–286.
- Tosi, D., Kokaj, R., & Rocchetti, M. (2024). 15 years of big data: A systematic literature review. *Journal of Big Data*, 11, Article 73. <https://doi.org/10.1186/s40537-024-00914-9>
- [28] Tufekci, Zeynep. (2013). *Big Data: Pitfalls, Methods and Concepts for an Emerging Field*
- Veer, C., Veer, D. K., & Khiste, G. P. (2018). Big data output in Scopus during 2012 to 2016: A bibliometric analysis. *Knowledge Librarian*.
- [29] Wodeyar, R. (2022). Big data research productivity: A scientometric study (Doctoral dissertation, Visvesvaraya Technological University).
- [30] Wodeyar, Mulla KR. R (2022). Big data research in India: A scientometric review. *Journal of Data Science, Informetrics, and Citation Studies*.
- [31] Young, L., Wilkinson, I., & Smith, A. (2015). Scientometric analysis of publications in the *Journal of Business-to-Business Marketing* 1993-2014. *Journal of Business-to-Business Marketing*
- [32] Serrano-Pozo, A., Aldridge, G. M., & Zhang, Q. (2017). Alzheimer's disease research (1975–2014): A bibliometric and scientometric overview of four decades *Journal of Alzheimer's disease*