

Retracted Data Science Publications in India: A Bibliometric Study (2010-2024)

Madhu S*

Department of Library and Information Science, GM University, Davangere, Karnataka, INDIA.

ABSTRACT

The exponential growth of data science research has raised increasing concerns about the integrity and reliability of published findings. This study investigates the nature and trends of retracted publications authored by Indian scholars in the field of data science between 2010 and 2024. Data were obtained from the Retraction Watch Database (RWD) by applying filters for "India," "retracted status," and the period "2010-2024." Analytical processing was conducted using Microsoft Excel to examine publication patterns, document types, publishers, and institutional affiliations. A total of 920 retracted papers were identified, with the highest number observed in 2022 (148 Articles), indicating a sharp escalation in research misconduct in recent years. Research articles accounted for 92.07% of all retractions, followed by conference papers (4.67%) and review articles (1.63%). The majority of cases were full retractions (98.04%), primarily associated with publishers such as IOS Press, Hindawi, and Springer Nature. Institutionally, Anna University, Chennai (15 articles) and Jawaharlal Nehru Medical College, Wardha (9 articles) recorded the most retractions. These findings highlight the urgent need to reinforce research ethics, strengthen editorial oversight, and promote awareness of publication integrity in the domain of data science.

Keywords: Bibliometrics, Data Science, Research Integrity, Retracted Publication, Retraction Watch, Scientific Misconduct.

Correspondence:

Madhu S

Department of Library and
Information Science, GM University,
Davangere-577006, Karnataka, INDIA.
Email: madhu.rey24@gmail.com

Received: 24-12-2025;

Revised: 19-02-2026;

Accepted: 08-04-2026.

INTRODUCTION

The rapid advancement of data science has significantly reshaped the global research landscape, influencing disciplines such as engineering, healthcare, business, and the social sciences (Martini, 2016; Cao, 2017). However, this unprecedented growth has also raised serious concerns regarding research ethics, data manipulation, and the credibility of scholarly outputs (Fanelli, 2018). Retractions serve as an important corrective mechanism in scholarly communication, reflecting the self-regulatory nature of science while highlighting lapses in methodological rigor or ethical standards (Bar-Ilan and Halevi, 2017). Despite growing international attention toward retracted publications, limited empirical evidence exists on the specific patterns and causes of retractions within the field of data science, particularly in the Indian research context (Kumar and Siwach, 2024). Given India's rapidly increasing research output and its expanding global footprint in data-driven disciplines, understanding the characteristics of retracted articles is essential for strengthening research integrity frameworks. Accordingly, this study presents

a comprehensive bibliometric analysis of 920 retracted data science publications indexed between 2005 and 2024, examining their distribution by document type, year, journal, publisher, and institutional affiliation (Madhu *et al.*, 2025; Madhu and Kannappanavar, 2020). The study aims to provide a holistic overview of retraction trends in Indian data science research, offering valuable insights for policymakers, academic institutions, and publishers to promote transparency, ethical compliance, and responsible research practices.

REVIEW OF LITERATURE

Several studies have examined the patterns, causes, and consequences of retracted publications across disciplines, highlighting the growing concern over research integrity. Retractions are often considered corrective mechanisms to maintain the credibility of the scientific literature, yet their frequency and underlying reasons reveal systemic challenges in scholarly publishing. Gundur and Sampath Kumar (2025) recently studied the 2,853 retracted publications by Indian scholars between 2010 and 2024. Their findings revealed that the majority of retractions occurred during 2021-2024, accounting for 57.55% of the total. The predominant cause of retraction was fake peer review, affecting 1,007 articles. The study also highlighted that eight major publishers, including Springer, Elsevier, and Taylor and Francis, were responsible for 73.78% of all retractions Basumatary *et al.*, (2025). 79 retracted Indian



DOI: 10.5530/jcitation.20260254

Copyright Information :

Copyright Author (s) 2026 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia, [www.mstechnomedia.com]

business research articles published between 2014 and 2023 using data from the *Retraction Watch* and *Scopus* databases. The analysis revealed a sharp rise in retractions, primarily due to fake peer reviews (46.83%) and plagiarism (31.64%). Tamil Nadu recorded the highest number of retractions ($n=44$), while *Springer* accounted for 41 cases and the *Journal of Ambient Intelligence and Humanized Computing* alone contributed 17. Most retractions occurred within one year of publication, indicating increasing editorial vigilance in addressing research misconduct Kocyigit *et al.*, (2023). Analysed 91 retracted publications from Kazakhstan, with 49 research articles (53.9%) and 39 conference papers (42.9%). Retractions were primarily due to fake-biased peer review (38 articles, 41.8%), plagiarism (25 articles, 27.5%), and duplication (14 articles, 15.4%), with Russia (24 articles, 26.4%) and China (5 articles, 5.5%) as the leading collaborative countries.

Sharma (2024) examined the 3,244 retracted publications by Indian researchers, revealing that 60% originated from private institutions, with fake peer review as the leading cause. Retractions due to data integrity primarily occurred in public and medical institutions, while plagiarism was common in conference proceedings and non-Scopus journals. Domestic collaborations accounted for 80% of retractions, mostly in Q1-Q2 journals (50%), whereas international collaborations (20%) were concentrated in Q1 journals, highlighting that high-quality journals are more affected by global collaborations Amit; Kumar & Siwach, (2023). Conducted a quantitative analysis of 93 retracted publications by Indian authors in Scopus due to data falsification or fabrication, with an average retraction time of 2.62 years. The articles, primarily in Biochemistry, Genetics, and Molecular Biology, had an average of 5.11 authors and 42.1 citations, with roughly 50% being open access. Despite retraction, citation rates remained largely unchanged, highlighting persistent dissemination of flawed research and the need for stronger monitoring and ethical safeguards Fang *et al.*, (2012). Review of 2,047 retracted biomedical and life-science articles indexed in PubMed (as of May 3, 2012) found that only 21.3% were retracted due to errors, whereas 67.4% were due to misconduct, including fraud or suspected fraud (43.4%), duplicate publication (14.2%), and plagiarism (9.8%). The study also noted a tenfold increase in fraud-related retractions since 1975 and highlighted distinctive temporal and geographic patterns, indicating systemic issues in research integrity and reporting practices.

OBJECTIVES OF THE STUDY

The present study aims to undertake a bibliometric analysis of retracted publications in data science associated with India between 2010 and 2024 using data from the “Retraction Watch Database (RWD)”. The specific objectives are:

1. To examine the year-wise growth and temporal trends of retracted data science publications in India.
2. To analyze the types of documents and nature of retractions to understand the patterns and causes of withdrawal.
3. To evaluate the subject-wise and disciplinary distribution of retracted articles, highlighting the domains most affected.
4. To identify the leading journals, publishers, and institutions associated with retracted publications in data science.
5. To assess the extent of international collaboration involved in retracted works co-authored by Indian researchers.

METHODOLOGY

The present study adopted a quantitative bibliometric approach to examine retracted publications in the field of data science authored by Indian researchers. Data were obtained from the Retraction Watch Database (RWD) (<https://retractiondatabase.org/>), a comprehensive global repository of retracted scholarly articles, accessed in October 2024. Filters were applied for the country “India”, the retraction status “retracted”, and the publication period “2010-2024”. The keyword “data science” was used to retrieve relevant records. A total of 920 retracted articles meeting these criteria were identified and exported to Microsoft Excel for data cleaning, categorization, and statistical analysis. Key bibliometric variables, including document type, year of retraction, journal source, publisher, country collaboration, and institutional affiliation, were analysed using descriptive statistics and frequency distributions. The methodological framework ensured a systematic and transparent assessment of retraction trends in Indian data science research, contributing to a deeper understanding of publication ethics and research integrity.

Ethical Considerations

This study relies exclusively on secondary data available in the public domain, specifically bibliographic and retraction metadata obtained from the Retraction Watch Database (RWD). The research does not involve human participants, experiments, or the collection of personal or sensitive information. Therefore, ethical approval was not required. All procedures were conducted in accordance with accepted ethical standards for bibliometric research and the responsible use of scholarly publication data.

DATA ANALYSIS AND INTERPRETATION

Publication Types

Table 1 reflects various types of documents that have been retracted in the data science publications within India contingent. The analysis found that the highest 847 articles (92.07%) were retracted documents that came in the form of research articles. Followed by, 43 papers (4.67%) were conference abstracts/papers

and 15 (1.63%) review articles. A smaller proportion of papers were found on 9 (0.98%) Letter/Review Article, 2 (0.22%) Article in Press/Research Article and Commentary/Editorial/Letter and one each paper found (0.11%) Article in Press, and Retracted Article comment. Overall, the analysis clearly indicates that research articles overwhelmingly constitute the dominant type of retracted documents in data science publications from India, reflecting the central role of primary research outputs in instances of publication withdrawal.

Year-Wise Trend of Retracted Publications in Data Science

Figure 1 presents the year-wise trend of retracted publications in data science within India from 2005 to 2024, encompassing a total of 920 articles. During the initial years (2005-2017), retractions were minimal, remaining in single digits with only one to four articles withdrawn annually. A gradual increase was observed between 2018 and 2020, when retraction counts rose to double digits with 18 articles in 2018, 67 in 2019, and 70 in 2020 indicating growing awareness of publication ethics and data verification practices. From 2021 onward, a substantial rise occurred, with retraction figures entering triple digits with 102 articles in 2021 and peaking at 306 articles in 2022, representing the highest level of retraction activity during the study period. In the following years, a decline was noted, with 210 articles retracted in 2023 and 120 in 2024, suggesting a gradual stabilization of editorial oversight and research integrity mechanisms. Overall, the trend highlights a pronounced escalation in retraction activity in recent years, reflecting both the rapid expansion of data science research and the increasing vigilance of the Indian scholarly community toward ethical publishing standards.

Retractions Delays in Data Science Publication

Figure 2 illustrates the retraction delay patterns in data science publications. Among the 920 articles analysed, the majority 598 articles (65%) were retracted within one to two years of publication, suggesting relatively prompt corrective measures by journals and publishers. Another 207 articles (22.5%) were withdrawn after three to four years, while 104 articles (11.3%) required five to six years to be officially retracted. A small number of cases exhibited extended delays, with five articles retracted after six to seven years and one case exceeding a decade from its original publication. Notably, five articles were retracted within less than a year, indicating rapid intervention in instances of detected misconduct or errors. Overall, the findings suggest that while most retractions in data science occur within a short timeframe, demonstrating strengthened post-publication oversight, a minority still face prolonged delays, highlighting the need for more timely detection and consistent corrective mechanisms across publishers.

Co-associated Subjects with Data Science Publications

Table 2 highlights the co-associated subject areas linked with retracted publications in data science within India. The analysis indicates that retractions were not confined solely to core data science disciplines but extended across several applied and interdisciplinary domains. The highest overlap was observed with Technology (548 articles) and Computer Science (532 articles), together accounting for a substantial share of retractions, underscoring the central role of these fields in data science research output. Within the health sciences domain, notable subject associations included Radiology/Imaging (117), Oncology (89), General Medicine (60), Neurology (59), and Cardiovascular Medicine (37), reflecting the growing integration of data-driven approaches in biomedical and clinical research. Similarly, retractions were also recorded in Electrical Engineering (73), Public Health and Safety (35), General Engineering (35), and Agriculture (30), indicating the multidisciplinary nature of data science applications. Other areas such as Infectious Diseases (29), Biostatistics/Epidemiology (25), Obstetrics and Gynaecology (25), Environmental Sciences (24), Psychology (24), and Business and Economics (23) also appeared among retracted works, showing the field's broad reach. Additionally, 75 retractions were distributed across miscellaneous subjects, representing diverse and emerging intersections of data science. Overall, the findings suggest that retractions in data science are predominantly concentrated in technology-oriented and health-related disciplines, emphasizing both the field's interdisciplinary expansion and the necessity for stronger ethical and methodological rigor across applied domains.

Leading Journals with Retracted Publications in Data Science

Table 3 presents the journal-wise distribution of retracted publications in data science within India from 2005 to 2024. The analysis reveals that retractions are highly concentrated in a few journals, with the *Journal of Intelligent and Fuzzy Systems* accounting for the largest share, comprising 330 articles

Table 1: Retracted Document Types.

Document Type	Total	Percentage
Research Article	847	92.07%
Conference Abstract/Paper	43	4.67%
Review Article	15	1.63%
Letter/Review Article	9	0.98%
Article in Press/Research Article	2	0.225
Commentary/Editorial/Letter	2	0.22%
Article in Press	1	0.11%
Retracted Article	1	0.11%
Grand Total	920	100.00%

(35.87%) of the total. This is followed by *Soft Computing* with 182 articles (19.78%), together representing more than half of all retractions in the dataset. Other journals with a notable number of retractions include *Computational Intelligence and Neuroscience* (39; 4.24%), *BioMed Research International* (37; 4.02%), and *Journal of Healthcare Engineering* (29; 3.15%). A moderate number of retractions were also found in *Security and Communication Networks* (25; 2.72%) and *Computational and Mathematical Methods in Medicine* (16; 1.74%). The remaining journals each contributed less than 2% of total retractions, while 92 other journals collectively accounted for 137 articles (14.89%), reflecting the dispersed yet diverse nature of retractions across publication venues. Overall, the data indicate that a small number of journals dominate retraction activity in data science, which may be attributed to higher publication volumes, greater scrutiny, or the presence of compromised peer-review practices in certain outlets.

Leading Publishers with Retracted Publications in Data Science

Table 4 illustrates the publisher-wise distribution of retracted data science publications originating from India between 2005 and 2024. The findings show that retractions were predominantly concentrated among a few major publishers. IOS Press (now acquired by SAGE publications in November 2023) accounted for the highest number of retractions with 330 articles (35.87%), indicating a significant share of the total. This was followed closely by Hindawi with 208 articles (22.61%) and Springer-Nature Publishing Group with 207 articles (22.50%), together representing over three-fourths of all retracted articles. Springer alone contributed 61 articles (6.63%), while IEEE (Institute of Electrical and Electronics Engineers) accounted for 27 articles (2.93%), and SAGE Publications (exclusive of IOS Press acquisitions) for 24 articles (2.61%). Other major publishers such as IOP Publishing (1.96%), Elsevier (1.63%), and Wiley (1.30%) showed relatively lower retraction counts. The remaining

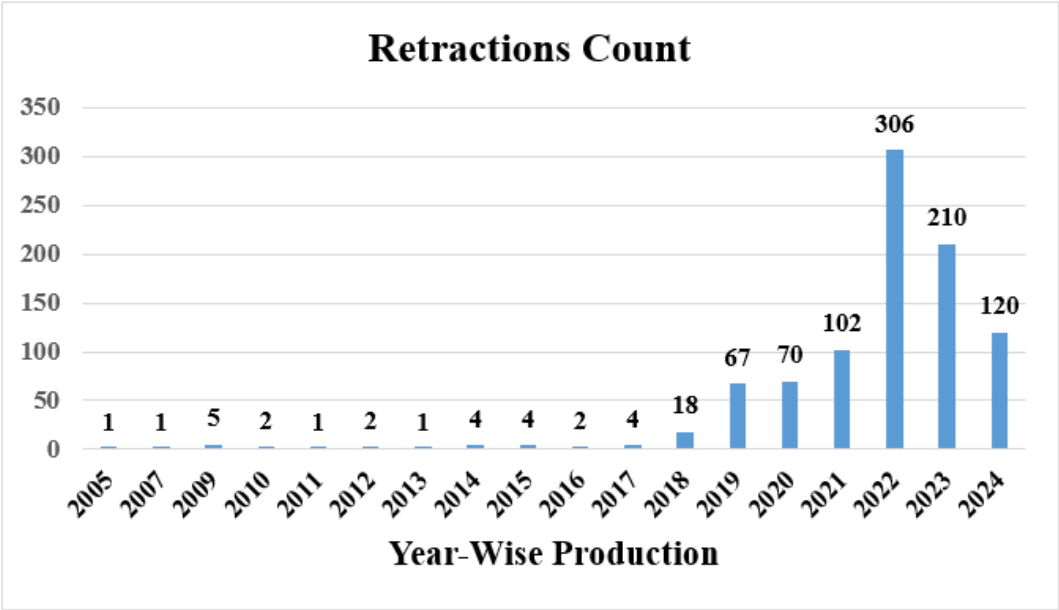


Figure 1: Year-Wise Trajectory of Retraction.

Table 2: Co-associated Subjects in Retracted Articles.

Sl. No.	Co-associated Subject	Total	Sl. No.	Co-associated Subject	Total
1	Technology	548	11	Engineering - General	35
2	Computer Science	532	12	Agriculture	30
3	Radiology/Imaging	117	13	Medicine - Infectious Disease	29
4	Medicine - Oncology	89	14	Energy	27
5	Engineering - Electrical	73	15	Biostatistics/Epidemiology	25
6	Medicine - General	60	16	Medicine - Obstetrics/Gynaecology	25
7	Medicine - Neurology	59	17	Environmental Sciences	24
8	Communications	38	18	Psychology	24
9	Medicine - Cardiovascular	37	19	Business - Economics	23
10	Public Health and Safety	35	20	Other subjects 75	452

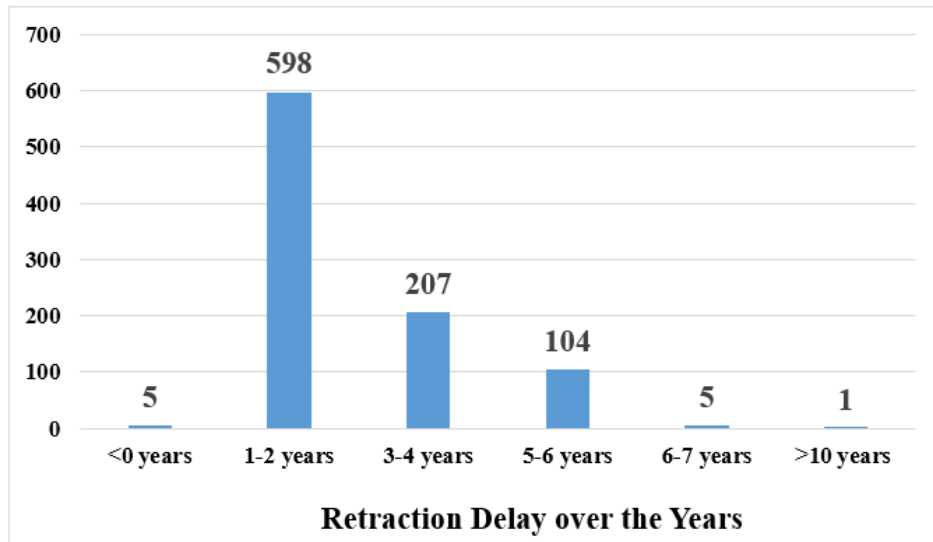


Figure 2: Retraction Delays.

Table 3: Journal-wise Frequency of Retracted Articles.

Sl. No.	Journal	Total	Percentage
1	Journal of Intelligent and Fuzzy Systems	330	35.87%
2	Soft Computing	182	19.78%
3	Computational Intelligence and Neuroscience	39	4.24%
4	BioMed Research International	37	4.02%
5	Journal of Healthcare Engineering	29	3.15%
6	Security and Communication Networks	25	2.72%
7	Computational and Mathematical Methods in Medicine	16	1.74%
8	Annals of Operations Research	15	1.63%
9	The International Journal of Electrical Engineering and Education	15	1.63%
10	Journal of Food Quality	13	1.41%
11	IOP Conference Series: Materials Science and Engineering	12	1.30%
12	Journal of Sensors	12	1.30%
13	Neurosurgical Review	11	1.20%
14	Applied Nanoscience	9	0.98%
15	Concurrent Engineering	9	0.98%
16	Advances in Materials Science and Engineering	8	0.87%
17	Wireless Communications and Mobile Computing	8	0.87%
18	The Journal of Supercomputing	7	0.76%
19	Neural Computing and Applications	6	0.65%
20	92 Journal	137	14.89%
Total		920	100.00%

publishers, including Taylor and Francis, MDPI, PLoS, Frontiers, and others, together constituted less than 2% of total retractions. Overall, the pattern highlights that the majority of retracted data science articles were concentrated within a small group of high-volume open-access and hybrid publishers, suggesting issues related to peer review integrity, publication pressure, or

compromised editorial practices within specific publication ecosystems.

Global Collaboration Patterns in Retracted Data Science Publications

Table 5 presents the country-wise distribution of co-authored retracted publications in data science involving Indian authors

during 2005-2024. The data indicate that India had the highest collaborative retraction cases with Ethiopia and Saudi Arabia, each accounting for 109 articles, reflecting strong research collaboration links between these nations. China (43 articles) and the United States (37 articles) also featured prominently, suggesting notable international engagement in data science research with Indian counterparts. Moderate levels of collaboration were observed with Iraq (24), Ghana (21), South Korea (20), and the United Kingdom (19). A smaller number of retracted articles were

associated with countries such as Bangladesh and Pakistan (17 each), Malaysia (16), and Oman (14). Additionally, collaborations with 52 other countries collectively contributed 142 retracted publications, highlighting the global nature of India's data science research network. Overall, the findings underscore that retractions are not confined to domestic collaborations but also extend across a diverse international landscape, emphasizing the shared responsibility of maintaining research integrity in global scientific partnerships.

Table 4: Publisher-wise Frequency of Retracted Articles.

Sl. No.	Publisher	TP	%TP
1	IOS Press (bought by Sage November 2023)	330	35.87
2	Hindawi	208	22.61
3	Springer - Nature Publishing Group	207	22.50
4	Springer	61	6.63
5	IEEE: Institute of Electrical and Electronics Engineers	27	2.93
6	SAGE Publications	24	2.61
7	IOP Publishing	18	1.96
8	Elsevier	15	1.63
9	Wiley	12	1.30
10	Taylor and Francis	4	0.43
11	MDPI	3	0.33
12	PLoS	3	0.33
13	Frontiers	2	0.22
14	AIMS Press	1	0.11
15	Elsevier - Cell Press	1	0.11
16	Emerald	1	0.11
17	SAE International	1	0.11
18	Scientific Research Publishing (SCIRP)	1	0.11
19	SPIE	1	0.11
	Grand Total	920	100.00

TP: Total Articles; %TP: Total Percentage.

Table 5: Country-wise Distribution in Co-authored Retracted Articles.

Sl. No.	Country	TP	Sl. No.	Country	TP
1	Ethiopia	109	11	Malaysia	16
2	Saudi Arabia	109	12	Oman	14
3	China	43	13	Egypt	13
4	United States	37	14	Turkey	13
5	Iraq	24	15	Norway	10
6	Ghana	21	16	Nepal	9
7	South Korea	20	17	Peru	9
8	United Kingdom	19	18	Kuwait	7
9	Bangladesh	17	19	Russia	7
10	Pakistan	17	20	Other 52 Countries	142

TP: Total Articles.

Nature of Retracted Publications in Data Science

Table 6 presents the nature of retraction observed in data science publications within the Indian context. Out of the total 920 retracted articles, the overwhelming majority, 902 articles (98.04%), were formally categorized as *retractions*, indicating complete withdrawal from the scholarly record due to significant ethical or methodological concerns. A smaller number of publications, 16 articles (1.74%), were marked as *expressions of concern*, typically issued as precautionary notices pending further investigation into possible research misconduct or data irregularities. Only 2 articles (0.22%) were issued as *corrections*, suggesting relatively minor issues that did not warrant full retraction. Overall, the data highlight that the dominant mode of post-publication correction in data science literature remains full retraction, underscoring the seriousness of the integrity issues leading to such editorial actions.

Quarter wise of Retracted Publications in Data Science

Table 7 highlights the Quarter wise distribution of retracted publications in data science reveals noticeable variation across different periods of the year. Out of the total 920 retractions, the highest number occurred in the second quarter (Quarter 2) with 298 articles (32.39%), followed by the first quarter (Quarter 1) with 267 articles (29.02%). The third (Quarter 3) and fourth quarters (Quarter 4) recorded relatively fewer retractions, with 179 (19.46%) and 176 (19.13%) articles, respectively. This pattern suggests that retraction activity tends to peak during the first half of the year, possibly corresponding with increased editorial reviews and post-publication audits typically conducted after annual reporting or evaluation cycles.

Table 6: Nature of Articles in Retractions.

Sl. No.	Nature of Retraction	Total Articles
1	Correction	2
2	Expression of concern	16
3	Retraction	902
Grand Total		920

Leading Institution-Wise Retracted Publications in Data Science

Table 8 highlights the institutional distribution of retracted publications in data science involving Indian affiliations. Among the listed institutions, Anna University, Chennai, recorded the highest number of retractions with 15 articles, followed by the Department of Research and Development, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, with 9 retractions. A few other prominent contributors include Sathyabama Institute of Science and Technology, Chennai, and the School of Computer Science and Engineering, Vellore Institute of Technology, each with 4 articles retracted. Institutions such as Chitkara University, Vidyasagar University, GITAM University, and SNS College of Technology each had three retractions. The remaining 853 retracted articles were distributed across 849 other affiliations, indicating that retraction incidents are widely dispersed across numerous academic and research institutions in India. This pattern suggests that while a few institutions show relatively higher concentrations of retractions possibly due to larger research output or stronger detection mechanisms of publication, integrity challenges remain a broad-based issue across the academic landscape.

DISCUSSION

The bibliometric analysis of 920 retracted data science publications reveals clear patterns in publication type, source, collaboration, and institutional contribution. The majority were research articles (847; 92.07%), followed by conference papers (43; 4.67%) and review articles (15; 1.63%), confirming that primary research outputs remain the most affected by integrity issues. In terms of the nature of retraction, 902 (98.04%) were full retractions, 16 (1.74%) expressions of concern, and only 2 (0.22%) corrections, emphasizing the predominance of serious misconduct cases. Quarter-wise analysis shows the highest retractions were found in Quarter 2 (298; 32.39%) journals, followed by Quarter 1 (267; 29.02%) journals, together contributing over 61% of total retractions. This may correspond to editorial audits and post-publication reviews typically conducted in the first half of the year. The country collaboration data indicate the strongest associations with Ethiopia (109) and Saudi Arabia (109), followed by China (43), the United States (37), and Iraq (24), underscoring the international scope of retracted works and the complexities of shared authorship accountability. Institutionally, Anna University

Table 7: Quarter wise Distribution of Retracted Articles.

Sl. No.	Quarters	Total Articles	Percentage
1	Quarter 1	267	29.02%
2	Quarter 2	298	32.39%
3	Quarter 3	179	19.46%
4	Quarter 4	176	19.13%
Grand Total		920	100%

Table 8: Institution-wise Frequency of Retracted Articles in Data Science.

Sl. No.	Institution	Total
1	Anna University, Chennai, India.	15
2	Department of Research and Development, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, India.	9
3	Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India.	4
4	School of CSE, Vellore Institute of Technology, Vellore, Tamil Nadu, India.	4
5	Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.	3
6	Department of Applied Mathematics with Oceanology and Computer Programming, Vidyasagar University, Midnapore, India.	3
7	Dept. of CSE, GITAM School of Technology, GITAM University, Bengaluru, India.	3
8	Dept. of CSE, SNS College of Technology, Tamil Nadu, Coimbatore, India.	3
9	University of Engineering and Management, Kolkata, India.	3
10	Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab, India.	2
11	College of CSE, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia.	2
12	Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India.	2
13	Dept. Mathematics, Rajalakshmi Engineering College, Thandalam, Chennai, India.	2
14	Dept. of CSE, Pimpri Chinchwad College of Engineering, Pune, India.	2
15	Dept. of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India.	2
16	Dept. of CSE, Institute of Aeronautical Engineering, Hyderabad, India.	2
17	Dept. of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.	2
18	Dept. of CSE, Lovely Professional University, Punjab, India.	2
19	Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.	2
20	Other 849 affiliations	853

(15) and Datta Meghe Institute of Higher Education and Research (9) led retraction counts, while 849 other institutions contributed 853 articles, showing widespread though uneven distribution across the Indian research landscape. At the publisher level, IOS Press (330; 35.87%), Hindawi (208; 22.61%), and Springer Nature (207; 22.50%) dominated, jointly representing over 80% of retractions, indicating systemic editorial vulnerabilities in high-volume open-access publishing. Overall, the results highlight increasing awareness and detection of unethical practices in data science research. Strengthening peer review, implementing stricter authorship verification, and promoting research ethics training are essential to ensure publication integrity and prevent future retractions.

CONCLUSION

The bibliometric analysis of 920 retracted data science publications from India (2005-2024) revealed that research articles (92.07%) dominated retraction cases, with most being full retractions (98.04%), highlighting serious ethical and methodological concerns. Retraction activity sharply increased after 2020, particularly among publishers such as IOS Press, Hindawi, and Springer Nature, indicating vulnerabilities in editorial and peer-review processes. Institutional and international collaboration patterns further suggest that issues of research integrity extend beyond individual authors or organizations. Overall, the findings emphasize the need for stronger ethical governance, transparent review mechanisms, and integrity training in data science research. Future studies should examine citation behavior before and after retraction, analyze author-level patterns, and assess the impact of institutional policies on improving publication ethics and research quality.

ACKNOWLEDGEMENT

None.

ABBREVIATIONS

RWD: Retraction Watch Database; **CSE:** Computer Science and Engineering; **CS and IT:** Computer Science and Information Technology; **IMSIU:** Imam Mohammad Ibn Saud Islamic University; **IEEE:** Institute of Electrical and Electronics Engineers; **SCIRP:** Scientific Research Publishing; **SPIE:** SPIE.

FUNDING

None.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest.

REFERENCES

- Bar-Ilan, J., and Halevi, G. (2017). Post retraction citations in context: a case study. *Scientometrics*, 113(1), 547-565. DOI: 10.1007/s11192-017-2242-0.

- Basumatary, B., Sinha, M. K., Kumar, V., and Verma, M. K. (2025). Retractions in Indian Business Research: A Closer Look at Trends, Reasons and Impacts of Retracted Articles. *DESIDOC Journal of Library and Information Technology*, 45(2), 131-140. DOI: 10.14429/djlit.45.2.20357.
- Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 60(8), 59-68. DOI: 10.1145/3015456.
- Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115(11), 2628-2631. DOI: 10.1073/pnas.1708272114.
- Fang, F. C., Steen, R. G., and Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028-17033. DOI: 10.1073/pnas.1212247109.
- Gundur, B. N., and Sampath Kumar, B. T. (2025). Trajectory of Research Retractions by Indian Scholars: Trends, Causes and Ethical Imperatives. *Journal of Data Science, Informetrics, and Citation Studies*, 4(2), 257-262. DOI: 10.5530/jcitation.20250207.
- Kocyigit, B. F., Zhaksylyk, A., Akyol, A., and Yessirkepov, M. (2023). Characteristics of Retracted Publications from Kazakhstan: An Analysis Using the Retraction Watch Database. *Journal of Korean Medical Science*, 38(46), 1-11. DOI: 10.3346/jkms.2023.38.e390.
- Kumar, A., and Siwach, A. K. (2023). Analysis of Retracted Publications by Indian Authors due to Falsification or Fabrication of Data. *Kelpro Bulletin*, 27(2), 29-44.
- Kumar, A., and Siwach, A. K. (2024). Analysis of Indian Retracted Publications: A Study Based on Scopus Data. *The Serials Librarian*, 85(7-8), 191-204. DOI: 10.1080/0361526X.2024.2306396.
- Madhu, S., and Kannappanavar, B. U. (2020). Bio-bibliometric Study of Prof. P Balaram contributions in the field of Bio-organic Chemistry and Molecular Biophysics. *Library Philosophy and Practice*, 1-15.
- Madhu, S., Kannappanavar, B. U., and Sab M, C. (2025). Scientometric Portrait of India's Renowned Chemist from Indian Institute of Science: Dr. Gautam Radhakrishna Desiraju. *Journal of Data Science, Informetrics, and Citation Studies*, 4(1), 47-53. DOI: 10.5530/jcitation.20250170.
- Martini, B. (2016). The Data Revolution. *Big Data, Open Data, Data Infrastructures and Their Consequences. Regional Studies*, 50(3). DOI: 10.1080/00343404.2015.1107987.
- Sharma, K. (2024). Over two decades of scientific misconduct in India: Retraction reasons and journal quality among inter-country and intra-country institutional collaboration. *Scientometrics*, 129(12), 7735-7757. DOI: 10.1007/s11192-024-05192-z.

Cite this article: Madhu S. Retracted Data Science Publications in India: A Bibliometric Study (2010-2024). *Journal of Data Science, Informetrics, and Citation Studies*. 2026;5(1):12-20.