Contemplating and Visualizing Plagiarism Through a Bibliometric Study

Vishal Bapte[#*] and Ritu Nagpal[&]

[#]*Sant Gadge Baba Amravati University, Amravati, Maharashtra, India*
[&]*O.P. Jindal Global University, Sonīpat-131 001, Haryana, India*
[*]*E-mail: vishal_bapte@rediffmail.com*

ABSTRACT

Plagiarism has become a buzz word in higher education sector graving as unlawful act in academic publishing. A lot of literature covering the various aspects of plagiarism and related issues has been published in the form of scientific communications. The article contemplates and visualizes the literature on plagiarism with the help of selective bibliometric parameters. The present study is focused upon concept of plagiarism with an intend to enrich the quality of research and bringing in more awareness on the topic by presenting detailed analysis on the quantum of research work from 1989-2022. During the study period, 3771 validated documents were found on the theme of plagiarism. An average citation per document is registered as 14.3. Wiwanikit, V; Rolg, M; Joob, B, and Marusic A had the highest publications on plagiarism literature. The analysis in the given study stated that USA, United Kingdom, China, Australia, and Canada have been most contributing countries in terms of the research output. Using Bradford's law, the top ten sources along with SJR value[13]from the core zone has been evaluated. The conceptual structure on theme of plagiarism is revealed through co-occurrence of keyword and thematic map. 'Plagiarism', academic dishonesty' and 'attitude' are found to be highly occurred keywords.The article written by Drummond, G. B. entitled 'Reporting ethical matters in The Journal of Physiology: standards and advice' is found to be highly cited in overall output of plagiarism literature. The stakeholder of this research would be benefited from the quantitative information on the theme of plagiarism. The list of prominent authors, core journals and multitude themes will help them to submerge in the subject of plagiarism to explore more the issue related it

Keywords: Plagiarism; Bibliometric study; R bibliometrix package; UGC's regulations 2018

1. INTRODUCTION

In this competitive digital age 'plagiarism' is a vital aspect of academic integrity. The concept of plagiarism is most often misunderstood. With the convenient availability of the digital content the technology plays a major role in the production of plagiarised content. In context to academic integrity the educational institutions give a lot of emphasis to this multifaceted notion. A well debated though sensitive issue of plagiarism deals in to probing the original work of the researchers which has been clearly analysed by Ali, Ismail, & Cheat[1] while working with the undergraduate students of University Putra Malaysia. The study cited that the concept is misinterpreted and needs a thorough understanding. Posing to the fact, with the growth of information in all fields credited to the advancement of technology[2] also calls it 'detrimental' in the educational perspectives which would lead to undue credits for the plagiarist. In a very recent study by Nundy, Kakar, and Bhutta[3], it is clearly stated that plagiarism can even be theft of someone's ideas, while at

the same time with the same objective UGC guidelines[4] also regulates a system of acquaintances and proficiency over plagiarism in the academic community. This setup would support the assessor's manual intervention by defining different levels of plagiarism and allows excluding quotations and bibliography defining item by item. UGC regulation[4] on plagiarism defines it as 'the practice of taking someone else's work or idea and passing them as one's own. Thus, to arbitrate on the given it reflects an 'acceptable percentage of similarity' for a manuscript that needs critical analysis. In Indian academic set-up, up to 10 % similarity is acceptable which is called as minor or Level 0.

2. LITERATURE REVIEW

A few studies forming similarities with the topic selected for the study have been presented as below in the form of review of literature.

Ali, Sultan and Abolmaged[5] did the bibliometric analysis of the literature on academic misconduct by using the VOS viewer software. By using the 'co-occurrence' function, the authors found five clusters consisting of

'plagiarism', 'contracting cheating', 'academic collusion', and 'scientific misconduct'. The theme of academic misconduct became vivid since 2000 with a sharp increase in publication output. The co-authorship analysis revealed sixty countries published 779 papers. The USA, the UK and Australia were the prolific countries in terms of co-authorship. There was little collaboration between developed and developing countries.

Bhakta and Bhui[6] did the bibliometric analysis of 2561 articles retrieved from WoS during 2000-2018. A continuous growth was observed in publication output with an average of 135 articles per year. Wiwanitkit V and Rolg, M was found to be the most prolific authors publishing on plagiarism. Most of the papers (1372) were single authored. The degree of collaboration was 0.46. *Nature* and *Science and Engineering Ethics* were the most relevant and productive titles which published 45 articles each. USA (981), England (628) and the Netherlands (214) published maximum documents in the area of Plagiarism.

Marques, Reis, & Gomes[7] carried out the bibliometric analysis on academic dishonesty research in two phases. The search query in the first search brought 503 articles on the themes related to plagiarism. This output was the result of the productivity of 1070 authors with an average of 2.6 authors per article. The article entitled 'Factors associated with cheating among college students: A review' by Whitley was the highly (113 citations, 22.5 %) cited article. Donald McCabe (12 articles) and Bernard Whitley (8 articles) were the most prolific authors on the theme of academic dishonesty. In the second phase of the search 829 documents were retrieved on the theme of plagiarism authored by 1798 authors with an average of 2.2 authors per article. Richard Marsh (15 articles) and Joshau D. Landau (14 articles) were the most prolific authors. The article entitled 'Text, ownership, memory and plagiarism' by Pennycooks (60 citations, 7.2 %) was the most cited paper on Plagiarism. Co-citation analysis revealed that some works on academic dishonesty was strongly associated with the theme of plagiarism.

Chauhan[8] analysed the Indian publications on plagiarism retrieved from Scopus during the period 2002 to 2016. Overall research output of 385 received 45.19 per cent citations. The annual growth rate of 38.12 per cent was registered. The single authorship dominated in terms of publications (34.55 %) with a citation rate of 50.38 %. The overall degree of collaboration was 0.65 with a mean of 0.0.48. Deepa Gupta (11) from Amrita University, Bangalore; D. K. Vani (9) from the same university, and P. Chaddah (7) of UGC-DAE Consortium of Scientific Research were found to be the most prolific authors. In all, 112 sources published 385 publications with *Current Science* (25) and *CEUR Workshop Proceeding* (Germany) being the prominent sources. However, *Lung India* has the highest (0.93) impact score.

Velmurugan and Natarajan[9] carried out the bibliometric analysis of 792 records retrieved for WoS during 2010-2014. This output was the result of the productivity of 688 research institutes. USA (25.3 %), UK (8.8 %) and Australia (5.6 %) have been the leading contributors. Among the institutes University of Hong Kong topped the list. Wiwanitkit, V (20); and Joob, B (7) had the higher research productivity as an individual author. India shared 2.1 % publications in overall research output. *Current Science, Biochemica Medica* and *Ethics and Behaviour* have published maximum documents on Plagiarism.

Garcia-Romero and Estrada-Lorenzo[10] analysed the citation pattern and the impact of the journal titles publishing the original documents. They also tried to assess the impact of the documents written by plagiarists. There was lower visibility and received less citations for plagiarism cases published in journals. It was found that full text similarity was higher in plagiarised texts. In self-plagiarism, this proportion was lower. A higher self-citation was observed for collaborative documents compared with documents produced without collaboration. The present study covers the latest time period with refined query. Here lies the significance of the study.

3. OBJECTIVES

Following are the objectives of the study:
1. To study the growth pattern of output during 1989-2022;
2. To find out the most productive authors writing on the theme of Plagiarism;
3. To find out the core journals publishing on plagiarism by using the Bradford's law of Scattering;
4. To understand the conceptual structure of plagiarism on the basis of co-occurrence of keywords and thematic map;
5. To find out highly cited document on the theme of plagiarism.

4. METHODOLOGY

The raw data for the present bibliometric study has been retrieved from Web of Science on May 20, 2022. The time span for the study was 1989 to 2022. The following search query was designed to get articles published on plagiarism.

Plagiarism (Topic) or "academic dishonesty" (Topic) or "literary theft" (Topic) or "intellectual dishonesty" (Topic) or "academic misconduct" (Topic) or "academic theft" (Topic) or "research misconduct" (Topic) or "research falsification" (Topic)

The search query resulted in 4231 records. These records were analysed with the help of 'bibliometrix[11]' package in R programming tool. However, 3771 records were authenticated in R tool. Academic integrity, contract cheating, data fabrication, data falsification, text recycling, piracy, evidence cooking, ghost-writing, proxy writing, research recycling, coping graphics are some of the words that are also associated with plagiarism. But adding of these words could have brought in many irrelevant documents. As such, the researcher built a search query which is scientifically constructed as above.

5. DATA ANALYSIS AND INTERPRETATION

5.1 Annual Scientific Production

During the period of 1989 to 2022, 3771 documents were published on plagiarism. Figure1 depicts the pattern of output in 8 blocks of 4 years each and last block of two years. A continuous growth is observed in the literature output on Plagiarism with a peak of 23.56 percent output in the block of 2017-2020. The annual percentage growth rate is registered as 3.47. The output in the last block of 2021-2022 is low because it is of one and the half years only.
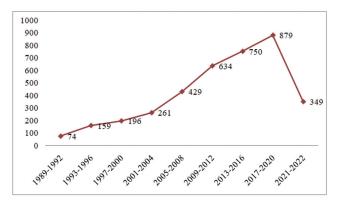
Figure 1. Scientific production.

5.2 Most Productive Authors

The total output was contributed by 6734 authors. The numbers of single authored documents were 1450 and multi-authored documents were 5284. The collaborative index is registered as 2.66. The number of documents is 3771.

Table 1 depicts 14 most productive authors along with their factionalised pattern of authorship. Wiwanitkit, V is the most prolific author with 39 documents. Rolg., M is in the second position with 16 documents. Joob, B and Marusic A are at the third and fourth position respectively with 15 documents. Fractionalised authorship, which takes into account the uniform contribution, shows some change in the list of leading prominent authors. Wiwanitkit, V who is at the first position position with score of 25.50. Rolg., M maintained his second position
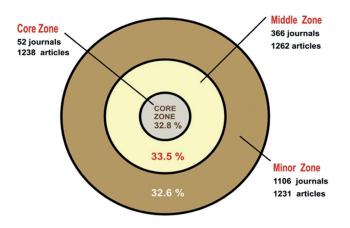
Figure 2. Source clustering through Bradford's Law.

with 12.58. Dyer, C and Dyer O are at the third (10.00) and fourth position (8.00).

5.3 Corresponding Author's Country

The leading 15 countries in terms of single country publications (SCP) and multiple country publications (MCP) are shown in table 2. USA leads the list of countries with 1215 articles in which 1131 are solely SCPs while 84 are MCPs. The MCP ratio for USA is 0.0691. However, Thailand, Iran, Germany, China, and Netherlands are the prolific countries in respect of MCP ratio.

5.4 Top 10 Core Journals Used for Publishing Using Bradford's Law

Bradford has predicted geometric distribution of periodicals in three successive zones: 1: n: n2. The observed distribution of literature on plagiarism is 52: 366: 1106 = 1524. The first zone which is considered as the core zone contained 52 journals and multiplier value 5.03 is the Bradford multiplier. The expected distribution is 52: 52 × 5.03: 52 × $(5.03)^2$ = 52:262:1316 (Rounded off to the nearest whole number) = 1630

$$\text{Percent error} = \frac{1630-1524}{1524} \times 100$$

= {106/1524}*100 = 6.95

The analysis reveals that the percentage of error is within the limit and hence Bradford's law is valid. The top ten sources along with SJR value[12] from the core zone is listed in Table 3.

The papers were published in 1524 journals published from different parts of the globe. Of these 581(15.4 %) were published in top 10 journals listed in Table 7 below along with their SJR values.[13] Figure 2 below shows different zones of Bradford's distribution.

6. CONCEPTUAL STRUCTURE

6.1 Co-occurrence of Keywords

Figure 3 depicts the network visualisation of the co-occurrence of keywords of the literature on Plagiarism. The full counting method is used with the minimum occurrence of keyword is five times. Of the 7124 keywords, 573 met the threshold. For each of the 573 keywords, the total strength of the co-occurrence links with other keywords is calculated. The keywords with the greatest total link strength have been selected. The dominant keywords are plagiarism (occurances-959; total link strength-3854), academic dishonesty (occurances-393; total link strength-2289), attitude (occurances-202; total link strength-1411), students (occurances-249; total link strength-1407), ethics (occurances-249; total link strength-1376), perception (occurances-190; total link strength-1273), research misconduct (occurances-256; total link strength-1235), cheating (occurances-172; total link strength-1103), integrity (occurances-164; total link strength-1088), misconduct (occurances-182; total link strength-

Table 1. Most productive authors

S. No.	Authors	Affiliation	Articles	Authors	Article fractionalised
1	Wiwanitkit V	Chandigarh University Mohali, Punjab, India	39	Wiwanitkit V	25.50
2	Rolg. M	McGill University, Canada	16	Rolg. M	12.58
3	Joob B	Sanitation1 Med Acad Ctr Bangkok, Thailand	15	Dyer C	10.00
4	Marusic A	University of Split Split, Croatia	15	Dyer O	8.00
5	Landau D J	Duke University Durham, Nc, USA	14	NA	8.00
6	Resnik D B	National Institutes of Health (NIH) - USA New York, NY, USA	14	Joob B	7.50
7	Marsh R L	Brown UniversityS Providence, RI, USA	12	Resnik D B	7.12
8	McCabe D L	Rutgers State University Newark, NJ, USA	12	McCabe D L	6.67
9	Dierickx K	KU Leuven Leuven, Belgium	10	Hu, G W	6.00
10	Dyer C	Scandal Hit Univ Hosp Morecambe Bay NHS Fdn Trust Kendal, England	10	Li Y Y	6.00
11	Hu G W	Wuhan University of TechnologyWuhan, Peoples R China	10	Smith R	5.92
12	Wager E	Sideview, Princes Risborough, England	10	Landau J D	5.58
13	Heidman K F	Philips Laboratory, Geophysics Directorate, Hanscom AFB, Massachusetts, USA	9	Whitley B E	5.08
14	Rauber R M	Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil	9	Marusic A	5.04

Table 2. Corresponding author's country

Country	Articles	Freq	SCP	MCP	MCP ratio
USA	1215	0.3800	1131	84	0.0691
United Kingdom	298	0.0932	253	45	0.1510
China	193	0.0604	152	41	0.2124
Australia	159	0.0497	130	29	0.1824
Canada	125	0.0391	101	24	0.1920
Spain	95	0.0297	78	17	0.1789
India	75	0.0235	70	5	0.0667
Germany	74	0.0231	58	16	0.2162
France	54	0.0169	47	7	0.1296
Iran	45	0.0141	34	11	0.2444
Netherlands	41	0.0128	33	8	0.1951
Thailand	39	0.0122	17	22	0.5641
Brazil	38	0.0119	32	6	0.1579
Turkey	38	0.0119	36	2	0.0526
Korea	36	0.0113	34	2	0.0556

Table 3. Top ten core sources

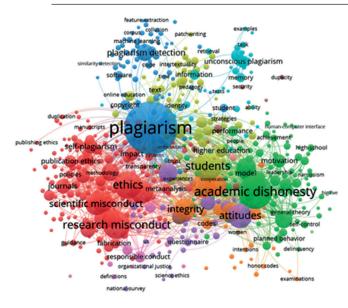| S. No.	Source	Country of origin	Articles	SJR 2021
1.	Science and Engineering Ethics	Netherlands	121	1.07
2.	Accountability in Research: Policies and Quality	United Kingdom	100	0.46
3.	Ethics & Behavior	United States	95	0.44
4.	British Medical Journal	United Kingdom	54	2.29
5.	Nature	United Kingdom	48	17.9
6.	Journal of Business Ethics	Netherlands	38	2.44
7.	Assessment & Evaluation in Higher Education	United Kingdom	37	2.03
8.	Plos One	United States	30	0.85
9.	Scientometrics	Netherlands	29	0.93
10.	Science	United States	29	14.59

Figure 3. Co-occurrence of keywords.

Figure 4. Thematic map.

10occurances-959; total link strength-385483), scientific misconduct (occurances-173; total link strength-995), and academic integrity (occurances-159; total link strength-917). The keywords like research integrity, fraud, scientific frauds, data falsification, ghost writing and cryptomnesia assures that that the query chosen for retrieving the literature on plagiarism was comprehensive and scientific one since the query designed for the study has searched all kinds of literature on it.

6.2 Thematic Map

Figure 4 reflects the thematic map which assists to understand one of the conceptual structures of the themes on plagiarism literature. The map is divided into four quadrants. The first quadrant given at the upper-right side de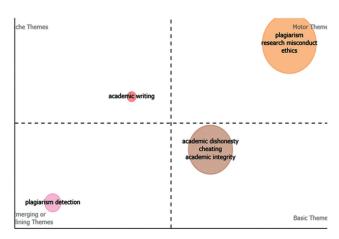notes the motor themes. These themes are highly important for researchers. The basic themes are indicated by the second quadrant and exactly given below the first quadrant. These are the established themes in any subject category. The emerging and declining themes are given in the 3rd quadrant. The highly developed themes are given in the 4th quadrant on the upper-left side. These fourth quadrants can also be highlighted in the form of four clusters namely academic writing (cluster frequency-28), Plagiarism (frequency-1684), academic dishonesty (cluster frequency–697) and Plagiarism detection (cluster frequency-55).

7. HIGHLY CITED PAPERS

Table 4 exhibits top highly cited documents on Plagiarism literature. Drummond, G.B. is highly cited author who has received 2040 citations for his title 'Reporting ethical matters in the Journal of Physiology: Standard and advice'. He is followed by Mazor, N. (2008), and Shewan, L.G. and Coats, A.J.S (2010).

Table 4. Highly cited papers

S. No.	Bibliographical details of the documents	Total citation	TC per year
1.	Drummond, G. B. (2009). Reporting ethical matters in The Journal of Physiology: standards and advice. The Journal of Physiology, 587(4), 713–719. https://doi.org/https://doi.org/10.1113/jphysiol.2008.167387	2040	45.71
2.	Mazor, N. (2008). Commentaries and Rejoinder to "The Dishonesty of Honest People." Journal of Marketing Research, 45(6), 645–653. https://doi.org/10.1509/jmkr.45.6.645	1321	145.71
3.	Shewan, L. G., & Coats, A. J. S. (2010). Ethics in the authorship and publishing of scientific articles. International Journal of Cardiology, 144(1), 1–2. https://doi.org/https://doi.org/10.1016/j.ijcard.2010.07.030	740	56.92
4.	Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLoS ONE, 4(5): e5738. https://doi.org/10.1371/journal.pone.0005738	694	49.57
5.	McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in Academic Institutions: A Decade of Research. Ethics \& Behavior, 11(3), 219–232. https://doi.org/10.1207/S15327019EB1103_2	549	24.95
6.	Drummond, G. B. (2009). Reporting ethical matters in The Journal of Physiology: standards and advice. The Journal of Physiology, 587(4), 713–719. https://doi.org/https://doi.org/10.1113/jphysiol.2008.167387	536	24.95
7.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences, 109(42), 17028–17033. https://doi.org/10.1073/pnas.1212247109	536	48.73
8.	McCabe, D. L., & Trevino, L. K. (1993). Academic Dishonesty: Honor Codes and Other Contextual Influences. The Journal of Higher Education, 64(5), 522–538. https://doi.org/10.2307/2959991	478	15.93
9.	McCabe, D. L., & Trevino, L. K. (1997). Individual and Contextual Influences on Academic Dishonesty: A Multicampus Investigation. Research in Higher Education, 38(3), 379–396. https://doi.org/10.1023/A:1024954224675	422	16.23
10.	Whitley, B. E. (1998). Factors associated with cheating among collegestudents: A Review. Research in Higher Education, 39(3), 235–274. https://doi.org/10.1023/A:1018724900565	383	15.32

8. CONCLUSION

The increasing amount of literature on plagiarism highlights the growing importance of the various issues related to it. The annual percentage growth rate is 3.47 for the publication output. Wiwanitkit, V (39), and Rolg., M (16) have been identified as the most prolific authors. Although Thailand, Iran, Germany, China and Netherlands had a few publications in comparison with USA and United Kingdom, these countries registered higher MCP ratio. The co-occurrence literature words confirmed that the query designed for retrieving the literature on the selected theme was scientific one since the retrieved records contained the literature representing different aspects of plagiarism. The thematic map showed the important sub-themes in the development of literature on plagiarism. The document written by Drummond, G. B. (2009) entiled 'Reporting ethical matters in The Journal of Physiology: standards and advice' published in the journal *The Journal of Physiology* is identified as the highly citedarticle**.** TheUGC's regulation gave a comprehensive description for responsible conduct of academic writing. It is the liability of the researcher to demonstrate the original piece of work avoiding violation.

The study assists the students and researchers to understand the growth of literature in the area of plagiarism, the most productive authors and core journals and various themes on plagiarism. This would help them to explore additional avenues in the area of plagiarism. The given work revealed that plagiarism can be termed differently but the conceptual structure is relevant in this present age of research, researchers, and research output. The work presented clearly indicates that understanding of the concept of plagiarism is vital to society and the efforts to impart the knowledge are not an alternative but the only solution to prevent it.

REFERENCES

1. Ali, W.Z.W.; Ismail, H. & Cheat, T.T. Plagiarism: to what extent it is understood*? Procedia-Social Behavio. Sci.*, 2012, **59**, 604-611. doi: 10.1016/j.sbspro.2012.09.320.
2. Franco-Salvador, M.; Rosso, P. & Montes-y-Gómez, M. A systematic study of knowledge graph analysis for cross-language plagiarism detection. *Infor. Process. Manage.*, 2016, **52**(4), 550-570. doi: 10.1016/j.ipm.2015.12.004.

3. Nundy, S.; Kakar, A. & Bhutta, Z.A. How to practice academic medicine and publish from developing countries? A practical guide. New Delhi, Springer Nature, 2022.

4. UGC Regulation 2018. Promotion of academic integrity and prevention of plagiarism in higher educational institutions regulations. https://www.ugc.ac.in/ugc_notices.aspx?id=MjA3OQ (Accessed on 15 June 2022).

5. Ali, Imran; Sultan, P. and Abolmaged, M. A bibliometric analysis of academic misconduct research in higher education: Current status and future research opportunities. *Accountability Res.*, 2021, **28**(6), 372-393.
    doi: 10.1080/08989621.2020.0836620.

6. Bhakta, J. and Bhui, T. Global Plagiarism output in the 21st century: A bibliometric study. *J. Indian Libr. Assoc.*, 2020, **56**(1), 1-9.

7. Marques, T.; Reis, N. & Gomes, J. A bibliometric study on academic dishonesty research. *J. Academic Ethics*, 2019, **17**(2), 169-191.
    doi: 10.1007/s10805-019-09328-2.

8. Chauhan, Suresh. Research on plagiarism in India during 2002-2006: A bibliometric analysis. *DJLIT*, 2018, **38**(2), 69-74.
    doi: 10.14429/djlit.38.2.12298.

9. Velmurugan, Ch. & Natarajan, R. Literature output of plagiarism: A scientometric approach thorough web of science. In Combating Plagiarism: A new role of librarian edited by M. Patra and S.K. Jena. S.K. Book Agency, New Delhi, 2015, 77-88.

10. Garcia-Romero, A. & Estrada-Lorenzo J.M. A bibliometric analysis of plagiarism and self-plagiarism thorough Dejavu. *Scientometrics*, 2014, **201**, 381-396.
    doi: 10.1007/s11192-014-1387-3.

11. Aria, M. & Cuccurullo, C. Bibliometrix: An R-tool for comprehensive science mapping analysis, *J. Informetrics*, 2017, **11**(4), 959-975.
    doi: 10.1016/j.joi.2017.08.007.

12. Scimago. 2022. Journal & Country Rank. https://www.scimagojr.com/index.php (Accessed on 20 July 2022)

CONTRIBUTORS

Mr Vishal D. Bapte is working as an Assistant Librarian in Sant Gadge Baba Amravati University, Amravati, Maharashtra (MS) since 2012. His research interests include: bibliometrics, scientometrics, collection development, library philosophy and practical applications in libraries.
In the current research work, he is responsible for conceptualizing, designing, analysis and overall writing of the paper.

Ms Ritu Nagpal is presently working as Joint Librarian with Global Library, O P Jindal Global University, India. As a Research and Outreach librarian her interest lies in Open Access, Open Source Library Automation, Digital Libraries, Data Management and also Information Literacy. She is a Review Board Member of International Journal of Advance Study and Research Work (IJASRW) and has also reviewed multiple research articles in eminent journals of high repute. She has written introductory part of this paper and assisted in technical writing of the different part of the research work.