GMR
The Original

ARTIFICIAL INTELLIGENCE ORIENTED BIBLIOMETRIC META ANALYSIS OF COVID ANALYSIS IN INDIA

Rana Tarannum Ziyauddin Shaikh[1], Dheva Rajan S[2], Gargi Tyagi[3]

[1]Research Scholar,,Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.
[2]Co-Supervisor, Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India
[3]Supervisor – Assistant Professor,Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India, rana.ansari@utas.edu.om.

ABSTRACT:
Background: SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is the source of the novel coronavirus disease, COVID-19, which remains a serious and immediate threat to global health. The virus that causes COVID-19 spreads when an infected person coughs or sneezes or when the infected person comes into contact with a healthy person. In many ways, science, technology, and artificial intelligence have been instrumental in overcoming this pandemic. **Methods:** A bibliometric, cross-sectional review of the literature on COVID-19-related papers from Web of Science journals was conducted with an emphasis on Indian research and trends. The current investigation employs bibliometric analysis, augmented by artificial intelligence approaches, to examine the extensive corpus of COVID-19-related publications from India. This study examined co-authorship maps, authors, institutions, nations, and keywords. Artificial intelligence, deep learning, and machine learning approaches have been used to analyze the COVID-19 mathematical model. The present investigation examines the publishing patterns of issues pertaining to the forecasting of COVID-19 numbers in India from 2016 to January 25, 2025, using the Web of Science (WOS) and Scopus databases. The AI approaches like random walk, clustering using Walktrap algorithm,Fruchterman-Reingold algorithm, networking distance approach has been utilized. **Findings:** The most often cited authors, their connected institutions, trends in COVID-19 publications and citations, and network visualizations of author co-authorship, co-occurrences, and co-citation of referenced sources are examined using bibliometric analysis. **Conclusion:** The findings provide valuable insights into predicting virus propagation using machine learning techniques, presenting additional opportunities for further investigation. The current study identifies key trends, notable studies, and emerging gaps in this field. The bibliometric meta-analysis yields new insights for future studies and encourages the exploration of the socio-economic impacts of the pandemic, utilizing the current results in a broader context.

KEYWORDS: COVID, prediction, Machine learning, Artificial intelligence, Mathematical model, Regression, Correlation

INTRODUCTION

The COVID-19 pandemic has highlighted how crucial scientific research is to solving world health emergencies. Even though COVID-19 started in 2019, the coronavirus is not new to the world (Cleveland, 2021). Although the coronavirus was initially discovered in the mid-1960s, the older human coronaviruses have likely been around for millennia. Alpha coronavirus 229E, beta coronavirus OC43, alpha coronavirus NL63, and beta coronavirus HKU1 are among them. "With the exception of HKU1, which may also induce a gastrointestinal illness," he says, "most of these earlier incarnations present with a moderate respiratory infection. It was stated by (UC San Diego Health, 2019) that the COVID-19 virus spread many months before it was announced in 2019. Scientists warned in 2007, "The presence of a large reservoir of SARS-CoV–like viruses in horseshoe bats… is a time bomb. The possibility of the re-emergence of SARS and other novel viruses… should not be ignored," as mentioned in (Cheng et al., 2007).

Hence, excluding articles from before 2019 is not the correct option. So, it has taken a decade of work related to covid 19 journey in India. In this regard, bibliometrics, the quantitative study of scientific literature, becomes an essential instrument. Bibliometric analysis facilitates the identification of key studies, patterns, and gaps in the vast body of COVID-related research by systematically evaluating research outputs. The present method enables politicians and medical professionals to make informed decisions based on the most significant and pertinent scientific facts, while also enhancing comprehension and accelerating discovery. As a result, bibliometric analysis is essential for negotiating and planning in the dynamic field of pandemic research.

1.1. Research Objectives

In the bibliometric analysis of COVID-19 literature from India, this study aims to integrate and analyze the use of Artificial Intelligence (AI) methods, including Machine Learning (ML) algorithms such as Support Vector

Machines (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). The objectives of the current research work are as follows:
A.Determining the most influential sources
B.Determining which authors have the most influence
C.Determining which documents have the most influence
D.Author clustering and any further methods
E.Identifying different structures, such as social, intellectual, and conceptual structures
Preprocessing and Data Analysis, Cluster analysis, network analysis, co-occurrence matrix construction, co-citation and coupling analysis, Statistical and network analysis techniques, which are fundamental to many AI applications but are specifically applied in a bibliometric context here, are used to visualize relationships between various entities in the bibliographic data, such as collaboration networks among authors or institutions.

2. REVIEW OF LITERATURE

A Bibliometric Analysis of COVID-19 across Science and Social Science Research was conducted by Aristovnik et al., specifically for these two fields. (Gautam et al., 2020) have done general work on bibliometric analysis and Insights on COVID-19, but the data collected were till April 10th, 2020, in the Scopus database. Patralekh et al., (2021) have done wonderful work on bibliometric analysis of COVID-19 in India, but it is related to publications only in orthopedic journals. An almost similar work by Wang & Wang, (2021) has been identified. However, it is based only on China, and it is only for robotic technologies, with only two research questions: one is about the role of robotics in combating COVID-19, and another one is about the promising support needed to combat COVID-19. Vaishya et al., (2023) have done a bibliometric analysis on only the top 100 highly cited papers from India on COVID-19 research. But Vaishya et al., (2023) concentrate only on core literature not specific to AI or robotics or statistical and ML models. An analysis performed by Karbasi et al., (2023) contained the word AI but in the context of Iran and with a database search until July 27, 2022. Salik Ata, (2023) has done a bibliometric analysis of Covid 19 using the Web of Science (WoS) Database, but it is towards studies published in Industry 4.0. Ameet & Rana, (2024) have done work on bibliometric analysis of COVID-19-based models of India, but the core contribution Ameet & Rana, (2024) is towards “Models,” especially between 2020-23, not specific to the current objective of AI. Singh et al., (2024) contributed to a bibliometric analysis by discussing the top 50 most-cited articles about COVID-19 and its complications. But Singh et al., (2024) concentrated on medically oriented articles that discuss respiratory illness causing pneumonia and ARDS, and its diverse complications extend to cardiovascular, gastrointestinal, renal, hematological, neurological, endocrinological, ophthalmological, hepatobiliary, and dermatological systems. Liew et al., (2024) have been discussing the 100 most-cited articles on COVID-19 using a bibliometric analysis. However, it is purely based on medical orientation, and the predominant keyword, COVID-19, is targeted only at China. Gaur et al., (2024) have done work on bibliometric analysis of COVID-19 till 12th February 2022, but Gaur et al., (2024) concentrated only on stochastic methods. A mind-blowing article has been published by Misra & Gupta, (2024), but it is about envisioning emerging research trends on the impact of COVID-19 on the Indian economy. The study by Misra & Gupta, (2024) applied latent semantic analysis and text mining techniques, which were two of the instruments used in the current study. Public and scholarly interest in social robots literature review has been performed by Mishra et al., (2024), but this is based only on robots and Google trends using the Scopus database and VoS viewer software. Hence, there is a research gap in combining ML, AI, robotics, and statistical models in bibliometric analysis and the proposed work fills the gap.

3.Data Collection, Materials and Methods

The Web of Science database is chosen to collect documents. The documents have been chosen based on the keywords under the topic covid as mandatory, with Boolean operators “and” for Prediction, Statistics, Mathematical model, Artificial intelligence, Boolean operator “or” Machine learning, ML, Regression, Correlation, SVM, KNN, ANN, AI. Upon executing the conditions mentioned, the information is collected from 11040 documents. When executing the Boolean operators with the mentioned titles, 3631774 documents were obtained. In that it is extracted only open access documents leads to 1800909 documents between the years 2016 and 2025. In finding article types it is obtained 1296311 documents in the Web of Science core collection on Science Citation Index Expanded (SCI-Expanded 1059508 documents), Social Science Citation Index (SSCI-211519 documents), Emerging Sources Citation Index (ESCI-177659 documents), Arts & Humanities Citation Index (A&HCI-3760 documents). As the country is restricted only to India, it obtained 47427 documents. The language is restricted to English, which only leads to 47418 documents. The research areas are restricted to Computer Science, Educational Research, Science and Technology, Other Topics, Information Science Library Science, Mathematics, Mathematical Methods in Social Science, Mathematical Computational Biology, and Robotics of the Web of Science. The list of categories leads to 11040 documents. Out of 11,040 documents, 3283 are in multidisciplinary, 2776 are in computer science information systems, 1963 are in engineering and electronics, 1664 computer science AI, 1373 are telecommunications, 849 are computer science interdisciplinary applications, 649 computer science theory methods, 482 Green sustainable science technology, 482 Material science Multidisciplinary, 460 Mathematics computational biology. Out of 11040 documents, 2941 documents are at the intersection of multidisciplinary and interdisciplinary sectors. The way the data screening was conducted is explained in Figure 1, the PRISMA diagram.

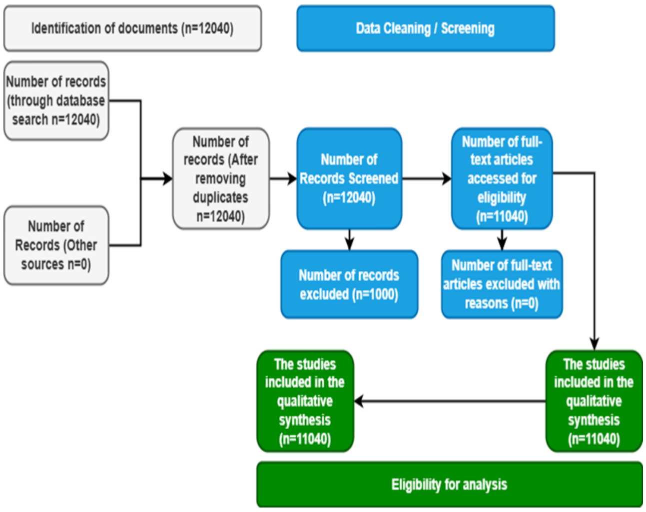

Fig. 1 PRISMA Diagram

The entire Bibliometric analysis is divided into six major parts and three subparts. For the most part, the structures are given below.

1.Overview of the data
2.Sources
3.Authors
4.Documents
5.Clustering
6.Structures
a.Conceptual Structure
b.Intellectual structure
c.Social Structures

4.ANALYSIS

4.1. Overview of the data

The overview section provides the main information, annual scientific production, average citations per year and three field plots. The details mentioned above are given in Figure 2, Main information.

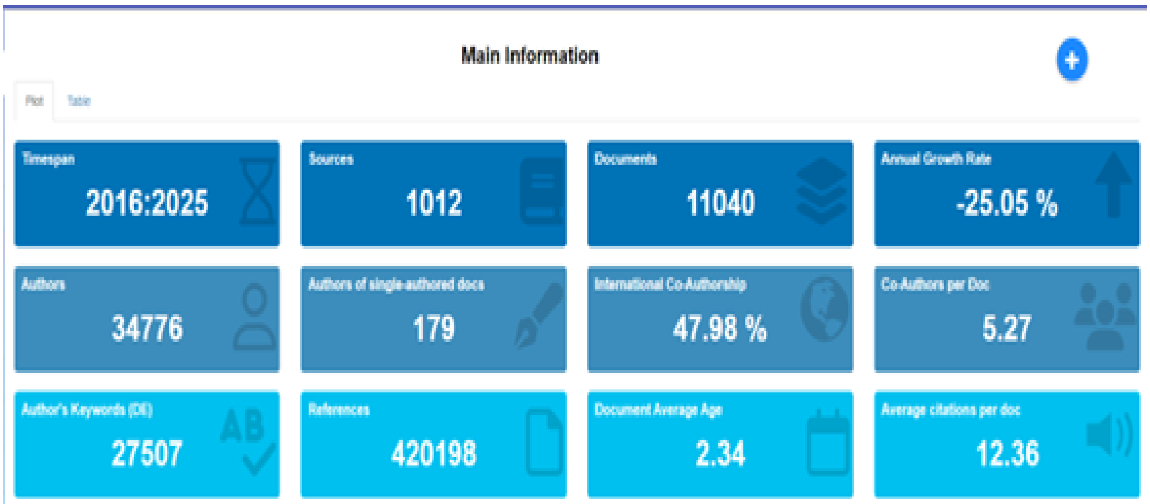

Fig. 2 Main Information

In the entire collection, 87 data papers, 30 proceedings papers, 71 retracted papers, 154 Early access papers and 11040 articles have been collected. In the collection, only the article type is chosen for further analysis. The completeness of the data from the analysis indicates that the keywords' plus 'and those stated as' poor ',' abstract ', 'corresponding author', 'cited references', and 'DOI' stated as 'good', along with all other parameters, are in an excellent state, as shown in Figure 3, Completeness of Metadata. The entire analysis in this work has been performed using R biblioshiney application of Bibliometry package using R programming (Aria & Cuccurullo, 2017) ensures the adoptability of AI in the current work.

Completeness of metadata -- **11040** docs from **Isi**

Metadata	Description	Missing Counts	Missing %	Status
C1	Affiliation	0	0.00	Excellent
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
WC	Science Categories	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
AB	Abstract	2	0.02	Good
RP	Corresponding Author	4	0.04	Good
CR	Cited References	13	0.12	Good
DI	DOI	20	0.18	Good
ID	Keywords Plus	2422	21.94	Poor
DE	Keywords	2454	22.23	Poor

Advice Report Save Close

Fig. 3 Completeness of Metadata

Figure 4 illustrates the annual scientific production of articles in the specified research area.

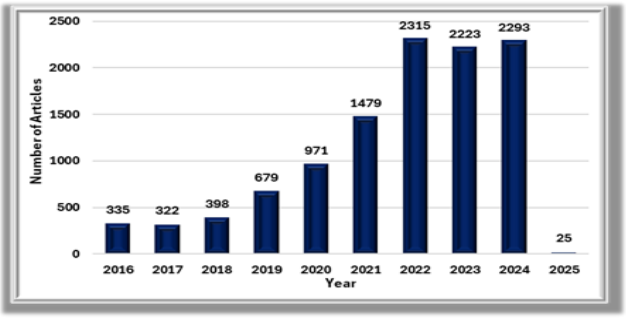

Fig. 4 Annual Scientific Production

In Figure 5, N is the total citations on the primary axis, and TC per year gives the average total citations per year denoted on the secondary axis.

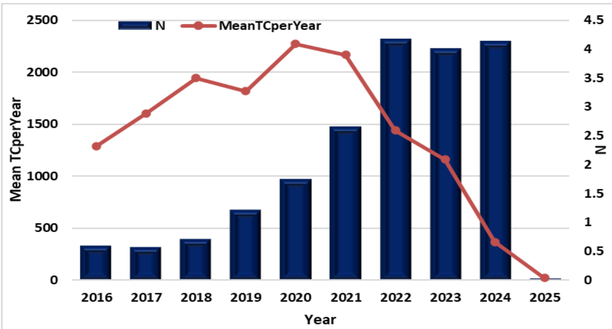

Fig. 5 Average Citations Per Year

The three field plots in Figure 6 show the connections between Cited references (CR), Authors(AU) and Keywordsplus(DR).

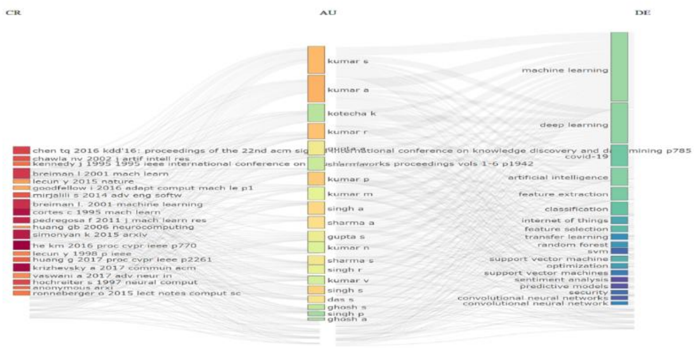

Fig. 6 Three Field Plot

4.2. Sources

The part, Sources, discusses the most relevant sources, the most locally cited sources, Bradford’s law, sources’ local impact, and sources' production over time. Figure 7 shows the most relevant sources on the said topic, with the y-axis denoting the frequency of documents on related topics. Scientific reports hold the highest position, with 1077 articles.

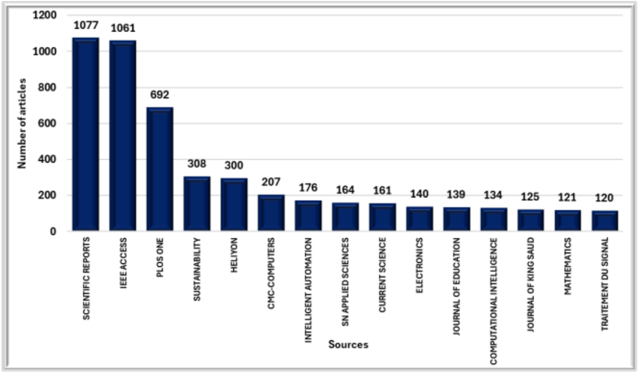

Fig. 7 Most Relevant Sources

Figure 8 shows the most locally cited sources. IEEE Access holds first place with 9211 documents, followed by ARXIV and One.

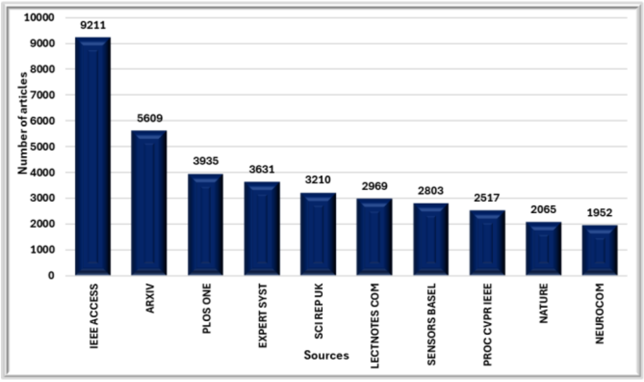

Fig. 8 Most Local Cited Sources

Bradford's Law is a principle used in bibliometrics to describe the scatter of literature in a particular field. Bradford's Law N=k ×BCn , where N is the total number of articles, k is a constant specific to each dataset or field of study, B is the base of the exponential function, which typically equals the number of core journals, C is the average number of articles in the core journals, n is the number of zones (or groups) into which the journals are divided based on the number of articles they contain. Breadford's law posits that the core of literature on a specific topic can be divided into a few key sources that are cited very frequently. As one expands the scope to include more sources, each is cited less frequently. This results in a pattern of diminishing returns in terms of the volume of literature as one moves away from the core. The shaded area labeled "Core Sources" identifies the region that contains the most significant number of articles relevant to the topic being studied. Beyond the core, there will be zones with progressively fewer relevant articles. Bradford described these as concentric zones of productivity that decrease as one moves away from the core. Hence, Bradford's law gives Scientific reports, IEEE Access, Plos One, Sustainability, Heliyon, and CMC as the core sources

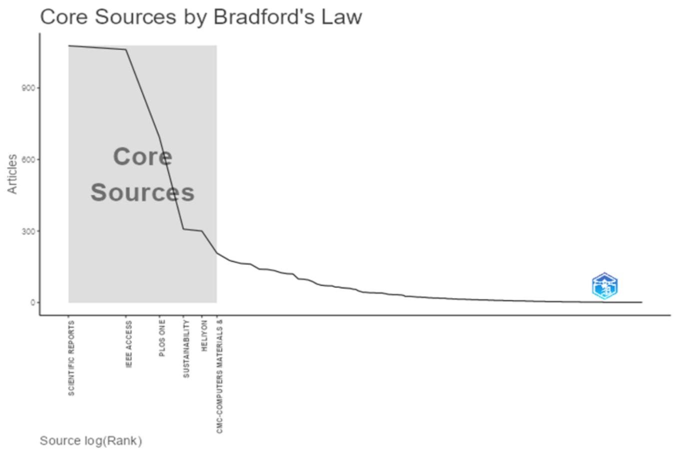

Figure 9: Bradford's Law

Figure 10 presents the local impact of the sources, including their h-index, g-index, and total citations. IEEE Access holds first place in all these aspects, with an h-index of 55, g-index of 98, and total citations per article.

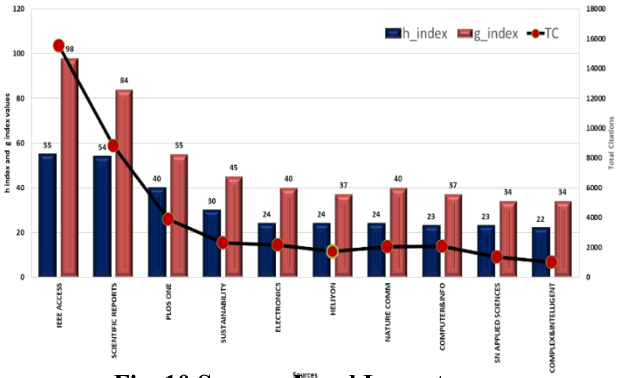

Fig. 10 Sources Local Impact

Figure 11 shows the sources' production over time, which is assessed between the years 2016 and 2025, at which point scientific reports show a steady increase. IEEE access started slowly in 2016, but it coincides with scientific reports in 2025. Heylon and sustainability have had a slow increase since 2016. Plos One will also have equal competition with scientific reports till 2021, meeting IEEE access in 2022; after that, Plos One will moderately increase.

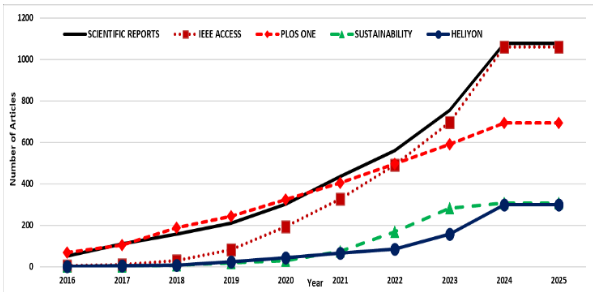

Fig. 11 Sources Production over Time

4.3. Authors

In this section, the author discusses the most relevant authors, the most locally cited authors, authors' production over time, Lotka's law, and the local impact of authors. Figure 12 lists the most relevant authors, with Kumars securing the top five places, followed by Singh and Sharma.

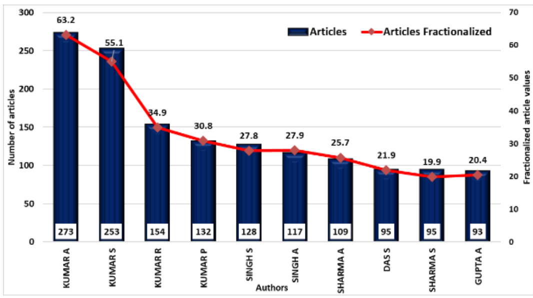

Fig. 12 Most Relevant Authors

Figure 13 gives us the most locally cited authors, with ALAZAB M holding first place with 123 citations, followed by GADEKALLU TR 109. KUMAR S holds third place with 96 citations, whereas he holds second place among the most relevant authors (as seen in Figure 12). KUMAR A acquired fifth position with 86 articles but holds first place among the most relevant authors (as seen in Figure 12).

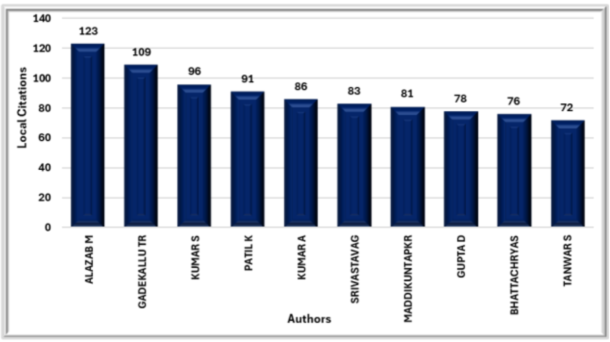

Fig. 13 Most Local Cited Authors

Figure 14 presents the authors' production overtime, as illustrated in Figure 12. There is no surprise in this as A. Kumar et al., (2024) holds his first place, followed by S. Kumar et al., (2024), R. N. Kumar et al., (2024) analyzing this, one can find that the articles written by the authors in 2024 have the highest citations, not in the earlier years.

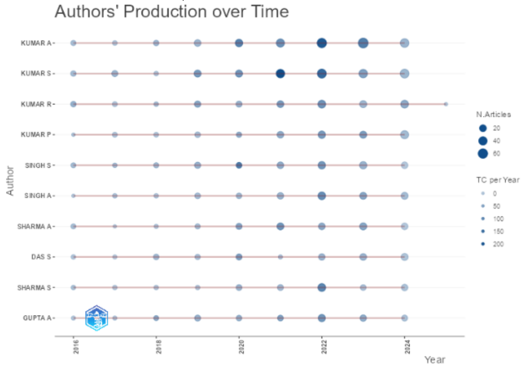

Figure 14: Authors Production Over Time

Lotkas law (named after Alfred J Lotka) is one of the varieties of Zipf's law used to describe the frequency of publications by various authors in the given field. Zipfs law states that word frequency is inversely proportional to the corresponding word rank. This Zipfs concept in mathematics is formalized in mathematics as a special probability distribution where the inverse power law is the rank of the frequency distribution. This Zipfs law is highly related to Benford's law and Pareto distribution. Lotkas law can be stated as Y' is proportional to X−k−1where Y' is the number of authors with atleast X publications, whereas Lotaks original publication he claimed only for k=2 and k varies depending on the discipline. When the number of authors is increasing, obviously the number of authors producing that many articles become very less. The author's productivity, as shown in Figure 15, is explained using Lotka's law.

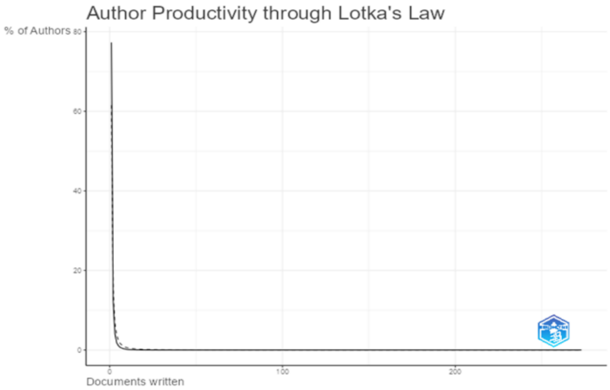

Figure 15: Lotka's Law

Figure 16 illustrates the local impact of the authors as presented in Figures 12 and 14. In this analysis, S. Kumar overtaking A. Kumar and R. Kumar. The most surprising number of authors for the article published in 2022 by A. Kumar In the Nature Journal, it is 603, with 177 citations and 44.25 total citations per year.

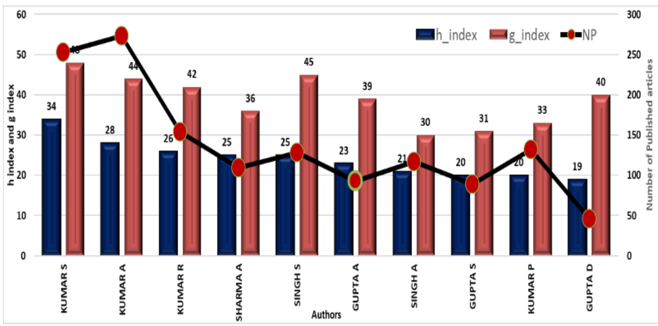

Figure 16: Authors Local Impact

Figure 17 shows the most relevant affiliation at which Indian Institute Of Technology System (IITs) holds the first place with 1559 documents, followed by the National Institute Of Technology (NITs) with 738 documents, Vellore Institute Of Technology (VITS) with 710 documents, VIT Vellore, 486 with documents, followed by Manipal Academy Of Higher Education (MAHE), Indian Council Of Agricultural Research (ICAR), and King Saud University (KSUNIV), SRM Institute Of Science And Technology Chennai (SRMC), Egyptian Knowledge Bank (EKB), Council Of Scientific And Industrial Research (CSIR). Firstly, IITs (1559) produced more than twice the number of articles as NITs (738).

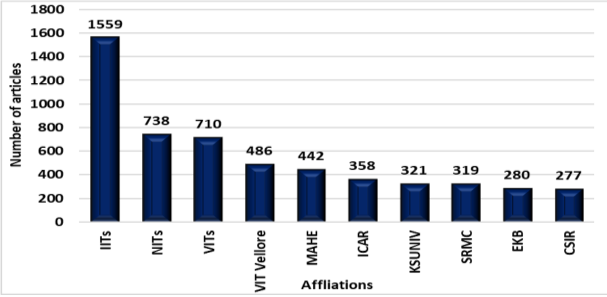

Figure 17: Most Relevant Affiliation

Figure 18 illustrates the growth of affiliations over time, during which IITs in India experienced exponential growth, followed by NITs. VIT groups and VIT Vellore are competing equally, and MAHE is in fifth place.

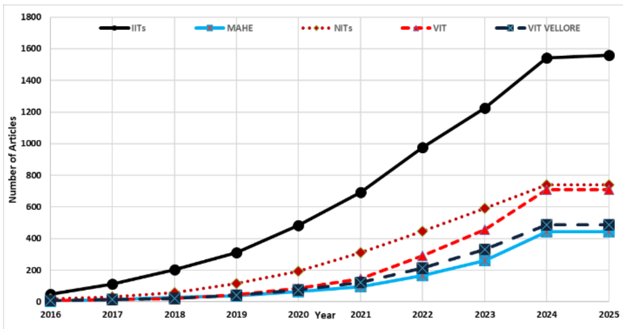

Figure 18: Affiliations' Production Over Time

Figure 19 shows the countries of the Corresponding Author. Without loss of generality, one can observe that India is at the top, followed by the USA, China, Ethiopia, the UK, Saudi, Korea, Australia, Malaysia, and Germany. Surprisingly, when discussing the published works with respect to India, China, and Korea in the order of the corresponding authors' countries, they are very interested in knowing the status of COVID-19 in India. This could be due to the collaboration of authors across the globe. Such collaboration is evident in the global heat map presented in Figure 20.

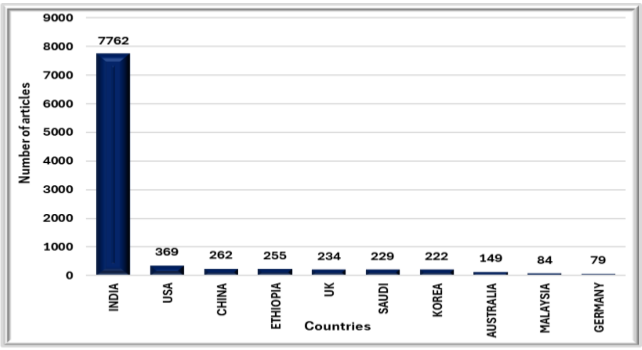

Figure 19: Corresponding Author's Countries

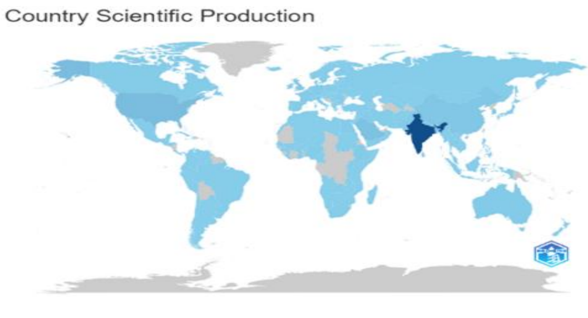

Figure 20: Countries' Scientific Production

Out of those scientific productions discussed using earlier figures 17,18,19, and 20, the country's production overtime is given in Figure 21. India secures its first place, being the primary country of analysis, followed by the USA and Saudi Arabia. The UK and China follow the Saudis in their scientific production in relation to India.

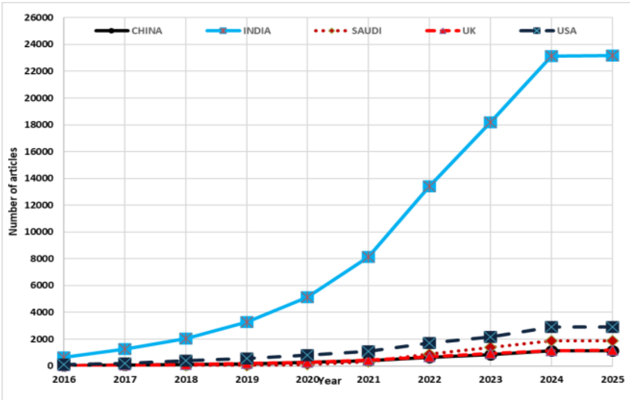

Figure 21: Countries Production Over Time

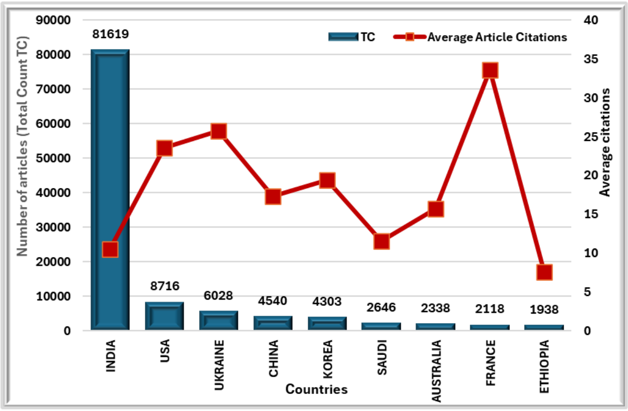

Figure 22: Most Cited Countries

4.4. Documents

In the most locally cited documents Mohan et al., (2019) on Heart Disease Prediction Using Hybrid ML Techniques hold the first place, followed by Vinayakumar et al., (2019) Deep Learning Approach for Intelligent Intrusion Detection System, Reddy et al., (2020) Dimensionality Reduction Techniques, Bhattacharya et al., (2021) Deep learning and medical image processing for coronavirus (COVID-19), Gadekallu et al., (2020) Diabetic using PCA and Deep Learning, Bhattacharya et al., (2020) PCA and Firefly, Sethy et al., (2020) COVID-19 based on Deep Learning and Support Vector Machine(SVD), Jain et al., (2021) deep learning, COVID-19, Tuli et al., (2020) COVID-19 ML, cloud computing and Bharti et al., (2021) Heart Disease, Machine Learning and Deep Learning. In these documents, most are related to medical issues, particularly those related to the heart. The most used ideas are deep learning, ML, AI and Principal Component Analysis methods.

Figure 23: Most Local Cited Documents

The most globally cited documents is given in figure 24. Kairouz et al., (2021) Advances and Open Problems in Federated Learning holds the first place followed by Dwivedi et al., (2021) challenges and opportunities in AI, Bernard et al., (2018) Deep Learning Techniques for MRI Cardiac, Dwivedi et al., (2023) With an Opinion Paper: “So What If ChatGPT Wrote It?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, Vinayakumar et al., (2019) Deep Learning Approach, Maddikunta et al., (2022) Industry 5.0, Hassija et al., (2019) Survey on IoT, Ren et al., (2022) Survey of Deep Active Learning, Saw et al., (2017) Topological defects in govern cells, Reddy et al., (2020) on Dimensionality Reduction. Most of these documents are in the fields of medicine, AI, deep learning, ML, and dimensionality reduction.

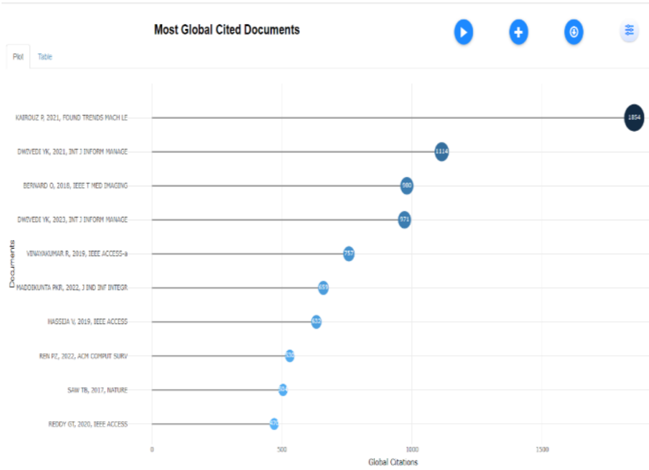

Figure 24: Most Global Cited Documents

Figure 24 gives the Most Global Cited Documents at which He et al., (2016) on Deep Residual Learning, Krizhevsky et al., (2017) ImageNet classification, Cortes & Vapnik, (1995) Support-vector networks, Simonyan & Zisserman, (2015) Very Deep Convolutional Networks, Breiman, (2001) Random Forests, Chen & Guestrin, (2016) XGBoost: A Scalable Tree Boosting System, Hochreiter & Schmidhuber, (1997) Flat Minima, Chawla et al., (2002) Synthetic Minority Over-sampling Techniques. Most of these documents are AI- and medical-oriented.

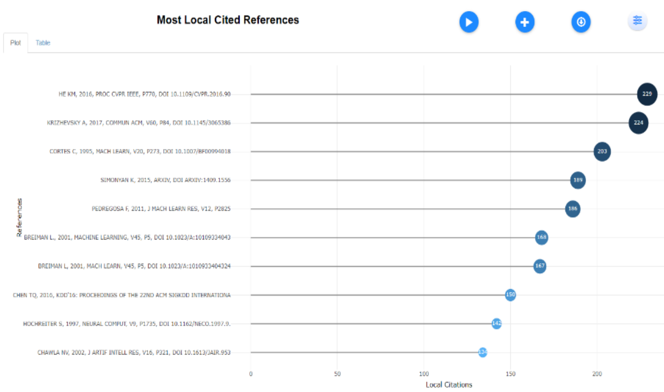

Figure 25: Most Local Cited References

Figure 26 gives the word cloud at which the words classification (657), model (471), prediction(406), Algorithm (379), system (367), performance (266), design (248), optimization (239), diagnosis (228), and impact (214).

Figure 26: Word Cloud

Figure 27 shows the frequency of the words classification, model, prediction, algorithm, system, performance, design, optimization, diagnosis, and impact over time.

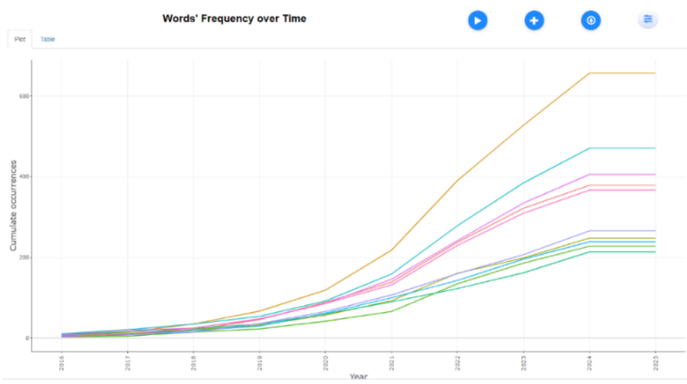

Figure 27: Words Frequency over time

Figure 28 shows the trending topics at which the topics with keywords classification, model, prediction, management, expression, growth, challenges, temperature, cancer, and AI hold its top positions in trending.

Figure 28: Trending Topics

Figure 29 shows the tree map of the keywords utilized in figures 27 and 28 and its percentage of share compared with all the other keywords.

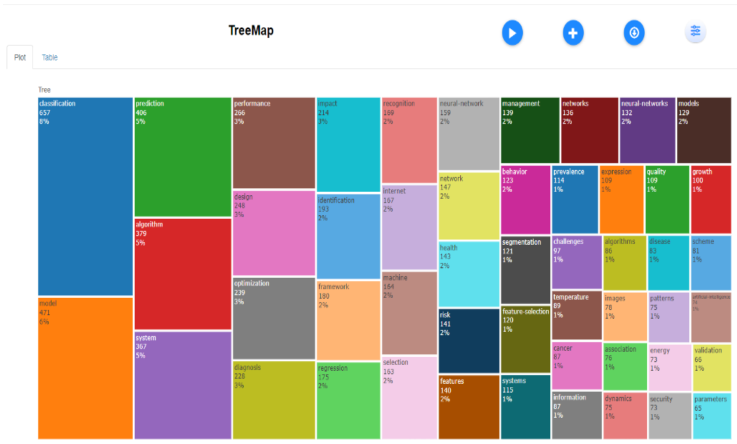

Figure 29: Tree map

4.5. Clustering

Figure 30 shows the clustering by document coupling, at which one can easily infer there were exactly two clusters, one with the keyword's classification, optimization and algorithm, and another cluster with classification, segmentation and recognition. In figure 31, Bhattacharya et al., (2021) is the center of cluster 1 with a normalized local citation score of 15.17.

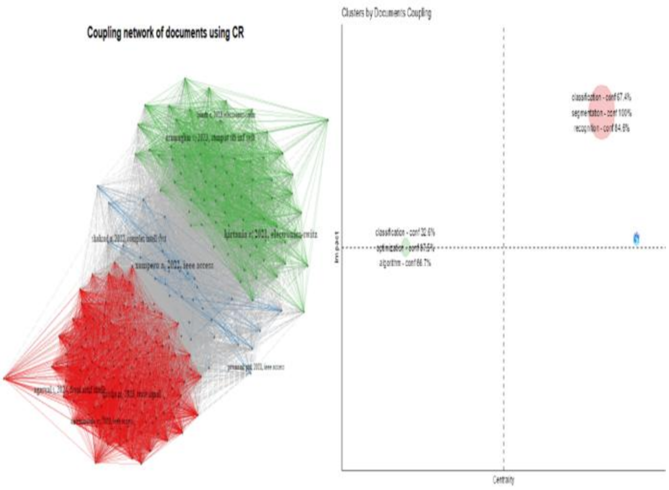

Figure 30: Clustering by Documents Coupling

4.6. Structures

The conceptual structure of the co-occurrence network diagram given in Figure 31 has been created for the keywords plus with the network layout of Fruchterman & Reingold with Walktrap clustering algorithm. Fruchterman method has a disadvantage of high run time. To overcome that issue, Walktrap algorithm the distance between the nodes is calculated by the random walks which is highly used in AI in the network analysis is used. Walktrap is an efficient algorithm that can run with Omn2 time complexity and On2 space complexity in the dreadful case.

Network analysis is based on the matrix methods, for instance, if there is a relationship, then the edges are connected with an edge. Hence,

Ai,j=1,if there is a relationship 0. if there is no relationship.

Normalization was done for the associations with 54 nodes, repulsion force 1, a minimum number of incidence edges 1, and the isolated nodes, which were removed for the betterment of the network diagram.

4.6.1. Conceptual Structures

The conceptual structure of the co-occurrence network diagram given in Figure 31 has been created for the keywords plus with the network layout of Fruchterman & Reingold with Walktrap clustering algorithm. Fruchterman & Reingold layout has advantages of producing good quality results with flixibillity, intuitiveness, simplicity, interactivity and with strong theoretical foundations, which is highly used in AI conceptual structure oriented graphical illustrations. Normalization was performed for the associations with 54 nodes, a repulsion force of 1, a minimum number of incidence edges of 1, and the isolated nodes, which were removed to improve the network diagram.

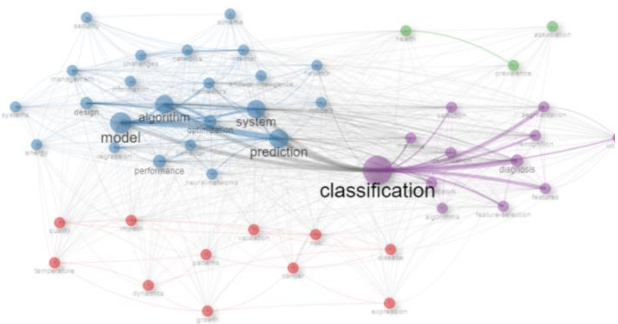

Figure 31: Co-occurrence Network Diagram

The thematic map of the keywords plus the number of words 50, minimum cluster frequency and label size set to 1 with a number of labels 10, community repulsion 0.3 with Walktrap clustering algorithm one can obtain Figure 32. In this, the keywords like model, prediction, algorithm, system, performance, design, optimization, regression, ML, selection, classification, diagnosis, recognition, features, and segmentation become the basic themes, health, risk prevalence, growth, challenges, internet become in centrality in development and relevance whereas, quality, temperature, cancer, disease, dynamics, patterns and validation become the niche themes.

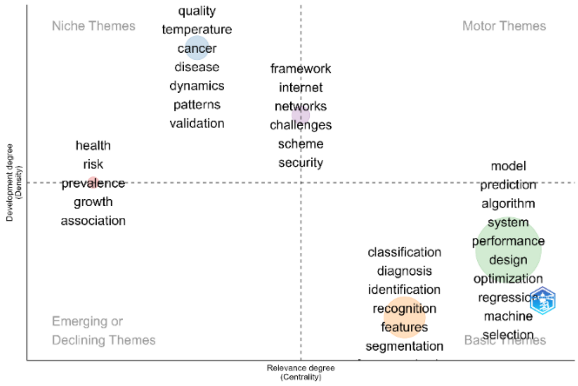

Figure 32: Thematic Map Niche Themes

4.6.2. Intellectual Structures

Figure 33 illustrates another co-citation network, providing a visualization of how different publications are frequently cited together within the research body. Central nodes like He et al., (2016) and Cortes & Vapnik, (1995) Figure 33 illustrates that, typically, key publications serve as a bridge connecting various research topics or communities, underscoring their pivotal role in the research landscape. In Figure 33, the node He et al., (2016) stands out due to its central position and large size, suggesting it is a seminal paper within the network, frequently cited across multiple research clusters.

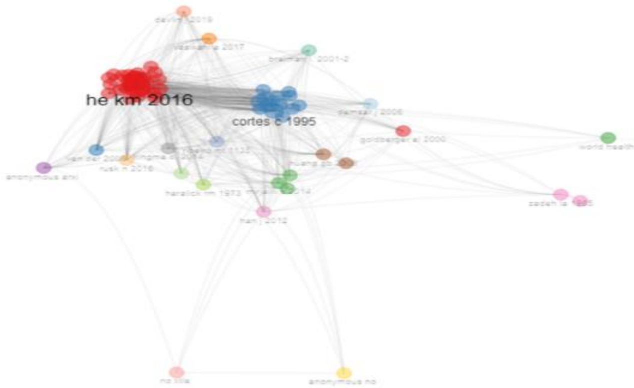

Figure 33: Co-citation Network

Figure 34 gives us a visualization of the historical development and interconnection of research publications over time. The strong clustering of the red nodes, like those by Bhattacharya et al., (2021) and Reddy et al., (2020), Gadekallu et al., (2020) , Bhattacharya et al., (2020) suggests a closely related set of studies that are foundational or influential in the field. The isolated blue node, Suryawanshi et al., (2022) could indicate emerging research that has yet to be heavily integrated into the main body of literature or represents a new direction or methodology in the field. This visualization helps researchers quickly understand the key contributors and influential papers within the field, as well as how current research trends are evolving.

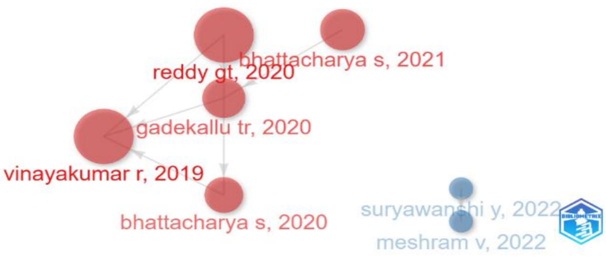

Figure 34: Historiograph

The thematic evolution graph in Figure 35 illustrates how the major keyword trend evolves over time. In 2024 and 2025, the word data is ruling, whereas in 2022, 2023 learning, 2020 2021 results, and 2016-2020, it was studied. Hence, studyresultslearningmodelStudydata is the evolution of the keywords over time.

Figure 35: Thematic Evolution over the period

4.6.3. Social Structures

A. Kumar et al., (2024) and S. Kumar et al., (2024) appears as a central in Figure 36, a highly influential author, indicated by the large node size and numerous connections. This suggests that A. Kumar et al., (2024) and S. Kumar et al., (2024) work is foundational and widely referenced across various research themes within the field. The distinct color groups, such as the purple and blue clusters, suggest thematic or disciplinary coherence among groups of co-cited authors, which might be explored further to understand specific subfields or collaborative networks.

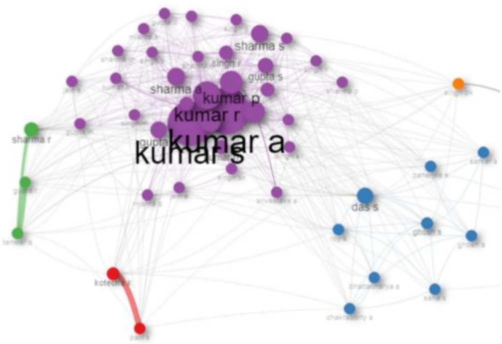

Figure 36: Collaborative Network

The dense network of lines in Figure 37, especially over Europe, North America, and parts of Asia, suggests a high level of collaboration between countries in these regions. The visual density of lines spanning across the Atlantic indicates robust transcontinental collaborations, possibly reflecting established academic and research connections.

Figure 37: Country Collaboration Map

The entire analysis could aid policymakers and researchers in identifying research trends and gaps, enhancing research collaboration, forecasting scenario planning, evidence-based policymaking, public health communication, and training and evaluation needs. Through the mechanisms discussed in this paper, this bibliometric analysis not only enhances the understanding of the research landscape but also serves as a strategic tool in managing the public health responses to pandemics, ultimately aiding in better preparedness and more informed policymaking.

5. CONCLUSION

The thorough bibliometric analysis, enhanced by AI techniques, has offered a hitherto unheard-of window into the field of COVID-19 research in India, revealing the complex web of scientific activity spanning numerous fields and approaches. By effectively mapping the conceptual, social, and intellectual components of the COVID-19 research landscape, the study demonstrated a strong network of scholarly collaboration across national boundaries. Important discoveries highlight the significant concentration of research activity in a small number of core journals, highlighting the crucial role that particular publications and writers have in influencing conversation. Additionally, the use of cutting-edge AI algorithms has not only improved data extraction and analysis but also brought attention to new trends and possible lines of inquiry. This work serves as a strategic resource for policymakers and healthcare professionals who want to use scientific insights to make informed decisions in health crises. It also enriches academic discourse by highlighting the most influential sources, prolific authors, and significant publications. Essentially, this study establishes a standard for future academic research on pandemic events and beyond by highlighting the revolutionary potential of incorporating AI into bibliometric studies. To highlight the challenge faced in the analysis, are the data availability and the dynamic nature of the data. As bibliometric analysis is the vast work, it is challenging to provide the entire calculation methodology. Hence, a future researcher may go ahead with any one or two types of analysis like author and document analysis to provide a lucid explanation. The knowledge gained from this analysis will likely guide future research projects, resulting in a more proactive, knowledgeable response to international health crises, policy holders. As it is used Fruchterman & Reingold layout, Walktrap clustering algorithm, the future researchers can experiment with other techniques for the enhancement.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors received no financial support for the research, authorship, and/or publication of this article.

Acknowledgements

The authors would like to express their gratitude to the College of Computing and Information Sciences (Mathematics section) at the University of Technology and Applied Sciences, Al Musannah, Sultanate of Oman, for supporting this research.

REFERENCES

1. Ameet, & Rana, C. (2024). Bibliometric Analysis of COVID-19-based Models of India. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4754962
2. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M. A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V. A., Krishnamurthi, G., Rohé, M.-M., … Jodoin, P.-M. (2018). Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging, 37(11), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502
4. Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Computational Intelligence and Neuroscience, 2021(1), 8387680. https://doi.org/10.1155/2021/8387680
5. Bhattacharya, S., Reddy Maddikunta, P. K., Pham, Q.-V., Gadekallu, T. R., Krishnan S, S. R., Chowdhary, C. L., Alazab, M., & Jalil Piran, Md. (2021). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable Cities and Society, 65, 102589. https://doi.org/10.1016/j.scs.2020.102589
6. Bhattacharya, S., S, S. R. K., Maddikunta, P. K. R., Kaluri, R., Singh, S., Gadekallu, T. R., Alazab, M., & Tariq, U. (2020). A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU. Electronics, 9(2), 219. https://doi.org/10.3390/electronics9020219
7. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
8. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
9. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
10. Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y., & Yuen, K. Y. (2007). Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. Clinical Microbiology Reviews, 20(4), 660–694. https://doi.org/10.1128/CMR.00023-07
11. Cleveland. (2021). Coronaviruses Have Been Around For Centuries: What Differentiates COVID-19? https://consultqd.clevelandclinic.org/coronaviruses-have-been-around-for-centuries-what-differentiates-2019-ncov
12. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
13. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary

perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
14. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642
15. Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., Ra, I.-H., & Alazab, M. (2020). Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. Electronics, 9(2), 274. https://doi.org/10.3390/electronics9020274
16. Gaur, N., Chaudhary, R., Yadav, M., Srivastava, S., Chaudhary, V., & Shah, M. A. (2024). A stochastic analysis and bibliometric analysis of COVID-19. Indian Journal of Microbiology Research, 11(4), 342–353. https://doi.org/10.18231/j.ijmr.2024.058
17. Gautam, P., Maheshwari, S., Kaushal-Deep, S. M., Bhat, A. R., & Jaggi, C. K. (2020). COVID-19: A Bibliometric Analysis and Insights. International Journal of Mathematical, Engineering and Management Sciences, 5(6), 1155–1169. https://doi.org/10.33889/IJMEMS.2020.5.6.088
18. Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P., & Sikdar, B. (2019). A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures. IEEE Access, 7, 82721–82743. https://doi.org/10.1109/ACCESS.2019.2924045
19. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90
20. Hochreiter, S., & Schmidhuber, J. (1997). Flat Minima. Neural Computation, 9(1), 1–42. https://doi.org/10.1162/neco.1997.9.1.1
21. Jain, R., Gupta, M., Taneja, S., & Hemanth, D. J. (2021). Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence, 51(3), 1690–1700. https://doi.org/10.1007/s10489-020-01902-1
22. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083
23. Karbasi, Z., Gohari, S. H., & Sabahi, A. (2023). Bibliometric analysis of the use of artificial intelligence in COVID-19 based on scientific studies. Health Science Reports, 6(5), e1244. https://doi.org/10.1002/hsr2.1244
24. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386
25. Kumar, A., Luthra, S., Yadav, S., Agrawal, R., & Frederico, G. F. (2024). Exploring the relationship between digitalization, resilient agri-food supply chain management practices and firm performance. Journal of Enterprise Information Management, 37(2), 511–543. https://doi.org/10.1108/JEIM-03-2022-0095
26. Kumar, A., Mishra, A., Malik, R., Hachiya, T., Jürgenson, T., Namba, S., Posner, D. C., Kamanu, F. K., Koido, M., Le Grand, Q., Shi, M., He, Y., Georgakis, M. K., Caro, I., Krebs, K., Liaw, Y.-C., Vaura, F. C., Lin, K., Winsvold, B. S., … Debette, S. (2022). Stroke genetics informs drug discovery and risk prediction across ancestries. Nature, 611(7934), 115–123. https://doi.org/10.1038/s41586-022-05165-3
27. Kumar, R. N., Chandan, K., Saadeh, R., Qazza, A., Karthik, K., Varun Kumar, R. S., Khan, U., Masmoudi, A., Abdou, M. M. M., Ojok, W., & Kumar, R. (2024). Predicting the thermal distribution in a convective wavy fin using a novel training physics-informed neural network method. Scientific Reports, 14(1), 7045. https://doi.org/10.1038/s41598-024-57772-x
28. Kumar, R., Puri, D., Kumar, S., Thakur, M. S., Fekete, G., Lee, D., & Singh, T. (2024). Performance analysis and modelling of circular jets aeration in an open channel using soft computing techniques. Scientific Reports, 14(1), 3140. https://doi.org/10.1038/s41598-024-53407-3
29. Kumar, S., Patil, S., Varadarajan, V., Mazhar, S. M., Sahibzada, A., Ahmed, N., Sinha, O., Shaw, K., & Kotecha, K. (2022). Explainable Artificial Intelligence for Intrusion Detection System. Electronics, 11(19), 3079. https://doi.org/10.3390/electronics11193079
30. Kumar, S., Uppar, R., & Dinesha, P. (2024). Yield optimization of nonedible vegetable oil-based bio-lubricant using design of experiments. Environment, Development and Sustainability, 26(11), 29557–29582. https://doi.org/10.1007/s10668-024-04529-1
31. Liew, Y. Y., Dong, Q., Lakshman, N., & Khajuria, A. (2024). The 100 most-cited articles in COVID-19: A bibliometric analysis. European Journal of Public Health, 34(4), 744–752. https://doi.org/10.1093/eurpub/ckae098
32. Maddikunta, P. K. R., Pham, Q.-V., B, P., Deepa, N., Dev, K., Gadekallu, T. R., Ruby, R., & Liyanage, M. (2022). Industry 5.0: A survey on enabling technologies and potential applications. Journal of Industrial Information Integration, 26, 100257. https://doi.org/10.1016/j.jii.2021.100257

33. Mishra, N., Bharti, T., Tiwari, A. K., & Pfajfar, G. (2024). Public and scholarly interest in social robots: An investigation through Google Trends, bibliometric analysis, and systematic literature review. Technological Forecasting and Social Change, 206, 123578. https://doi.org/10.1016/j.techfore.2024.123578
34. Misra, P., & Gupta, J. (2024). Envisioning Emerging Research Trends on Impact of COVID-19 on Indian Economy: A Bibliometric Study Applying Latent Semantic Analysis and Text Mining Technique. The Indian Economic Journal, 00194662241238588. https://doi.org/10.1177/00194662241238588
35. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. IEEE Access, 7, 81542–81554. https://doi.org/10.1109/ACCESS.2019.2923707
36. Patralekh, M. K., Iyengar, K. P., Jain, V. K., & Vaishya, R. (2021). Bibliometric analysis of COVID-19 related publications in Indian orthopaedic journals. Journal of Clinical Orthopaedics and Trauma, 22, 101608. https://doi.org/10.1016/j.jcot.2021.101608
37. Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. IEEE Access, 8, 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942
38. Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., & Wang, X. (2022). A Survey of Deep Active Learning. ACM Computing Surveys, 54(9), 1–40. https://doi.org/10.1145/3472291
39. Salik Ata, N. (2023). Bibliometric Analysis of Covid 19 and Industry 4.0 Studies Published In Web of Science Database. İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 12(4), 2344–2362. https://doi.org/10.15869/itobiad.1344615
40. Saw, T. B., Doostmohammadi, A., Nier, V., Kocgozlu, L., Thampi, S., Toyama, Y., Marcq, P., Lim, C. T., Yeomans, J. M., & Ladoux, B. (2017). Topological defects in epithelia govern cell death and extrusion. Nature, 544(7649), 212–216. https://doi.org/10.1038/nature21718
41. Sethy, P. K., Behera, S. K., Ratha, P. K., & Biswas, P. (2020). Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine. International Journal of Mathematical, Engineering and Management Sciences, 5(4), 643–651. https://doi.org/10.33889/IJMEMS.2020.5.4.052
42. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556
43. Singh, T., Rao Padubidri, J., Shetty, P. H., Antony Manoj, M., Mary, T., & Thejaswi Pallempati, B. (2024). The top 50 most-cited articles about COVID-19 and the complications of COVID-19: A bibliometric analysis. F1000Research, 13, 105. https://doi.org/10.12688/f1000research.145713.3
44. Suryawanshi, Y., Patil, K., & Chumchu, P. (2022). VegNet: Dataset of vegetable quality images for machine learning applications. Data in Brief, 45, 108657. https://doi.org/10.1016/j.dib.2022.108657
45. Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things, 11, 100222. https://doi.org/10.1016/j.iot.2020.100222
46. UC San Diego Health. (2019). Novel Coronavirus Circulated Undetected Months before First COVID-19 Cases in Wuhan, China. UC San Diego Health. https://health.ucsd.edu/news/press-releases/2021-03-18-novel-coronavirus-circulated-undetected-months-before-first-covid-19-cases-in-wuhan-china/
47. Vaishya, R., Gupta, B. M., Misra, A., Mamdapur, G. M. N., Walke, R., & Vaish, A. (2023). Top 100 highly cited papers from India on COVID-19 research: A bibliometric analysis of the core literature. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 17(11), 102898. https://doi.org/10.1016/j.dsx.2023.102898
48. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep Learning Approach for Intelligent Intrusion Detection System. IEEE Access, 7, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334
49. Wang, X. V., & Wang, L. (2021). A literature survey of the robotic technologies during the COVID-19 pandemic. Journal of Manufacturing Systems, 60, 823–836. https://doi.org/10.1016/j.jmsy.2021.02.005