Environ Res Tec, Vol. 7, Issue. 3, pp. 356–377, September 2024

Environmental Research and Technology
https://ert.yildiz.edu.tr - https://dergipark.org.tr/tr/pub/ert
DOI: https://10.35208/ert.1434390

Research Article

Bibliometric analysis of Indian research trends in air quality forecasting research using machine learning from 2007–2023 using Scopus database

Asif ANSARI*, Abdur Rahman QUAFF

Department of Civil Engineering, National Institute of Technology, Patna, India

ARTICLE INFO

Article history
Received: 09 February 2024
Revised: 10 March 2024
Accepted: 20 April 2024

Key words:
Air quality; Air pollution; Bibliometric analysis; Prediction; Machine learning; Vosviewer; R-package

ABSTRACT

Machine-learning air pollution prediction studies are widespread worldwide. This study examines the use of machine learning to predict air pollution, its current state, and its expected growth in India. Scopus was used to search 326 documents by 984 academics published in 231 journals between 2007 and 2023. Biblioshiny and Vosviewer were used to discover and visualise prominent authors, journals, research papers, and trends on these issues. In 2018, interest in this topic began to grow at a rate of 32.1 percent every year. Atmospheric Environment (263 citations), Procedia Computer Science (251), Atmospheric Pollution Research (233) and Air Quality, Atmosphere, and Health (93 citations) are the top four sources, according to the Total Citation Index. These journals are among those leading studies on using machine learning to forecast air pollution. Jadavpur University (12 articles) and IIT Delhi (10 articles) are the most esteemed institutions. Singh Kp's 2013 "Atmospheric Environment" article tops the list with 134 citations. The Ministry of Electronics and Information Technology and the Department of Science and Technology are top Indian funding agency receive five units apiece, demonstrating their commitment to technology. The authors' keyword co-occurrence network mappings suggest that machine learning (127 occurrences), air pollution (78 occurrences), and air quality index (41) are the most frequent keywords. This study predicts air pollution using machine learning. These terms largely mirror our Scopus database searches for "machine learning," "air pollution," and "air quality," showing that these are among the most often discussed issues in machine learning research on air pollution prediction. This study helps academics, professionals, and global policymakers understand "air pollution prediction using machine learning" research and recommend key areas for further research.

Cite this article as: Ansari A, Quaff AR. Bibliometric analysis of Indian research trends in air quality forecasting research using machine learning from 2007–2023 using Scopus database. Environ Res Tec 2024;7(3)356–377.

INTRODUCTION

Annually, the deleterious effects of air pollution on the environment, human health, and the global economy render it a pervasive hazard with far-reaching implications for all individuals. According to the World Health Organisation, the annual number of preventable deaths caused by indoor and outdoor pollution exceeds seven million [1]. This phenomenon can be attributed to the elevated mortality rates associated with various health conditions, such as stroke [2], coronary heart disease [3], chronic obstructive pulmonary disease [4], lung cancer [5], and acute respiratory infections [6, 7]. Furthermore, it is worth noting that in the year 2019, a significant majority of the global population, specifically 99%, resided in geographical areas that failed to meet the air quality standards set out by the World Health Organisation

*Corresponding author.
*E-mail address: asifa.ph21.ce@nitp.ac.in

(WHO). Furthermore, as stated by the World Health Organisation (WHO), the deleterious impacts of air pollution on the environment in Southeast Asia account for a staggering 91% of avoidable fatalities in nations with lower and moderate income levels. In a study conducted in 2013, led by the International Agency for Research on Cancer of the World Health Organisation, it was found that particulate matter components have been classified as carcinogenic to humans [8]. These components are believed to be the primary contributors to the increasing prevalence of cancer, particularly lung cancer [9]. There is a need to enhance public knowledge regarding the development and dissemination of pollution maps that provide timely alerts for hazardous air pollutants.

In recent times, numerous noteworthy occurrences of natural and anthropogenic pollution have had severe detrimental effects on both human well-being and the ecological system [10–14]. The prevalent natural air pollutants include ozone (O_3), sulphur dioxide (SO_2), carbon monoxide (CO), particle matter (PM), and nitrogen dioxide (NO_2). The primary sources of air pollution resulting from human activities encompass electricity generation, emissions from stationary vehicles, industrial emissions, agricultural emissions, emissions from home heating systems, including aquatic surfaces, cooking activities, and so forth. Multiple studies [9, 15, 16] have indicated that air pollution in any given place is influenced by both regional and international sources, as well as adjacent local sources. The levels of air pollution vary significantly across different locations as a result of variances in factors such as the quantity, composition, fuel type, emission control technology, and concentration of several sources. The temporal patterns of air pollutant concentrations exhibit considerable fluctuation due to weather restrictions that vary on a daily, weekly, and yearly basis, as well as the presence of several sources. Various machine learning techniques have been employed in the domain of air pollution to forecast air pollution levels [17, 18], determine the sources of pollution [19, 20], and monitor air pollution [21–23], among other diverse applications.

Scholars employ a range of qualitative and quantitative methodologies in their literature reviews to evaluate and structure their texts and discoveries. Bibliometrics is a commonly employed method for assessing the existing state of knowledge on a certain subject [24–26]. This discipline offers a methodical, replicable, statistically rigorous, and transparent approach to conducting reviews. This approach involves utilising specific elements such as titles, abstracts, author names, journal names, keywords, affiliations, references, and other relevant information sourced directly from academic databases, including but not limited to Scopus and WoS. Consequently, the analysis derives the issues that have been examined, identifies the most prominent institutions and scholars, along with enduring patterns, detects shifts in the disciplinary parameters through time, and provides a comprehensive outlook on the topic. A significant increase in bibliometric research has been observed across various academic areas, such as tourism [27], health and infection [28] and educational administration [29, 30].

In order to gain clarity and establish a clear direction for addressing the global problem of air pollution, it is necessary to do a bibliometric study, as suggested by author [31]. Numerous bibliometric studies have been conducted in the past, exploring various aspects of air pollution. In the study conducted by author [32], an analysis was performed on a comprehensive dataset of global scientific papers pertaining to pollution research spanning the years 2000 to 2016. The analysis encompassed both quantitative and qualitative aspects, allowing for a comprehensive understanding of the research landscape in this field. In their study, Kumar [33] performed an analysis encompassing all the studies on air pollution published in the Web of Science (WoS) database over the period from 2005 to 2014. In their study, M. Kumar [34] undertake a thorough examination of the existing body of research pertaining to the health consequences associated with the exposure of young individuals to air pollution. Yang [35] conducted a comprehensive analysis of the existing literature pertaining to the factors contributing to air pollution throughout the period from 2006 to 2015. The authors conducted a bibliometric analysis of scholarly works pertaining to respiratory health issues associated with outdoor air pollution [36]. Furthermore, a comprehensive analysis was conducted on several scholarly articles pertaining to the utilisation of machine learning techniques in the context of air pollution-related applications. Dogra [37] have undertaken a bibliometric examination of statistical forecasting and prediction methodologies pertaining to air pollution. The researchers employed the Markov chain and evolving trees methodologies to project forthcoming advancements in the study of significant air contaminants. Rybarczyk [38] conducted a comprehensive literature review on the utilisation of machine learning in the context of air pollution. The study revealed that researchers tend to favour regression techniques for estimation purposes, while prediction applications commonly employ neural network algorithms and support vector machines. This finding underscores the prevalent preferences within the research community regarding the selection of machine learning methods for addressing air pollution challenges. In their study, Guo [39] conducted a comprehensive search of the Web of Science database to identify all relevant scholarly articles. Subsequently, they employed CiteSpace 5.8.R1, a widely used software tool, to analyse various aspects such as countries, organisations, authors, keywords, and references. The primary objective of this analysis was to identify prominent areas of research and emerging trends pertaining to the application of artificial intelligence in the domain of air pollution. The purpose of this endeavour was to identify areas of concentrated activity and emerging trends in the field of artificial intelligence. The topic of predicting air pollution has been extensively investigated [40]. Statistical forecasting, numerical forecasting methodologies, and artificial intelligence were employed to classify the forecasting models. Most studies that employ time-series and machine learning techniques to predict air quality commonly utilise multilayer neural networks. However, it is worth noting that these networks were originally designed for tasks other than time-series modelling [41, 42]. The utilisation of machine learning and time series analysis in the prediction of air quality has been investigated by researchers. There have been proposed methodologies utilising data-driven techniques for the purpose of predicting air pollution levels. Zong [43] propose a methodology that integrates meteorological features and air quality data to develop a predictive model for air quality with a lead time of two days. In 2019, Cabaneros [44] conducted a comprehensive assessment of 139 research articles pertaining to the prediction and estimation of ambient air pollution levels. P. Guo [45] conducted a comprehensive analysis of the research landscape pertaining to construction dust, including its distribution, emerging fields of investigation, and potential avenues for future study, utilising the CiteSpace programme. The primary objective of this study was to examine scholarly articles published in the Web of Science (WoS) database from 2010 onwards, with a specific emphasis on those that addressed the subject of "construction dust." The study presents an analysis of many properties of these publications, including their quantity trend, quality, author cluster, related institution, and journal category. Additionally, the analysis includes article co-citation and keyword co-occurrence. A bibliometric analysis was conducted using CiteSpace 5.7.R3 to examine literature related to ozone pollution in the Web of Science (WoS) database [46]. The authors have reached the conclusion that significant emphasis has been placed on elucidating the mechanism of ozone formation and source allocation, characterising ozone pollution, modelling ozone dispersion at various scales, and assessing the risks posed to humans and plants due to short- and long-term exposure. However, it is necessary to conduct a comprehensive quantitative analysis of various academic articles that cover all potential domains where machine learning methods might be employed for addressing air pollution. This study conducts a bibliometric analysis of the existing research on "Air Pollution Prediction Using Machine Learning" with the aim of providing valuable recommendations for future studies and practical applications.

To the best of the author's knowledge, a limited number of scholars have utilised the WoS database for conducting bibliometric evaluations pertaining to the domains of air pollution and machine learning for global research. Several studies have been conducted on air pollution prediction using machine learning approaches [47–49]. Nevertheless, it is important to acknowledge that the Scopus database has not been employed in any bibliometric assessments pertaining to this subject matter. Additionally, there is no documented evidence of any researcher utilising bibliometric analysis to study machine learning techniques for the specific objective of air quality prediction from an Indian perspective.

The primary aim of this research is to acquire a full comprehension of the utilisation of machine learning methodologies in the examination of air pollution. The researchers conducted an analysis of the literature available in the Scopus online database, focusing on subject categories, article volumes, and journal kinds. This analysis aimed to obtain a thorough grasp of the current status of progress in the field. The aforementioned documents were exported to the VOSviewer and Biblioshiny applications for the purpose of analysis. Through the establishment of collaborative partnerships among nations, authors, and institutions, we have successfully recognised worldwide research needs and fostered cooperative linkages. This study enables air pollution experts to realise the following different research questions: (i) Has there been an increase or reduction in research on the prediction of air pollution using machine learning? (ii) What is the annual growth rate of scholarly papers on this issue? (iii) What are the essential terms associated with the topic of "air pollution and machine learning" as identified in the existing literature? (iv) How do distinguished scientists collaborate with each other? Which academic journals and universities have the most significant impact? The remaining sections are categorised as follows: The subsequent section provides a comprehensive discussion of both the materials used and the processes employed. The findings of the bibliometric analysis conducted using VOSviewer and biblioshiny are presented and analysed in the third section of this study. The study concludes in Section 4, wherein prospective areas for further research are also highlighted.

MATERIALS AND METHODS

In recent years, there has been a notable surge in the volume of research undertaken within technical disciplines. As a result of this phenomenon, the effort of staying abreast of the pertinent literature pertaining to a specific field is becoming more challenging. The aforementioned requirement calls for the development of quantitative bibliometric approaches that possess the capability to effectively handle extensive volumes of data, identify the most influential publications, and reveal the fundamental structure of the field [50]. These methodologies should possess the capability to effectively manage and process large volumes of data. We conducted a comprehensive review of the existing literature on the application of machine learning in air pollution forecasting, with a focus on identifying trends and key factors influencing this subject. To accomplish this, we employed the bibliometric approach proposed by Garfield [51]. The aforementioned technique was employed by our research team. Zupic [52] assert that bibliometric methods employ a quantitative approach to classify, evaluate, and monitor published research. The purpose of this endeavour is to implement a systematic, clear, and replicable process for conducting reviews, with the ultimate goal of enhancing the overall quality of the reviews. Scholars sometimes find it beneficial to derive their conclusions from a corpus of accumulated bibliographic data provided by fellow academics in the discipline who express their perspectives through written works, citations, and collaborative efforts. Scholars use this methodology to draw conclusions. In contrast to the research conducted on air pollution projections utilising machine learning techniques, bibliometrics has been extensively employed in many disciplines, including administration, biology, nutrition, engineering, and various medical specialties. As a result of this, a search was done on August 7, 2023, in the Scopus database using a Boolean search strategy to identify pertinent literature published throughout the timeframe of 2007 to 2023. To determine the keywords for the present study, the research team utilised their own pre-existing expertise on the subject matter, as well as previously conducted research, keyword analysis of specific databases, and assessment of relevant studies conducted in other locations. The search queries include [TITLE-ABS-KEY ("air pollution" OR "air quality") AND TITLE-ABS-KEY ("prediction" OR "forecasting") AND TITLE-ABS-KEY ("machine learning" OR "artificial intelligence")] with the longest period permitted by the database to encompass all possible articles. Zupic [52] conducted a process of selecting and importing relevant texts from a database using biblioshiny, a web interface designed for the bibliometrix programme. In order to assess subtle distinctions, an examination was conducted on the most reliable Indian scholarly publications within the English-language domains of Article, Review, Conference Paper, Conference Review, and Book Chapter. The results were organised based on the number of citations, resulting in a total of 326 instances. Figure 1 shows a flow chart diagram for the collection of literature (identification, screening, and eligibility).

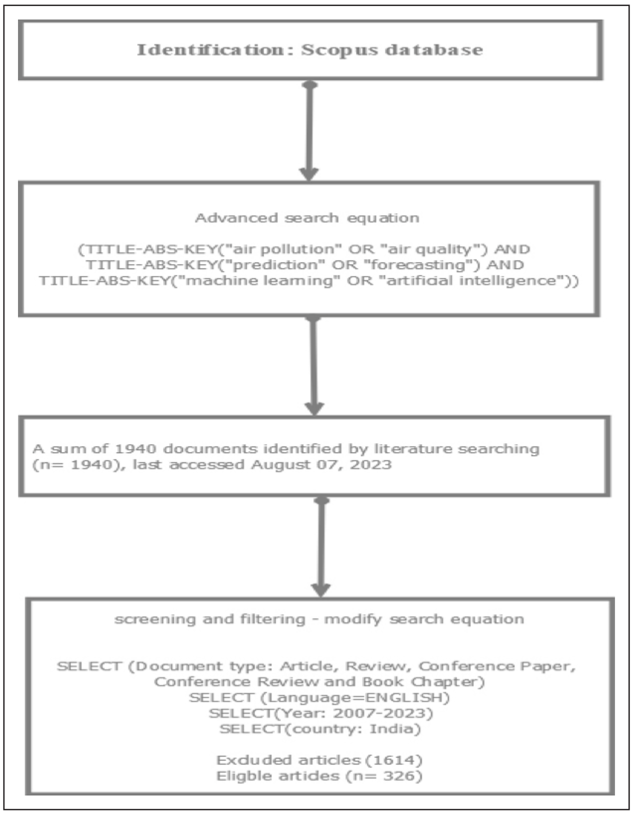

Figure 1. Flow chart diagram for the collection of literature (identification, screening, and eligibility).

VOSviewer has also performed document analysis to enhance the comprehensibility of network diagrams and overlay diagrams. The methodology incorporates performance assessment and science mapping as key components [53–56]. As a component of performance analysis, scholarly publications are scrutinised with respect to the authors, countries, and institutions involved. In contrast, science mapping uses bibliometric approaches to identify and analyse patterns within the realm of scientific study. The review articles by Abafe [31] and Velasco-Muñoz [50] contribute quantitative rigour to the assessment of subjective literature judgements and provide empirical evidence for theoretically defined categories. The following indicators were specifically examined: (i) A comprehensive summary encompassing pertinent information such as important details, annual scientific output, and the annual average of citations (ii) Various sources, including those commonly referenced within a certain locality, the most relevant sources, and the impact of sources The focus of this inquiry is on the primary relationships and their evolutionary progression. (iiib) The analysis also focuses on authors, including their relevance to the topic, the frequency of citations received locally, the impact they have had on the field, and their productivity over a period of time. In this section, we will discuss the study of author keyword cooccurrence, the identification of the most commonly used terms, and the examination of trending issues. The corpus comprises various texts, including those that have garnered the highest number of citations on national scales.

In this study, S-curve analysis was utilized to show how the work changed over time. Several modeling approaches have been used to forecast the future of invention. However, researchers have used the S-curve to predict technological advancements [15, 57], and inventions and technologies typically follow it. The four stages of a technical structure's transformation can be simulated using an S-curve simulation: emergence, growth, maturity, and saturation or decline [58]. Using S-curve analysis, we could conduct a quantitative study on the prediction of air quality in the future using a variety of machine learning models. We employed the logistic model, where k and an act as the determinants of the shape of the curve, and x denotes the time bucket. An analysis of variance (ANOVA) and an independent t-test were conducted in order to look into the study's potential implications further.

RESULTS AND DISCUSSIONS

Descriptive Analysis

The dataset, which spans the years 2007 to 2023 and is derived from 231 sources, including books and journals, is shown in Table 1 and has a total of 326 documents. The figures show a strong yearly growth rate of 32.1% over this time period. The dataset's documents are 1.74 years old on average and have 6.911 citations on average per document. The total number of references cited in all papers is 9053. The dataset includes a wide range of information in terms of content. To classify and describe the papers, 1746 keywords and 808 author-specific keywords were included. Only six of the 984 authors in the sample contributed to documents with a single author. The data also shows a trend towards author collaboration, with 3.6 co-authors on average per

Table 1. Main information about data

Description	Results
Timespan	2007:2003
Sources (Journals, Books, etc)	231
Documents	326
Annual growth rate %	32.1
Document avarage age	1.74
Average citations per doc	6.911
References	9053
Keywords plus	1746
Author's keywords	808
Authors	984
Authors of single-authored docs	6
Single-authors per doc (collabration)	6
Co-authors per doc (collabration)	3.6
International co-authorships %	14.42
Article	133
Book chapter	22
Conference paper	161
Review	10

document. The percentage of international co-authorships in collaborations is about 14.42%. The dataset has 133 articles (41%), 22 book chapters (7%), 161 conference papers (49%), and 10 reviews (3%), among other document types. The various document types employed in the study are shown in Figure 2a. The dataset's multidisciplinary nature and its ability to contribute to numerous fields of study are reflected in the wide variety of document formats. Overall, the dataset offers a thorough selection of academic papers that span a wide range of topics and exhibit notable development and collaboration within the academic community.

Figure 2b displays the annual number of papers published along with the publication patterns from 2007 to 2023. An intriguing pattern of intellectual activity can be seen in the distribution of articles over several years. There were few or no publications published in the years from 2007 to 2012, indicating a comparative lull in activity. However, in later years, this pattern changed. A single essay published in 2013 served as the catalyst for a revitalised scholarly production. Three articles in 2015 helped the momentum pick up, and two pieces each in 2016 and 2017 helped it continue to gain ground. The genuine uptick started in 2018 with six articles, then increased significantly in 2019 with 27 articles. With 49 articles, the year 2020 saw a notable increase in scholarly contributions; this trend continued into the next year, 2021, with 58 articles. The maximum number of articles—91—was recorded in 2022, marking the peak of this expansion. While 2023 saw a small decline with 86 publications, the overall trend highlights a tremendous uptick in research activity, indicating a vibrant and alive environment for academic inquiry and information sharing.

S-Curve Analysis of Publications

Figure 2c shows S-curve plot of cumulative publication over time. Using a logistic growth model, the investigation sought to simulate the growth of total articles from 2007 to 2023. The parameters of this model, which give information about publication trends, stand for the carrying capacity (K), the growth rate (a), and the inflection point (t0). Estimates of these parameters were obtained by fitting the logistic function to the data. The carrying capacity (K) was roughly estimated at 468.21, reflecting the highest publishing level at which the model converges. It was calculated that the growth rate (a), which represents the speed at which publications approach the carrying capacity, is approximately 0.7790. The transition year into a phase of more rapid growth, designated as the inflection point (t0), was determined to be about 2021.94. The coefficient of determination (R2) was calculated to assess the model's performance, yielding a value of about 0.9992. This R2 score implies that the logistic growth model fits the observed data with extraordinary strength. The logistic growth model's high R2 value suggests that it accurately represents the trends in cumulative publications throughout the selected years. The analyses' deep understanding of how publications grow over time makes it easier to understand the patterns and trends in academic work over the time period that was looked at.

Most Influential Journals

Table 2 provides a list of variables related to several air pollution journals and the metrics they are associated with. There is information on the starting year (PY_start), total number of citations (TC), number of articles (NP), g-index, h-index, and m-index. The top 10 most pertinent sources out of 231 sources, as determined by the total number of documents, are an indication of their significant contributions to their respective disciplines. Leading this compilation is "Lecture Notes in Networks and Systems," a noteworthy source with 12 pages, a strong h-index of 3, an excellent g-index of 5, a significant m-index of 0, and a total of 30 citations since it was first published in 2020. Similar to "Lecture Notes in Electrical Engineering," which started its significant journey in 2020, "Lecture Notes in Electrical Engineering" stands out with 7 documents, an h-index of 2, a respectable g-index of 3, and a balanced m-index of 0.5. Since its debut in 2018, "Procedia Computer Science" has maintained its position with six documents, as evidenced by astounding h-index and g-index totals of 6, a notable m-index of 1, and a noteworthy citation total of 251. Furthermore, with 6, 6, and 5 documents each, journals like "Advances in Intelligent Systems and Computing," "Environmental Monitoring and Assessment," and "Communications in Computer and Information Science" continue to be important sources that offer insightful information in their fields. Not to be disregarded, "Atmospheric Pollution Research," "Atmospheric Environment," "Air Quality, Atmosphere, and Health," and "Journal of Cleaner Production," each with 4, 3, 3, and 3 papers, strengthen their places as key sources,

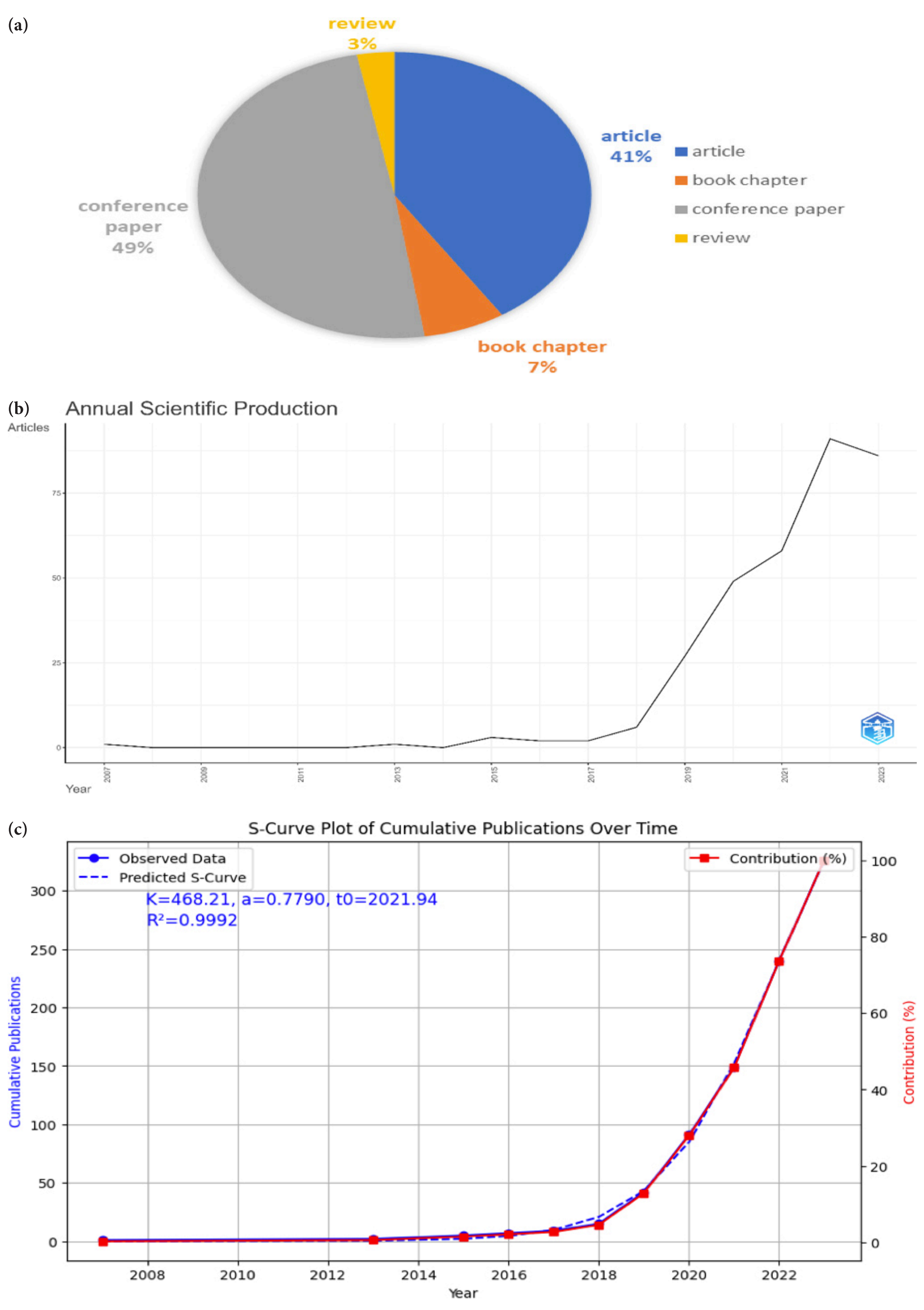

Figure 2. **(a)** Document type. **(b)** Annual scientific publication distribution. **(c)** S-curve plot of cumulative publication over time.

Table 2. Number of publications (NP), total citations (TC), publication year (PY)

Most relevant sources	NP	h_index	g_index	m_index	TC	PY_start
Lecture notes in Networks and Systems	12	3	5	0.75	30	2020
Lecture notes in Electrical Engineering	7	2	3	0.5	11	2020
Procedia Computer Science	6	6	6	1	251	2018
Advances in Intelligent Systems and Computing	6	2	2	0.4	7	2019
Environmental Monitoring and Assessment	6	2	2	1	7	2022
Communications in Computer and Information Science	5	1	2	0.25	6	2020
Atmospheric Pollution Research	4	4	4	0.444	233	2015
Atmospheric Environment	3	3	3	0.273	263	2013
Air Quality, Atmosphere and Health	3	2	3	0.4	93	2019
Journal of Cleaner Production	3	3	3	1	62	2021

Table 3. Sources local impact by total citation (TC) index

Sources local impact by total citation (TC) index	TC	h_index	g_index	m_index	NP	PY_start
Atmospheric Environment	263	3	3	0.273	3	2013
Procedia Computer Science	251	6	6	1	6	2018
Atmospheric Pollution Research	233	4	4	0.444	4	2015
Air Quality, Atmosphere and Health	93	2	3	0.4	3	2019
Plos One	83	1	1	0.143	1	2017
Journal of Cleaner Production	62	3	3	1	3	2021
Sensors and Actuators, B: Chemical	51	1	1	0.25	1	2020
Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018	48	1	1	0.167	1	2018
Environmental Science and Pollution Research	46	2	2	0.5	2	2020
Journal of Ambient Intelligence and Humanized Computing	46	2	2	1	2	2022

encouraging a comprehensive awareness of relevant topics. The top 10 sources, ranked by the quantity of documents published, are shown in Figure 3a.

Similar to Table 2, Table 3 lists a variety of elements related to several air pollution journals and their related metrics. The starting year (PY_start), overall number of citations (TC), total number of publications (NP), and g-, h-, and m-indices are all given. The summary provides a brief overview of the top 10 sources (out of 231) according to the Total Citation (TC) index, which rates them according to their local influence. "Atmospheric Environment" is in first place with a respectable TC index of 263, closely followed by "Procedia Computer Science" with 251 citations and "Atmospheric Pollution Research" with 233 citations. While "Plos One" and "Journal of Cleaner Production" show significant local impact with 83 and 62 citations, respectively, "Air Quality, Atmosphere, and Health" maintains its dominance with 93 citations. With 51 citations, "Sensors and Actuators, B: Chemical" is next, followed by "Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018," "Environmental Science and Pollution Research," and "Sensors and Actuators, B: Chemical," which each have 48. Last but not least, the regional influence is mirrored by the "Journal of Ambient Intelligence and Humanised Computing," with 46 citations. Together, these sources demonstrate their power within their spheres of expertise and make a substantial impact on the academic scene. According to the Total Citation (TC) index, Figure 3b displays the top 10 sources that have the most influence.

Anova Analysis

We used an analysis of variance (ANOVA) to investigate differences in the mean total citations (TC) across various document formats (Table 4). A statistical test called ANOVA identifies significant mean differences between many groups or categories. Our goal was to determine whether there is a discernible difference between publication document categories and the mean TC. The results of our ANOVA show a significant difference in mean TC between various document categories. The calculated F-statistic, which is roughly 3.12, measures how much the TC means vary from type to type of document. If the variability within each group is compared to the F-statistic, the difference between group averages is likely to be more evident. Additionally, the accompanying p-value, which is roughly 0.0264, is crucial in

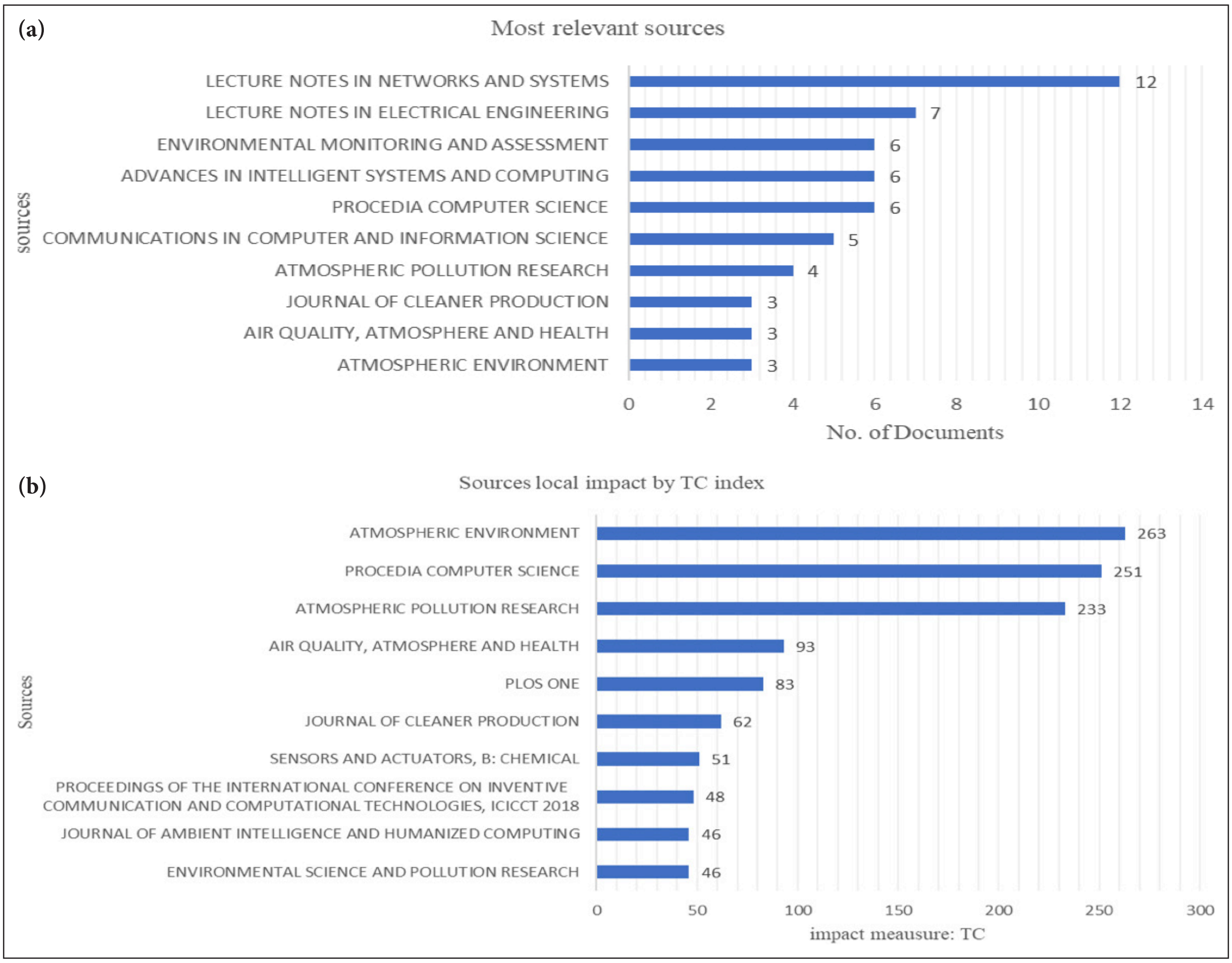

Figure 3. **(a)** Most relevant sources. **(b)** Sources local impact by total citations (TC).

Table 4. ANOVA analysis

Source	Sum of squares	Df	Mean square	F	Sig.
Between groups	2596.517	3	865.5056	3.115282	0.02642
Within groups	89459.9	322	277.8258		
Total	92056.42	325			

proving statistical importance. It denotes how likely it is that such large mean discrepancies will arise simply by chance. The null hypothesis is strongly refuted when the p-value is appreciably small (usually below the selected significance level, frequently set at 0.05). The extremely small p-value in this case clearly suggests that random chance is extremely unlikely to account for the observed mean discrepancy. As a result, our ANOVA results clearly show that the mean number of citations (TC) for different types of documents is different. This finding has consequences for scholars and decision-makers, highlighting the need to take document type into account when evaluating and interpreting citation data. The necessity for careful examination is highlighted by the possibility that various publication types would exhibit distinctive citation behavior patterns.

T-Test Analysis

To determine whether there is a statistically significant difference in securing citations between open access and subscription journals, a t-test was used in this investigation. The alternative hypothesis (H1) proposed that there is a major distinction, whereas the null hypothesis (H0) asserted that there is no significant difference. The t-test produced a t-statistic of around 2.2742 and a corresponding p-value of roughly 0.0236 when using an alpha level (or threshold) of significance (0.05) as the reference point. According to the interpretation of these results, open access and subscription publications secure statistically significantly fewer citations than each other. In particular, the p-value of 0.0236 is less than the chosen level of significance. This means that it is not likely that the changes seen in the number of citations between these types of journals are just random. As a result, we have enough statistical data to reject the null hypothesis and confirm that the availability of a journal (open access or subscription) has a significant impact on the number of citations it receives. It is important to recognize that even though statistical significance has been demonstrated, there may still be practical disparities in situations that occur in real life. We are unable to firmly declare these practical distinctions, however, in light of the conducted study and data. In conclusion, the t-test results strongly imply that there is a statistically significant difference in citation rates between open access and subscription journals.

We also performed a t-test in this analysis to assess whether there is a statistically significant difference in the number of citations received for various document types, including articles, conference papers, conference reviews, reviews, and book chapters. In contrast to the null hypothesis (H0), which claimed there is no significant variation in the number of citations received among different document categories, the alternative hypothesis (H1) stated that there is a substantial difference in citation counts. After applying the t-test to all pairwise comparisons of document categories, we found that the p-values for each pairwise comparison were all greater than the conventional significance level of 0.05. Because we were unable to reject the null hypothesis for any of the comparisons, it is clear that there is no statistically significant difference in the number of citations obtained between any of the document categories included in this research. Conclusion: Based on our data, the number of citations a publication receives is not significantly influenced by the form of document it picks, be it an article, conference paper, review, book chapter, or another kind. The number of citations for academic work appears to be unaffected by the kind of article writers and researchers choose, giving them peace of mind.

Authors

According to Indian Perspective, a total of 984 authors contributed 326 publications to the study on machine learning-based air quality prediction. A graph of author productivity using Lotka's law is shown in Figure 4a, which illustrates how authors contribute to the generation of documents. It demonstrates that, in line with Lotka's distribution pattern, the majority of authors (87.7%) have created just one document, while fewer authors have written more. For example, 7.7% of authors produced two documents, and lower percentages (1.5%) created three to seven documents. This pattern demonstrates a concentrated contribution from a small number of authors, which exemplifies Lotka's Law [59].

According to the number of documents, Table 5 identifies the top 10 authors, and Figure 4b also shows an analysis of their output. The summary offers information about various writers' publication histories. With 15 papers, Kumar A emerges as the most prolific author, followed by Singh S with 7. Six articles each have been submitted by Roy S, Sharma S, Dutta M, Gupta S, Kapoor NR, Kumar P, Marques G, and Middya AI. The dataset receives a significant amount of work from these writers that covers a wide range of issues and topics.

Figure 4c shows the total number of citations and the local effect of the author. The synopsis is a list of local citations for different authors. With 12 local citations apiece, Middya AI, Nath P, and Roy S have strong local recognition. Thomas B. and Babu S. come in second and third, with nine local citations each. With seven local citations, Bhosale A, Gokul PR, Matthew A, and Nair AT have had a significant local influence. Four local citations from Dilliswar Reddy P. bring up the rear of the list. Collectively, these authors demonstrate varying degrees of regional sway, which adds to the dataset's varied intellectual contributions.

Figure 4d displays the top 10 most productive authors from 2007 to 2023. The overview provides a thorough look at the production data for the top 10 authors throughout time. The size of the bubbles represents the volume of published materials. A small one stands in for one publication, and a large one for two. The number of citations every year is directly related to the colour intensity. Notably, in 2020 and 2022, writers like Dutta M. received citations with different total citation counts (TC) and TC per year ratios. Similar to this, authors like Kapoor NR received citations in 2022 and 2023, while Gupta S witnessed high citation activity in 2013. Citations for Kumar A increased significantly in 2022 and 2023, indicating an expanding effect. Between 2020 and 2022, Kumar P received citations, and in 2022, Marques G and Middya AI both received a significant number of citations. While Sharma S's work received citations over the years, Roy S's work received citations in 2022 and 2023. The number of citations for Singh S peaked in 2019, and sporadic activity persisted in the following years.

Distribution of the Most Productive Affiliation

The most pertinent affiliations and associated article counts of Indian research institutions are shown in Figure 5. Notably, Jadavpur University's computer science and engineering department in Kolkata takes the top spot with 12 articles, indicating its prodigious research output. With 10 articles, the Indian Institute of Technology Delhi's centre for atmospheric sciences has a commanding position and demonstrates extensive research activity. Additionally, Sri Sivasubramaniya Nadar College of Engineering and the department of computer science and engineering at the National Institute of Technology Durgapur, both of which have seven articles each, are highlighted for their important contributions to the subject. The department of computer science and engineering at Koneru Lakshmaiah Education Foundation, the University of Engineering & Management in Kolkata, the department of computer science and engineering at Sathyabama Institute of Science and Technology, the department of information science and engineering at M S Ramaiah Institute of Technology, the applied cognitive science lab at Indian Institute of Technology Mandi, and the department of inf This collection of several institutions highlights their significant contributions to India's comput-

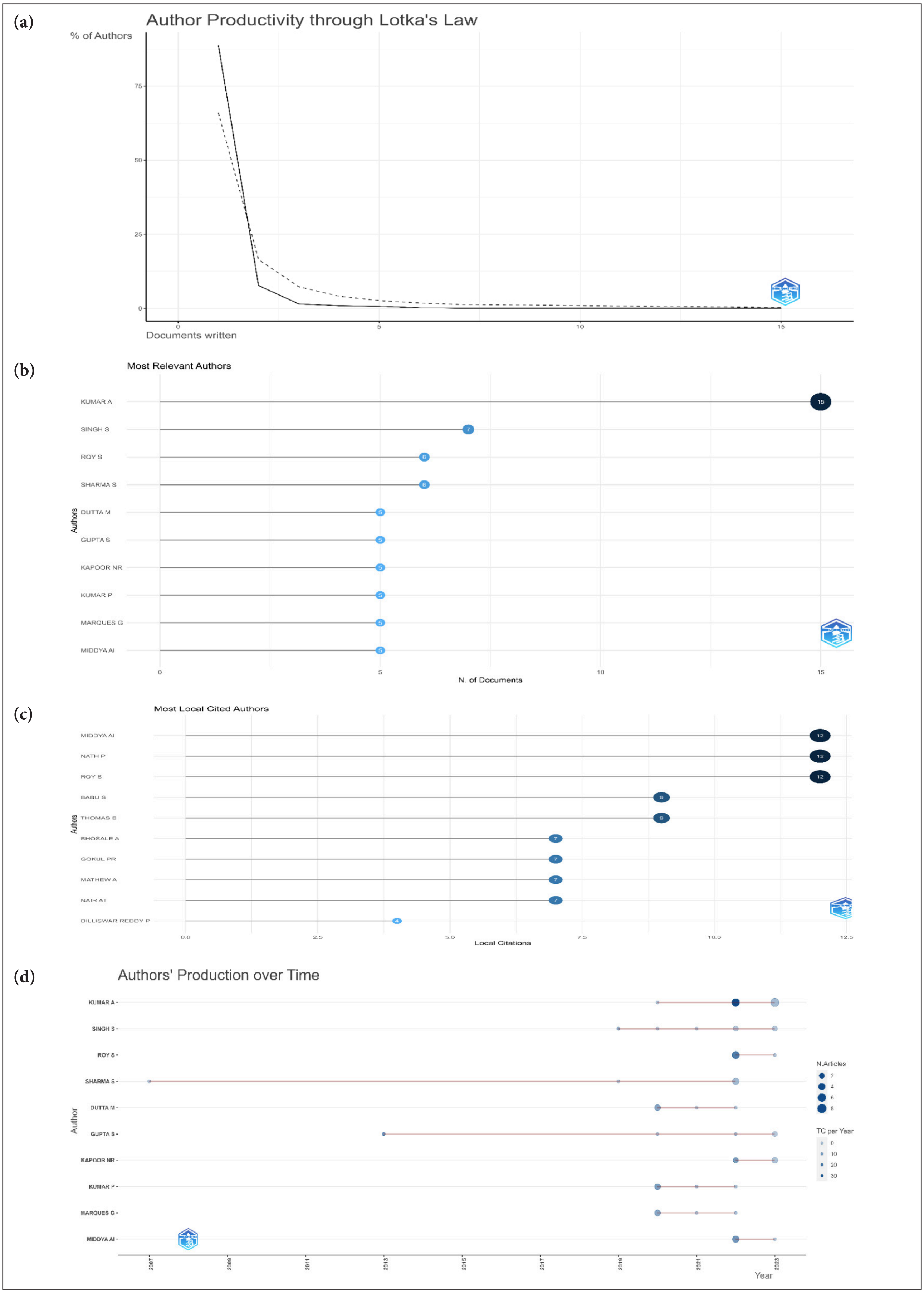

Figure 4. **(a)** Author productivity through Lotka's law. **(b)** The top ten most relevant authors. **(c)** Author local impact. **(d)** Top author's production from 2007–2023.

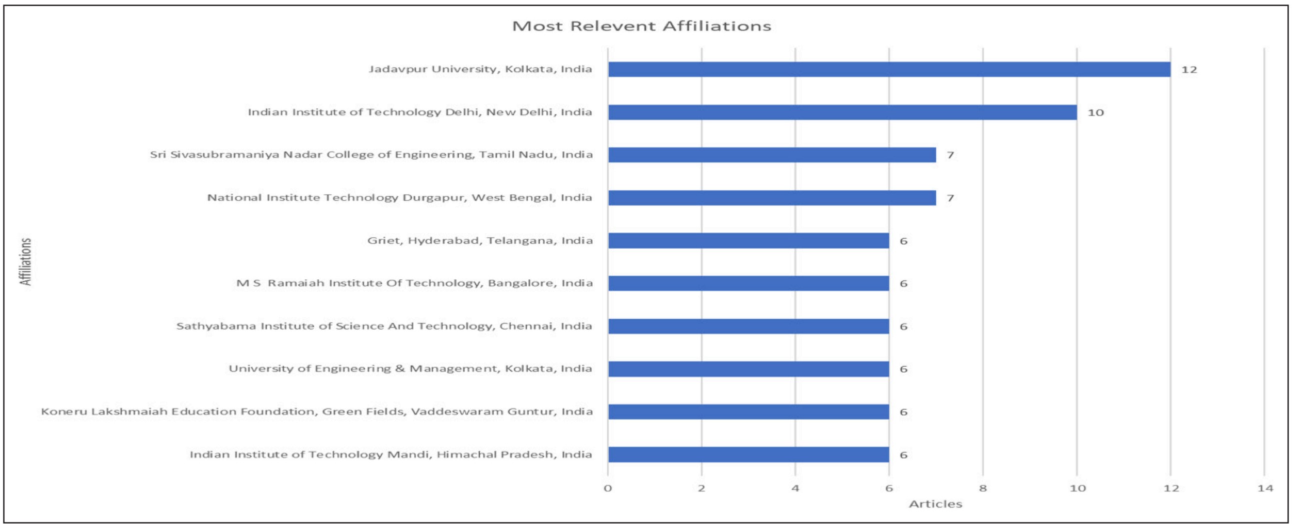

Figure 5. Most relevant affiliations.

Table 5. The top ten leading authors on machine learning-based air pollution prediction

Authors	Articles	Articles fractionalized
Kumar A	15	3.45
Singh S	7	2.17
Roy S	6	2.03
Sharma S	6	1.51
Dutta M	5	1.67
Gupta S	5	1.33
Kapoor NR	5	0.93
Kumar P	5	1.07
Marques G	5	1.67
Middya AI	5	1.53

er science and engineering fields.

Most Global Cited Documents

Understanding how this research stream has developed requires finding the articles that have contributed to the machine learning-based air quality forecast from an Indian perspective. Similar to this, using machine learning to study the patterns of citations in air pollution prediction may offer important suggestions about the direction of future research. Each of the 326 documents, with an average age of 1.74, that were part of the analysis averaged 6.91 citations. The top 10 documents, as shown in Table 6 and Figure 6a, have had a substantial impact in their respective domains and have received broad notice and recognition for their contributions. The article by Singh Kp from 2013, which was published in "Atmospheric Environment," is at the top of the list with 134 citations. Following closely after with 103 citations is the paper by Mishra D., which was also published in "Atmospheric Environment" in 2015. Krishan M.'s research on "Air Quality, Atmosphere, and Health" received 91 citations in 2019. The study by Doreswamy from 2020, which has an impressive 79 citations, is presented in "Procedia Computer Science," strengthening the list even more. The 2015 publication of Mishra D's study in "Atmospheric Pollution Research" received 63 citations and made a notable impact on the subject. It is clear that Rubal's article in "Procedia Computer Science" in 2018 had a significant impact because it was mentioned 54 times. 51 citations support the importance of Acharyya S's article from "Sensors and Actuators, B: Chemical" from 2020. With Ayele Tw's work obtaining 48 citations, "Proceedings of the International Conference on Inventive Communication and Computational Technologies (ICICCT) 2018" is added to the list. Masood A's work from 2021 was cited 46 times in "Journal of Cleaner Production," whereas Amuthadevi C's paper from 2022 has been quoted 39 times in "Journal of Ambient Intelligence and Humanised Computing." These widely cited papers serve as an example of the influence that research has on the greater scientific community and has had on the development of their respective subjects.

Most Local Cited Documents

The top 10 articles that are locally cited most frequently are shown in Table 7 and Figure 6b. There are 11 local citations for "Mahalingam U's" 2019 submission to the International Conference on Wireless Communications, Signal Processing, Networking, and WISPNET. Similar to this, Ayele Tw's article from the 2018 International Conference on Inventive Communication and Computational Technologies (ICICCT) has gotten nine local citations, highlighting its influence in that particular community. Pasupuleti Vr's work from the 2020 International Conference on Advanced Computing and Communication Systems (ICACCS) has also received seven local citations, demonstrating its importance in the neighbourhood. A total of four local citations have been made for "Yarragunta S's" contribution to the 2021 International Conference on Intelligent Computing and Control Systems (ICICCS). Three local citations have been awarded to Sur S's work at the IEEE International Conference on Convergence in Engineering (ICCE), and

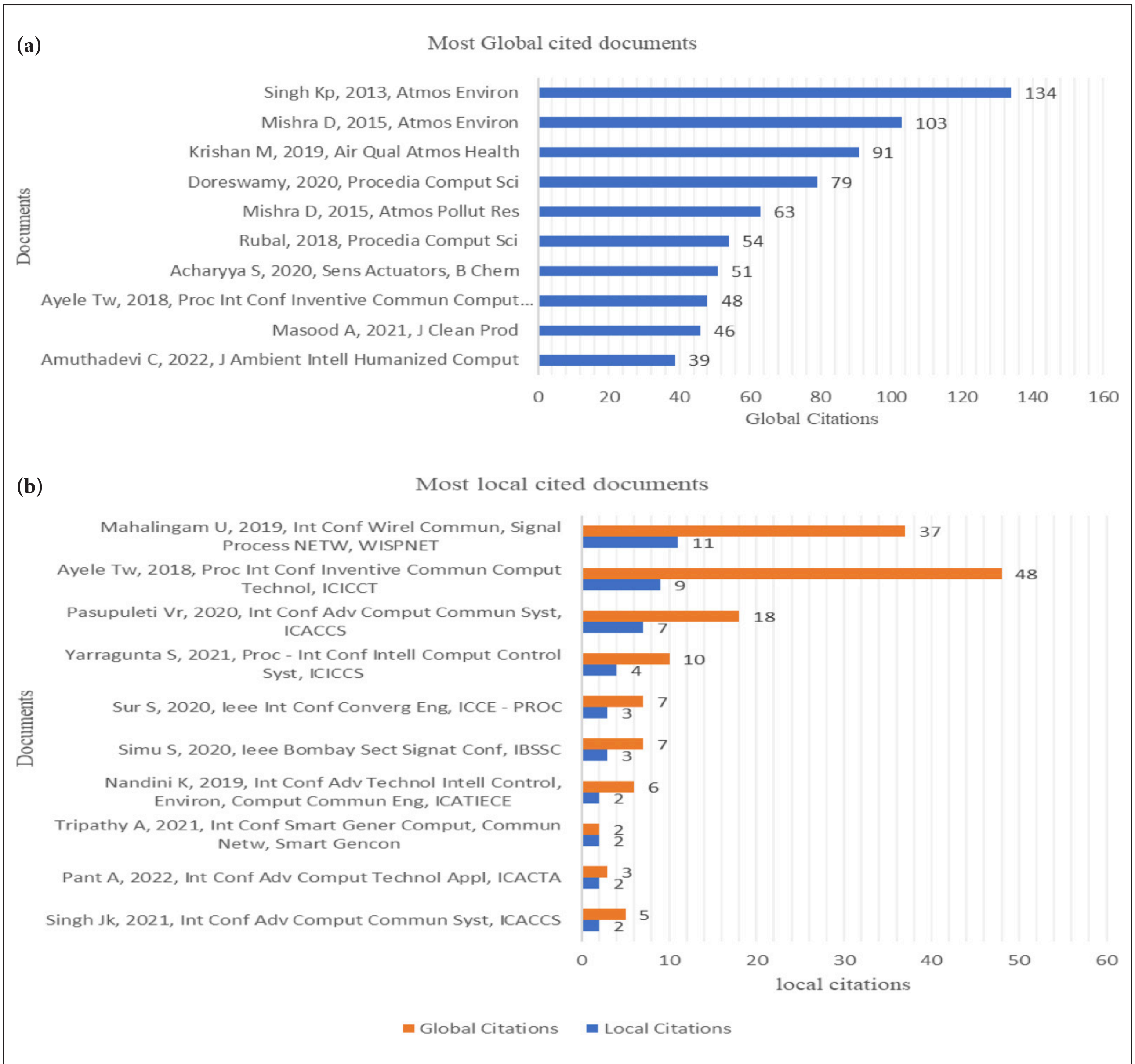

Figure 6. **(a)** Documents citation: Top 10 most global cited documents. **(b)** Documents citation: Top 10 Most local cited documents.

Table 6. Top 10 most global cited documents

Paper	DOI	Total citations	TC per year	Normalized TC	Ref
Singh Kp, 2013, Atmos Environ	10.1016/j.atmosenv.2013.08.023	134	12.18	1.00	[60]
Mishra D, 2015, Atmos Environ	10.1016/j.atmosenv.2014.11.050	103	11.44	1.75	[61]
Krishan M, 2019, Air Qual Atmos Health	10.1007/s11869-019-00696-7	91	18.20	8.87	[62]
Doreswamy, 2020, Procedia Comput Sci	10.1016/j.procs.2020.04.221	79	19.75	5.77	[63]
Mishra D, 2015, Atmos Pollut Res	10.5094/APR.2015.012	63	7.00	1.07	[64]
Rubal, 2018, Procedia Comput Sci	10.1016/j.procs.2018.05.094	54	9.00	1.71	[65]
Acharyya S, 2020, Sens Actuators, B Chem	10.1016/j.snb.2020.128484	51	12.75	3.72	[66]
Ayele Tw, 2018, Proc Int Conf Inventive Commun Comput Technol, Icicct	10.1109/ICICCT.2018.8473272	48	8.00	1.52	[67]
Masood A, 2021, J Clean Prod	10.1016/j.jclepro.2021.129072	46	15.33	8.42	[68]
Amuthadevi C, 2022, J Ambient Intell Humanized Comput	10.1007/s12652-020-02724-2	39	19.50	15.70	[69]

three local citations have been awarded to Simu S's paper that was presented at the IEEE Bombay Section Signature Conference (IBSSC) in 2020. " Two local citations for Singh Jk's work from the International Conference on Advanced Computing and Communication Systems (ICACCS) in 2021 attest to its importance in the neighbourhood. Similar to Pant A, Tripathy A's work at the International Conference on Smart Generation in Computing, Communication, and Networking (Smart Gencon) in 2021 received two local citations, as did Pant A's contribution to the International Conference on Advanced Computing Technology and Applications (ICACTA) in 2022. Last but not least, Nandini K.'s article won two local citations for its presentation at the 2019 International Conference on Advanced Technology, Intelligent Control, Environment, Computing, and Communication Engineering (ICATIECE). These locally referenced works indicate their significance within certain geographical contexts and reflect their influence on the regional academic scene.

The Local Citations to Worldwide Citations Ratio (LC/GC Ratio) is a useful indicator that illustrates how important documents are in relation to one another in local contexts versus on a worldwide scale. This ratio shows how much of a document's influence is felt locally or regionally compared to how much of an impact it has globally. A lower ratio means that the text's influence is more evenly spread across local and international audiences, whereas a greater LC/GC Ratio indicates that the work holds considerably more relevance within its local community than its global recognition. We can see that the LC/GC Ratios for each document differ when we look at the examples that have been given. For instance, "Tripathy A's" paper from the 2021 International Conference on Smart Generation in Computing, Communication, and Networking (Smart Gencon) has a remarkable LC/GC Ratio of 100.00%, indicating that this paper's impact is solely local and all of its citations are from the same local community. However, "Pasupuleti Vr's" work from the International Conference on Advanced Computing and Communication Systems (ICACCS) in 2020 has an LC/GC Ratio of 38.89%, indicating that it keeps a significant local effect while also attracting attention on a global level. The LC/GC Ratio aids in comprehending the document's reach and resonance throughout distinct academic communities and offers a nuanced view on the value of research within particular contexts.

Analysis and Co-Occurrence Network of Keywords

The co-occurrence of 808 author keywords was examined in order to use machine learning to highlight the research hotspots in the air pollution forecast area. The top 10 author keywords, along with the number of times each term appeared, are shown in Figure 7a. Notably, the term "machine learning" is mentioned 127 times, indicating its importance. The term "air pollution" appears 78 times, showing that it is frequently brought up. 41 times, the term "air quality index" is mentioned, most likely in connection with tracking and rating air quality. Both "air quality" and "deep learning" occur 32 times, indicating an emphasis on enhancing air quality and utilising cutting-edge machine learning methods. 29 instances of the word "prediction" are present, presumably indicating a focus on predictive modelling. There are 23 references to "LSTM," a type of neural network that may be related to time-series analysis. The terms "forecasting" and "random forest" both appear 22 times, indicating that both methods are probably used in air quality investigations. Finally, the phrase "air quality index (AQI)" is cited 18 times, referring to a particular metric usually brought up while discussing how to assess air quality.

The VOSviewer co-occurrence analysis demonstrates significant partnerships between authors' terms in the fields of air pollution, air quality, machine learning, forecasting, and related topics [54]. 50 of the 808 keywords meet the requirement of having at least five occurrences. These terms serve as hubs for study and information sharing in the field. Figure 7b depicts the author's Keywords as a co-occurrence network, and Figure 7c displays a co-occurrence overlay network. The number of times the highlighted terms appeared in the text was represented by the size of the circles. The larger the circle, the more the author's keyword has been co-selected in the literature on air pollution prediction. The distances between the elements of each pair were used to visually show the similarity and relative intensity of each topic. Different phrase clusters were given different circle colours. 78 instances and a total connection strength of 170 support the research on "air pollution," which is an important issue. This subject includes a number of different elements, including "air pollution forecasting," "monitoring," "prediction," and "particulate matter." The terms "air quality" and its different aspects, such as "air quality index (AQI)," "indoor air quality," and "PM2.5," highlight how crucial it is to comprehend and improve air quality. With a startling 127 occurrences and a total connection strength of 260, machine learning appears to be a crucial technology. It is a crucial tool for tackling the problems caused by air pollution. The use of cutting-edge algorithms in the analysis and forecasting of air quality trends is highlighted by terms like "deep learning," "LSTM," "random forest," "regression," "support vector machine (SVM)," and "XGBoost". Interdisciplinary fields are included in collaborative initiatives as well. The terms "artificial intelligence," "internet of things (IoT)," "smart city," and "COVID-19" show how technology advancement, urban growth, and public health connect. Additionally, techniques like "time series forecasting" and "ensemble learning" show that academics are focused on reliable model performance and precise forecasts. The VOSviewer co-occurrence analysis highlights a dynamic network of author keyword collaborations, highlighting the complicated interplay between air pollution research, machine learning, data analysis methods, and rising trends like IoT. These partnerships promote information transfer and development, which eventually helps people make well-informed decisions and find solutions to urgent environmental problems [79].

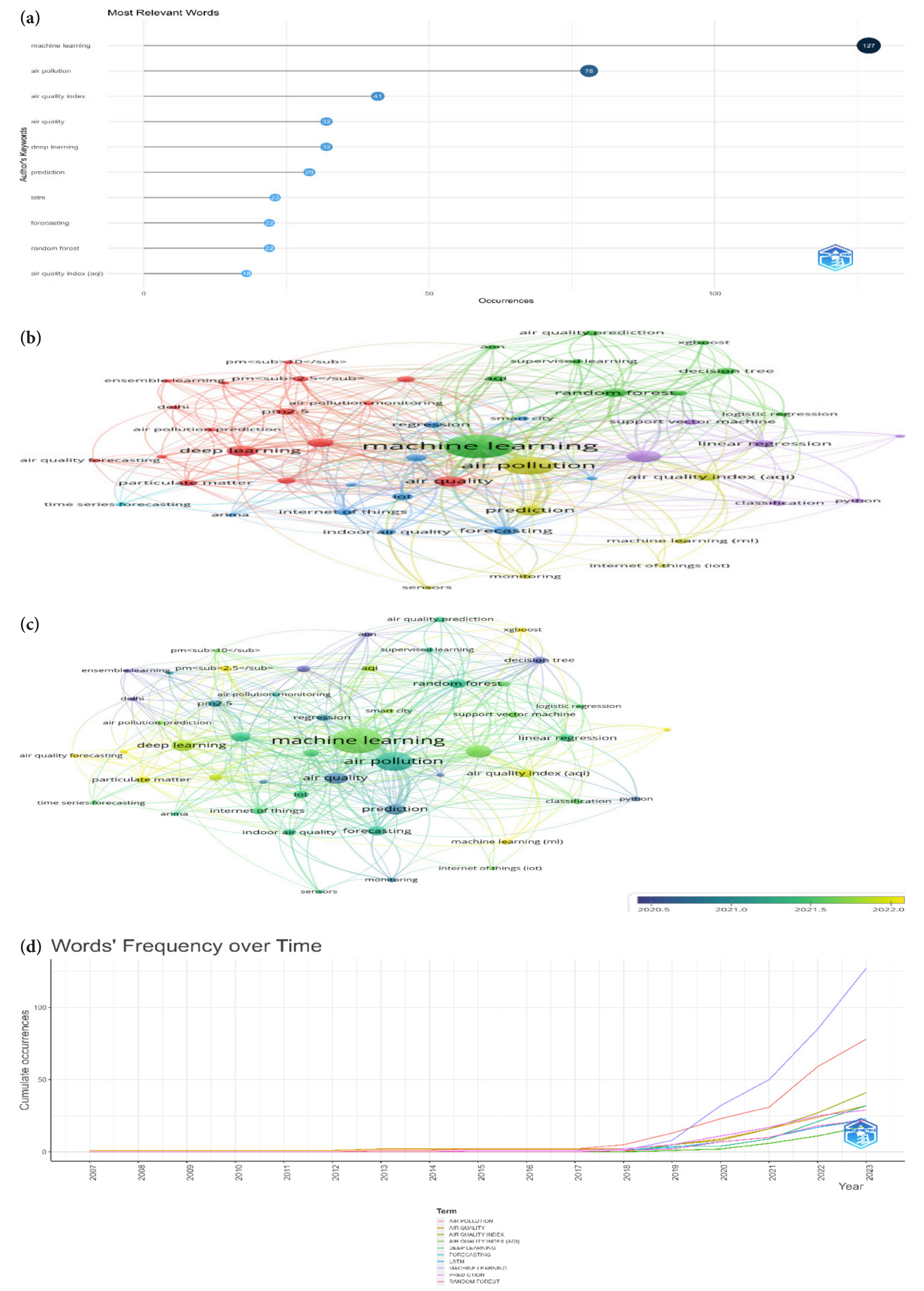

Figure 7. **(a)** Co-occurrence network of top 10 most relevant authors keywords. **(b)** Co-occurrence network visualization of author's keywords. **(c)** Co-occurrence overlay network of author keyword. **(d)** Words frequency over time.

Table 7. Top 10 most local cited documents

Document	DOI	Year	Local citations	Global citations	LC/GC ratio (%)	Normalized local citations	Normalized global citations	Ref
Mahalingam U, 2019, Int Conf Wirel Commun, Signal Process NETW, WISPNET	10.1109/WiSPNET45539.2019.9032734	2019	11	37	29.73	17.47	3.61	[70]
Ayele Tw, 2018, Proc Int Conf Inventive Commun Comput Technol, ICICCT	10.1109/ICICCT.2018.8473272	2018	9	48	18.75	4.91	1.52	[67]
Pasupuleti Vr, 2020, Int Conf Adv Comput Commun Syst, ICACCS	10.1109/ICACCS48705.2020.9074431	2020	7	18	38.89	18.05	1.31	[71]
Yarragunta S, 2021, Proc - Int Conf Intell Comput Control Syst, ICICCS	10.1109/ICICCS51141.2021.9432078	2021	4	10	40.00	23.20	1.83	[72]
Simu S, 2020, Ieee Bombay Sect Signat Conf, IBSSC	10.1109/IBSSC51096.2020.9332184	2020	3	7	42.86	7.74	0.51	[73]
Sur S, 2020, Ieee Int Conf Converg Eng, ICCE - PROC	10.1109/ICCE50343.2020.9290698	2020	3	7	42.86	7.74	0.51	[74]
Singh Jk, 2021, Int Conf Adv Comput Commun Syst, ICACCS	10.1109/ICACCS51430.2021.9441902	2021	2	5	40.00	11.60	0.91	[75]
Pant A, 2022, Int Conf Adv Comput Technol Appl, ICACTA	10.1109/ICACTA54488.2022.9753636	2022	2	3	66.67	60.67	1.21	[76]
Tripathy A, 2021, Int Conf Smart Gener Comput, Commun Netw, Smart Gencon	10.1109/SMARTGENCON51891.2021.9645787	2021	2	2	100.00	11.60	0.37	[77]
Nandini K, 2019, Int Conf Adv Technol Intell Control, Environ, Comput Commun Eng, ICATIECE	10.1109/ICATIECE45860.2019.9063845	2019	2	6	33.33	3.18	0.58	[78]

Table 8. Author keywords cumulative frequency over time

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Machine learning	0	0	0	0	0	0	0	0	0	0	0	1	8	32	50	85	127
Air pollution	0	0	0	0	0	0	1	1	2	2	2	5	13	23	31	59	78
Air Quality Index	0	0	0	0	0	0	0	0	0	0	0	0	5	8	16	27	41
Air quality	1	1	1	1	1	1	2	2	2	2	2	2	5	9	16	24	32
Deep learning	0	0	0	0	0	0	0	0	0	0	0	1	4	4	9	21	32
Prediction	0	0	0	0	0	0	0	0	1	1	1	2	5	11	17	25	29
LSTM	0	0	0	0	0	0	0	0	0	0	0	1	3	7	10	17	23
Forecasting	0	0	0	0	0	0	0	0	0	0	0	1	3	7	10	17	22
Random forest	0	0	0	0	0	0	0	0	0	0	0	1	2	7	10	18	22
Air Quality Index	0	0	0	0	0	0	0	0	0	0	0	0	1	2	6	11	18

Table 9. Top 10 most global cited documents

S. No.	Type of modelling/approach	Location(s)	Air pollutants examined	Study period	Ref.
1	A three-layer neural network model with a hidden recurrent layer	Delhi, India	SO_2	1997–1998	[80]
2	Neural network model with backpropagation	Kolkata, India	NO_2	1997–2002	[81]
3	ANN	New Delhi, India	NO_2	Two-year data	[82]
4	Neuro-fuzzy models	Delhi City, India	CO	One-hour average CO concentration data have been obtained from CPCB for a period of 2 years, i.e., January 2004 to December 2005	[83]
5	Neural networks (NN) and multiple-regression (MR) analysis.	New Delhi, India	O_3	1999–2004	[84]
6	Recurrent neural network model	New Delhi, India	CO; NO_2; NO; O_3; SO_2; $PM_{2.5}$	January 2009–June 2010 (1.5 years)	[85]
7	ANN-PCA	Kolkata, India	O_3	1997–2002	[86]
8	Partial least squares regression (PLSR), multivariate polynomial regression (MPR), and artificial neural network (ANN)	Lucknow, India	PM_{10}; SO_2; NO_2	Fve years (2005–2009)	[87]
9	ANN–MLP (Multi layer perceptron) Forecasting model	Agra, India	NO_2	Central Pollution Control Board (CPCB) at Sanjay Place, Agra, India sampled the relevant data. The sampling period was 18 November to 27 November 2013 at Sanjay Place, Agra (10 days)	[64]
10	TSA (Time series analysis), ANN, ANFIS (Adaptive neuro-fuzzy inference system)	Kolkata, India	NO_2	CPCB for the year 2010–2018	[88]

Analysis and Co-Occurrence of Authors Keyword Over Time

The dynamics of word recurrence through time provide fascinating new perspectives on the changing trends and emphases in the area. The cumulative frequency of the top 10 writers' keywords is shown in Table 8 and Figure 7d. Terms like "Machine Learning" and "Air Pollution" had a minimal presence in earlier years, such as 2007 to 2011, representing a period of relatively restrained interest in these subjects. However, as time goes on, particularly after 2012, there is a noticeable increase in debates about "Air Pollution," reflecting a growing awareness of environmental issues. The term "Air Quality Index" began to acquire popularity in 2013, and up to 2023, its total mentions grew steadily. This increased trend is consistent with the growing initiatives to thoroughly monitor and effectively inform the public about air quality conditions. In a similar vein, the word "Air Quality" first appeared in 2012 and has a steady growth pattern, underscoring the ongoing interest in analysing and resolving problems linked to air quality. Around 2018, "Deep Learning" and "Prediction" started to become noticeable in the conversation, and their combined appearances continued to rise over the following years. This problem reflects the growing use of complex methods like deep learning for predictive modelling. In particular, "LSTM," a specialised deep learning architecture, acquired popularity starting in 2018, indicating an increase in its use across numerous domains, including air quality prediction. In 2018, terms like "Forecasting" and "Random Forest" started gaining attention as well, indicating a growing interest in predictive modelling techniques. It's interesting to note that the cumulative incidence pattern of "Random Forest" closely resembles that of "Machine Learning," highlighting its importance as a prominent strategy within the larger environment. In conclusion, the cumulative dynamics of word occurrences provide a vivid account of the evolving research scene across time. This story highlights how important subjects including machine learning, deep learning, air pollution, and predictive modelling have become the preeminent areas of study and application in the field of air quality research.

Analysis of Funding Agency

Organisations that are essential to advancing research and innovation in India are highlighted in the top 10 Indian funding agencies list (Fig. 8). The Ministry of Electronics

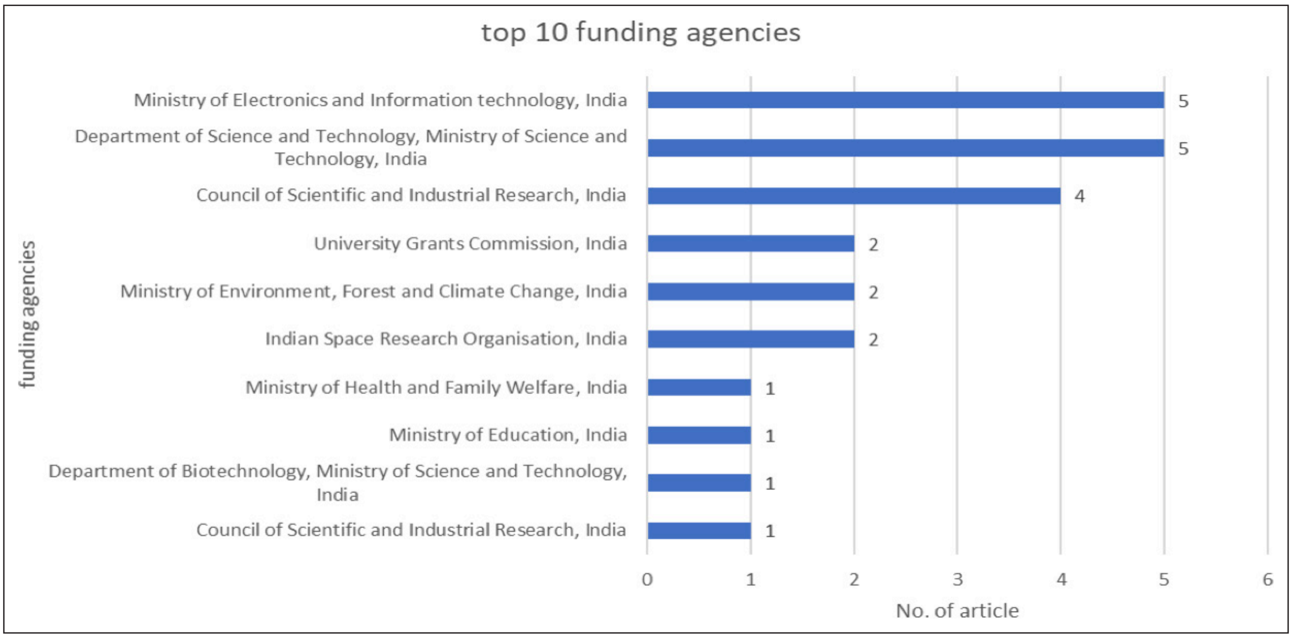

Figure 8. Top 10 funding agency.

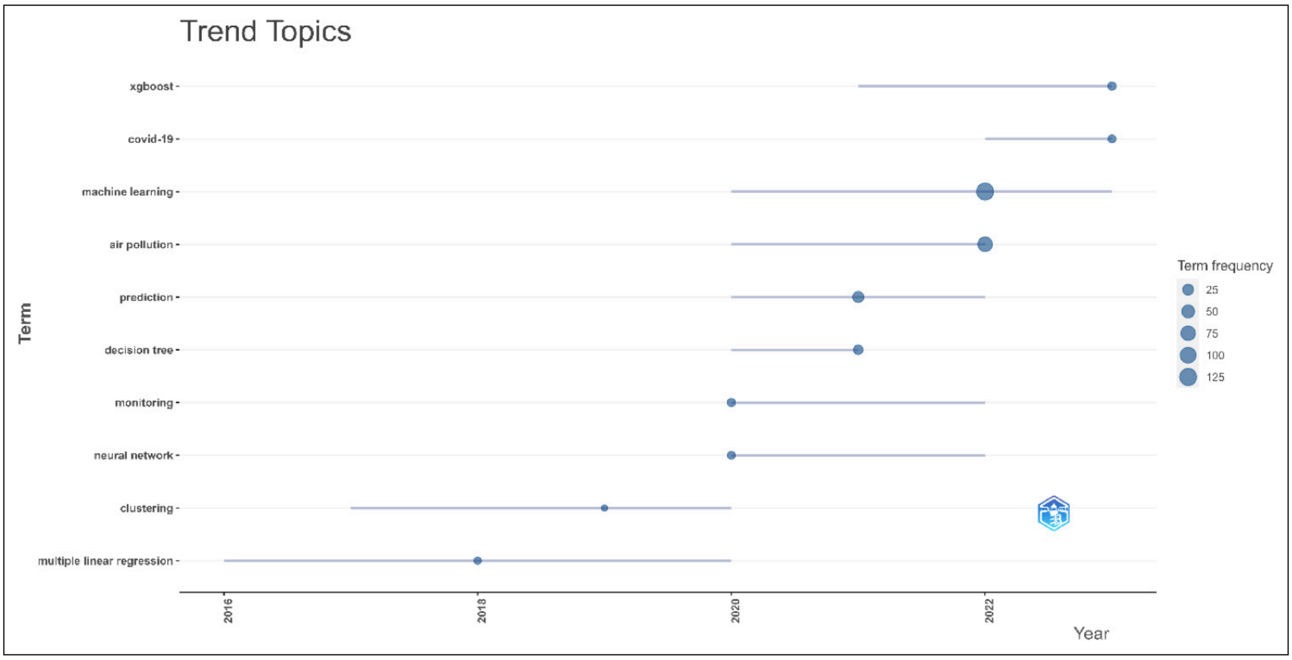

Figure 9. Trendy topics from 2016–2022.

and Information Technology and the Department of Science and Technology are at the top of the list, each receiving five units of funding, underscoring their dedication to advancing technology. With four financing organisations, the Council of Scientific and Industrial Research has a major presence and is clearly supporting numerous research projects. Two funding units have been given to the Indian Space Research Organisation, highlighting its contribution to space-related research. The Ministry of Education, the Ministry of Health and Family Welfare, the University Grants Commission, the Department of Biotechnology, and the Ministry of Environment, Forestry, and Climate Change all allocate one unit of funding each to show their commitment to advancing various fields of research in the nation. Together, these funding organisations create an active research ecosystem in India that encourages a variety of scientific inquiry and advancement.

Analysis of Machine Learning Techniques on Forecasting Air Quality

The Table 9 gives a brief summary of various air pollution modeling methodologies and techniques that have been adjusted for Indian settings. It covers study locations, air contaminants evaluated, and study times, offering light on key approaches for evaluating and forecasting air quality in the context of India.

The Future of Research into Machine Learning in the Field of Air Pollution

Figure 9 displays co-occurrence network mappings organised by topic area or publication date to highlight current hot topics and potential future directions in the study of air pollution. The analysis of trends in a variety of topics offers important insights into how research has changed over time. "Multiple Linear Regression" was one of the issues that received the most attention, with its frequency rising from 2016 to 2020. This indicates that the use of linear regression models for multi-dimensional analysis in air quality studies is expanding. From 2017 to 2020, "clustering" was consistently present, demonstrating a continued interest in combining related air quality data points for further study. The years 2020 to 2022 will see a clear emphasis on cutting-edge methods, particularly "Monitoring" and "Neural networks," demonstrating a heightened desire for precise and immediate data analysis. With a significant frequency of 29, the word "Prediction" appears as the prominent subject. Its prominence increases from 2020 to 2022, underscoring a spike in initiatives to forecast air quality situations using predictive models. Similar to how "Decision Tree" receives a lot of attention between 2020 and 2021, this is a reflection of its function in offering understandable insights into complicated air quality information. A critical turning point occurs in 2020, when "Machine Learning" undergoes a dramatic increase in frequency and dominates the conversation through 2023. This highlights a paradigm shift towards applying machine learning methods to problems related to air quality. In line with this, "Air Pollution" hits its peak in 2020 and continues to air often through 2022, highlighting continuous efforts to understand and reduce pollution levels. From 2021 to 2023, "XGBoost," a well-known boosting algorithm, emerges strongly, indicating its use for improving prediction accuracy. Last but not least, "COVID-19" becomes an important topic starting in 2022, showing the significance of the global context on air quality studies, maybe in response to the effects of the pandemic on environmental conditions.

CONCLUSION

Researchers can employ bibliometric analysis to consider many criteria, including the productivity of the field, different nations, the most relevant journals, authors, institutions, and so on, when making decisions about what and where to publish. Additionally, it is beneficial to examine the trends and patterns in publications in order to gain insight into the nature and productivity of a certain academic field. This research endeavour pertaining to the prediction of air pollution through the utilisation of machine learning encompasses a corpus of around 326 scholarly articles authored by 984 scholars and disseminated throughout 231 academic journals spanning the temporal domain from 2007 to 2023. Due to the significant significance of this area of research, it is unsurprising that there has been a notable increase in the number of articles published on these subjects since 2018, exhibiting an annual growth rate of 32.1%. The findings indicate that prominent scholarly journals have assumed a prominent role in advancing the field of machine learning-based air pollution prediction. Notably, Atmospheric Environment, with 263 citations; Procedia Computer Science, with 251 citations; Atmospheric Pollution Research, with 233 citations; and Air Quality, Atmosphere, and Health, with 93 citations, have emerged as key contributors in this area of research, as evidenced by their respective Total Citation Index scores. Jadavpur University, with its 12 articles, and IIT Delhi, with its 10 articles, are widely regarded as two of the most prestigious academic institutions in India. Singh Kp's (2013) essay titled "Atmospheric Environment" holds the highest position on the list, having accumulated a total of 134 citations. The Ministry of Electronics and Information Technology, along with the Department of Science and Technology, each earn a total of five units, which serves as a clear indication of their dedication and support towards the advancement of technology. The findings of the authors' analysis indicate that machine learning, air pollution, and air quality index are the most frequently occurring keywords in the keyword co-occurrence network mappings, with 127, 78, and 41 occurrences, respectively. The inclusion of phrases such as "air pollution," "machine learning," and "air quality" in our Scopus database search queries indicates that these topics are frequently examined in scholarly investigations pertaining to machine learning-driven predictions of air pollution. One of the emerging concepts in this field that indicates potential future advancements is the integration of XGBoost, neural networks, and machine learning techniques. However, there remain certain gaps that require completion. The necessity for further comparative studies in the aforementioned nations that are now underrepresented is arguably of utmost significance. Due to the extensive body of literature on the subject of air pollution prediction research and the inherent limitations of relying on a single database to offer a comprehensive overview of a research domain that holds substantial global significance, it is worthwhile to explore alternative avenues for investigation. These may include the integration of the two primary bibliographic databases, namely Web of Science (WoS) and Scopus, or the utilisation of supplementary databases. This paper aims to provide professionals, scholars, and worldwide policymakers with an understanding of the current status of the "air pollution prediction using machine learning" field while also highlighting certain areas that necessitate further investigation. This review offers a comprehensive guide for writers, reviewers, and journal editors to consider while contemplating their future work, its value, and the various challenges that may arise in the publication process.

DATA AVAILABILITY STATEMENT

The author confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

USE OF AI FOR WRITING ASSISTANCE

Not declared.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] World Health Organization (WHO), "Exposure & health impacts of air pollution," Available at: https://www.who.int/teams/environment-climate-change-and-health/air-quality-energy-and-health/health-impacts/exposure-air-pollution#:~:text=The%20combined%20or%20joint%20effects,cancer%20and%20acute%20respiratory%20infections. Accessed on Jul 19, 2024.

[2] Y. C. Hong, J. T. Lee, H. Kim, and H. J. Kwon, "Air pollution: A new risk factor in ischemic stroke mortality," Stroke, Vol. 33(9), pp. 2165–2169, 2002. [CrossRef]

[3] R. Ruckerl, A. Ibald-Mulli, W. Koenig, A. Schneider, G. Woelke, J. Cyrys…, and A. Peters, "Air pollution and markers of inflammation and coagulation in patients with coronary heart disease," American Journal of Respiratory and Critical Care Medicine, Vol. 173(4), pp. 432–441, 2006. [CrossRef]

[4] Z. J. Andersen, "Chronic obstructive pulmonary disease and long-term exposure to traffic-related air pollution: A cohort study," American Journal of Respiratory and Critical Care Medicine, Vol. 183(4), pp. 455–461, 2011. [CrossRef]

[5] F. Nyberg, P. Gustavsson, L. Jarup, T. Bellander, N. Berglind, R. Jakobsson, and G. Pershagen, "Urban air pollution and lung cancer in Stockholm," Epidemiology, Vol. 11(5), pp. 487–495, 2000. [CrossRef]

[6] M. Ezzati, and D. M. Kammen, "Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: An exposure-response study," Lancet, Vol. 358(9282), pp. 619–624, 2001. [CrossRef]

[7] L. A. Darrow, M. Klein, W. D. Flanders, J. A. Mulholland, P. E. Tolbert, and M. J. Strickland, "Air pollution and acute respiratory infections among children 0–4 years of age: An 18-year time-series study," The American Journal of Epidemiology, Vol. 180(10), pp. 968–977, 2014. [CrossRef]

[8] D. Loomis, W. Huang, and G. Chen, "The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: Focus on China," Chinese Journal of Cancer, Vol. 33(4), pp. 189–196, 2014. [CrossRef]

[9] D. Loomis, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa…, and K. Straif; International Agency for Research on Cancer Monograph Working Group, "The carcinogenicity of outdoor air pollution," The Lancet Oncology, Vol. 14(13), pp. 1262–1263, 2013. [CrossRef]

[10] Mokhtari, W. Bechkit, H. Rivano, and M. R. Yaici, "Uncertainty-aware deep learning architectures for highly dynamic air quality prediction," IEEE Access, Vol. 9, pp. 14765–14778, 2021. [CrossRef]

[11] M. Kampa, and E. Castanas, "Human health effects of air pollution," Environmental Pollution, Vol. 151(2), pp. 362–367, 2008. [CrossRef]

[12] E. Tagaris, K. J. Liao, A. J. DeLucia, L. Deck, P. Amar, and A. G. Russell, "Potential impact of climate change on air pollution-related human health effects," Environmental Science and Technology, Vol. 43(13), pp. 4979–4988, 2009. [CrossRef]

[13] M. I. Qureshi, A. M. Rasli, U. Awan, J. Ma, G. Ali, A. Alam…, and K. Zaman, "Environment and air pollution: Health services bequeath to grotesque menace," Environmental Science and Pollution Research, Vol. 22, pp. 3467–3476, 2015. [CrossRef]

[14] H. Orru, K. L. Ebi, and B. Forsberg, "The interplay of climate change and air pollution on health," Currrent Environmental Health Reports, Vol. 4, pp. 504–513, 2017. [CrossRef]

[15] H. Du, D. Liu, Z. Lu, J. Crittenden, G. Mao, S. Wang, and H. Zou, "Research development on sustainable urban infrastructure from 1991 to 2017: A bibliometric analysis to inform future innovations," Earth's Future, Vol. 7(7), pp. 718–733, 2019. [CrossRef]

[16] D. L. Crouse, N. A. Ross, and M. S. Goldberg, "Double burden of deprivation and high concentrations of ambient air pollution at the neighbourhood scale in Montreal, Canada," Social Science & Medicine, Vol. 69(6), pp. 971–981, 2009. [CrossRef]

[17] J. Kerckhoffs, G. Hoek, L. Portengen, B. Brunekreef, and R. C. H. Vermeulen, "Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces," Environmental Science and Technology, Vol. 53(3), pp. 1413–1421, 2019. [CrossRef]

[18] W. Wang, C. Men, and W. Lu, "Online prediction model based on support vector machine," Neurocomputing, Vol. 71(4–6) pp. 550–558, 2008. [CrossRef]

[19] R. S. Batth, M. Gupta, K. S. Mann, S. Verma, and A. Malhotra, "Comparative study of tdma-based mac protocols in vanet: A mirror review," Proceedings of the 2019 International Conference on Innovative Computing and Communications (ICICC). Ostrava, Czech Republic, 2019.

[20] M. Kaur, and S. Verma, "Flying ad-hoc network (FANET): Challenges and routing protocols," Journal of Computational and Theoretical Nanoscience Vol. 17(6) pp. 2575–2581, 2020. [CrossRef]
[21] T. Sharma, and S. Verma, "Prediction of heart disease using cleveland dataset: A machine learning approach," International Journal of Recent Research Aspects, Vol. 4(3), pp. 17–21, 2017.
[22] X. Tian, Y. Huang, S. Verma, M. Jin, U. Ghosh, K. M. Rabie, and D. T. Do, "Power allocation scheme for maximizing spectral efficiency and energy efficiency tradeoff for uplink NOMA systems in B5G/6G," Physical Communication, Vol. 43, Article 101227, 2020. [CrossRef]
[23] G. Ghosh, M. Sood, and S. Verma, "Internet of things based video surveillance systems for security applications," Journal of Computational and The-oretical Nanoscience, Vol. 17(6), pp. 2582–2588, 2020[CrossRef]
[24] V. P. Diodato, and P. Gellatly, "Dictionary of Bibliometrics," Routledge, 2013
[25] R. N. Broadus, "Toward a definition of 'bibliometrics,'" Scientometrics, Vol. 12, pp. 373–379, 1987. [CrossRef]
[26] Pritchard, "Statistical bibliography or bibliometrics," Journal of Documentation, Vol. 25(4), pp. 348–349, 1969. [CrossRef]
[27] M. A. Koseoglu, R. Rahimi, F. Okumus, and J. Liu, "Bibliometric studies in tourism," Annals of Tourism Research Vol. 61(1) pp. 180–198, 2016. [CrossRef]
[28] Y. Yu, Y. Li, Z. Zhang, Z. Gu, H. Zhong, Q. Zha…, and E. Chen, "A bibliometric analysis using VOSviewer of publications on COVID-19," Annals of Translational Medicine, Vol. 8(13), pp. 816–816, 2020. [CrossRef]
[29] P. Hallinger, and J. Kovačević, "A bibliometric review of research on educational administration: Science mapping the literature, 1960 to 2018," Review of Educational Research, Vol. 89(3), pp. 335–369, 2019. [CrossRef]
[30] P. Hallinger, and C. Chatpinyakoop, "A bibliometric review of research on higher education for sustainable development, 1998–2018," Sustainability, Vol. 11(8), Article 2401, 2019. [CrossRef]
[31] E. A. Abafe, Y. T. Bahta, and H. Jordaan, "Exploring biblioshiny for historical assessment of global research on sustainable use of water in agriculture," Sustainability, Vol. 14(17), Article 10651, 2022. [CrossRef]
[32] H. Babbar, S. Rani, M. Masud, S. Verma, D. Anand, and N. Jhanjhi, "Load balancing algorithm for migrating switches in software-defined vehicular networks," Computers, Materials & Continua, Vol.67(1), pp. 1301–1316, 2021. [CrossRef]
[33] S. Kumar, R. Shanker, and S. Verma, "Context aware dynamic permission model: A retrospect of privacy and security in android system," Procedings of the 2018 International Conference on Intelligent Cir-cuits and Systems (ICICS). Phagwara, India, 2018.
[34] M. Kumar, K. S. Raju, D. Kumar, N. Goyal, S. Verma, and A. Singh, "An efficient framework using visual recognition for IoT based smart city surveillance," Multimedia Tools and Applications, Vol. 80, pp. 31277–31295, 2021. [CrossRef]
[35] G. Yang, M. A. Jan, A. U. Rehman, M. Babar, M. M. Aimal, and S. Verma, "Interoperability and data storage in internet of multimedia things: Investigating current trends, research challenges and future directions," IEEE Access, Vol. 8, pp. 124382–124401, 2020. [CrossRef]
[36] S. Dash, S. Verma, Kavita, S. Bevinakoppa, M. Wozniak, J. Shafi, and M. F. Ijaz, "Guidance image-based enhanced matched filter with modified thresholding for blood vessel extraction," Symmetry (Basel), Vol. 14(2), Article 194, 2022. [CrossRef]
[37] V. Dogra, A. Singh, S. Verma, Kavita, N. Z. Jhanjhi, and M. N. Talib, "Analyzing DistilBERT for sentiment classification of banking financial news," Proceedings of the 2021 Intelligent Computing and Innovation on Data Science (ICTIDS). Ahmedabad, India, 2021. [CrossRef]
[38] Y. Rybarczyk, and R. Zalakeviciute, "Machine learning approaches for outdoor air quality modelling: A systematic review," Applied Sciences, Vol. 8(12), Article 2570, 2018. [CrossRef]
[39] Q. Guo, M. Ren, S. Wu, Y. Sun, J. Wang, Wang Q…, and Y. Chen, "Applications of artificial intelligence in the field of air pollution: A bibliometric analysis," Frontiers in Public Health, Article 2972, 2022. [CrossRef]
[40] L. Bai, J. Wang, X. Ma, and H. Lu, "Air pollution forecasts: An overview," International Journal of Environmental Research and Public Health, Vol. 15(4), Article 780, 2018. [CrossRef]
[41] X. Li, Y. Choi, B. Czader, A. Roy, H. Kim, B. Lefer, and S. Pan, "The impact of observation nudging on simulated meteorology and ozone concentrations during DISCOVER-AQ 2013 Texas campaign," Atmospheric Chemistry and Physics, Vol. 16(5), pp. 3127–3144, 2016. [CrossRef]
[42] C. Vitolo, Y. Elkhatib, D. Reusser, C. J. A. Macleod, and W. Buytaert, "Web technologies for environ-mental Big Data," Environmental Modelling & Soft-ware, Vol. 63(3), pp. 185–198, 2015. [CrossRef]
[43] Z. Zong, Y. Chen, C. Tian, Y. Fang, X. Wang, G. Huang…, G. Zhang,"Radiocarbon-based impact assessment of open biomass burning on regional carbonaceous aerosols in North China," Science of the Total Environment Vol. 518–519, pp. 1–7, 2015. [CrossRef]
[44] S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, "A review of artificial neural network models for ambient air pollution prediction," Environmental Modelling & Software, Vol. 119, pp. 285–304, 2019. [CrossRef]
[45] P. Guo, W. Tian, H. Li, G. Zhang, and J. Li, "Global characteristics and trends of research on construction dust: Based on bibliometric and visualized analysis," Environmental Science and Pollution Research, Vol. 27, pp. 37773–37789, 2020. [CrossRef]

[46] Y. Hou, and Z. Shen, "Research Trends, hotspots and frontiers of ozone pollution from 1996 to 2021: A review based on a bibliometric visualization analysis," Sustainability, Vol. 14(17), Article 10898, 2022. [CrossRef]

[47] S. Jain, N. Kaur, S. Verma, Kavita, A. S. M. S. Hosen, and S. S. Sehgal, "Use of machine learning in air pollution research: A bibliographic perspective," Electronics, Vol. 11(21), Article 3621, 2022. [CrossRef]

[48] Y. Li, Z. Sha, A. Tang, K. Goulding, and X. Liu, "The application of machine learning to air pollution research: A bibliometric analysis," Ecotoxicology and Environmental Safety, Vol. 257, Article 114911, 2023. [CrossRef]

[49] K. Mehmood, Y. Bao, Saifullah, W. Cheng, M. A. Khan, N. Siddique…, R. Naidu, "Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives," Journal of Cleaner Production, Vol. 379(Part 2), Article 134656, 2022. [CrossRef]

[50] J. F. Velasco-Muñoz, J. A. Aznar-Sánchez, L. J. Belmonte-Ureña, and I. M. Román-Sánchez, "Sustainable water use in agriculture: A review of worldwide research," Sustainability, Vol. 10(4), Article 1084, 2018. [CrossRef]

[51] E. Garfield, and I. H. Sher, "New factors in the evaluation of scientific literature through citation indexing," American Documentation, Vol. 14(3), pp. 195–201, 1963. [CrossRef]

[52] Zupic, and T. Čater, "Bibliometric methods in management and organization," Organizational Research Methods, Vol. 18(3), pp. 429–472, 2015. [CrossRef]

[53] M. Aria, and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis," Journal of Informetrics, Vol. 11(4), pp. 959–975, 2017. [CrossRef]

[54] N. J. van Eck, and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," Scientometrics, Vol. 84(2), pp. 523–538, 2010. [CrossRef]

[55] N. J. Van Eck, and L. Waltman, "Visualizing Bibliometric Networks." Edited by Ding, Y., Rousseau, R., and Wolfram D. Measuring Scholarly Impact, Springer. pp. 285–320, 2014.

[56] N. J. Van Eck, and L. Waltman, "Text mining and visualization using VOSviewer," arXiv Prepr. 2011.

[57] H. Ernst, "The use of patent data for technological forecasting: The diffusion of CNC-technology in the machine tool industry," Small Business Economics, Vol. 9, pp. 361–381, 1997. [CrossRef]

[58] G. Mao, H. Hu, X. Liu, J. Crittenden, and N. Huang, "A bibliometric analysis of industrial wastewater treatments from 1998 to 2019," Environmental Pollution Vol. 275, Article 115785, 2021. [CrossRef]

[59] J. Daniels, and P. Thistlethwaite, "Measuring Scholarly Impact," Cambridge University Press, 2022.

[60] K. P. Singh, S. Gupta, and P. Rai, "Identifying pollution sources and predicting urban air quality using ensemble learning methods," Atmospheric Environment, Vol. 80, pp. 426–437, 2013. [CrossRef]

[61] D. Mishra, P. Goyal, and A. Upadhyay, "Artificial intelligence based approach to forecast PM2. 5 during haze episodes: A case study of Delhi, India," Atmospheric Environment, Vol. 102, pp. 239–248, 2015. [CrossRef]

[62] M. Krishan, S. Jha, J. Das, A. Singh, M. K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India," Air Quality Atmosphere Health, Vol. 12(8), pp. 899–908, 2019. [CrossRef]

[63] K. S. Harishkumar, K. M. Yogesh, and I. Gad, "Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models," Procedia Computer Science, Vol. 171, pp. 2057–2066, 2020. [CrossRef]

[64] D. Mishra, and P. Goyal, "Development of artificial intelligence based NO_2 forecasting models at Taj Mahal, Agra," Atmospheric Pollution Research, Vol. 6(1), pp. 99–106, 2015. [CrossRef]

[65] Rubal, and D. Kumar, "Evolving differential evolution method with random forest for prediction of Air Pollution," Procedia Computer Science, Vol. 132, pp. 824–833, 2018. [CrossRef]

[66] S. Acharyya, B. Jana, S. Nag, G. Saha, and P. K. Guha, "Single resistive sensor for selective detection of multiple VOCs employing SnO_2 hollowspheres and machine learning algorithm: A proof of concept," Sensors and Actuators B Chemical, Vol. 321, Article 128484, 2020. [CrossRef]

[67] T. W. Ayele, and R. Mehta, "Air pollution monitoring and prediction using IoT," Proceedings of the 2018 International Conference on Inventive Communication and Computational Technologies (ICICCT). New Delhi, India, 2018. [CrossRef]

[68] Masood, and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance," Journal of Cleaner Production, Vol. 322, Article 129072, 2021. [CrossRef]

[69] Amuthadevi, D. S. Vijayan, and V. Ramachandran, "Development of air quality monitoring (AQM) models using different machine learning approaches," Journal of Ambient Intelligence and Humanized Computing, Vol 13. pp. 33–34, 2021. [CrossRef]

[70] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A machine learning model for air quality prediction for smart cities," Proceedings of the 2019 International Conference on Wireless Communications Signal Processing and Networking (WISPNET). Chennai, India, 2019. [CrossRef]

[71] V. R. Pasupuleti, P. Kalyan, and H. K. Reddy, "Air quality prediction of data log by machine learning," Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India, 2020. [CrossRef]

[72] S. Yarragunta, and M. A. Nabi, "Prediction of air pollutants using supervised machine learning," Proceedings of the 5th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India, 2021. [CrossRef]

[73] S. Simu, V. Turkar, and R. Martires, "Air pollution prediction using machine learning," Proceedings of the 2020 IEEE Bombay Section Signature Conference (IBSSC). Mumbai, India, 2020. [CrossRef]

[74] S. Sur, R. Ghosal, and R. Mondal, "Air pollution hotspot identification and pollution level prediction in the City of Delhi," Proceedings of the 1st International Conference for Convergence in Engineering (ICCE), Kolkata, India, 2020. [CrossRef]

[75] J. K. Singh, and A. K. Goel, "Prediction of air pollution by using machine learning algorithm," Proceedings of the 7th International conference on advanced computing and communication Systems (ICACCS), Coimbatore, India, 2021. [CrossRef]

[76] Pant, S. Sharma, M. Bansal, and M. Narang, "Comparative analysis of supervised machine learning techniques for AQI prediction," Proceedings of the 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), Coimbatore, India, 2022. [CrossRef]

[77] Tripathy, D. Vaidya, A. Mishra, S. Bilolikar, and V. Thoday, "Analysing and predicting air quality in Delhi: Comparison of industrial and residential area," Proceedings of the 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021. [CrossRef]

[78] K. Nandini, and G. Fathima, "Urban air quality analysis and prediction using machine learning," Proceedings of the 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2019. [CrossRef]

[79] Janik, A. Ryszko, and M. Szafraniec, "Scientific landscape of smart and sustainable cities literature: A bibliometric analysis," Sustainability, vol. 12(3), Article 779, 2020. [CrossRef]

[80] B. Chelani, C. V. C. Rao, K. M. Phadke, and M. Z. Hasan, "Prediction of sulphur dioxide concentration using artificial neural networks," Environmental Modelling & Software, Vol. 17(2), pp. 159–166, 2002. [CrossRef]

[81] B. Chelani, R. N. Singh, and S. Devotta, "Nonlinear dynamical characterization and prediction of ambient nitrogen dioxide concentration," Water, Air, and Soil Pollution, Vol. 166, pp. 121–138, 2005. [CrossRef]

[82] S. M. S. Nagendra, and M. Khare, "Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions," Ecological Modelling, Vol. 190(1–2) pp. 99–115, 2006. [CrossRef]

[83] S. Jain, and M. Khare, "Adaptive neuro-fuzzy modeling for prediction of ambient CO concentration at urban intersections and roadways," Air Quality, Atmosphere & Health, Vol. 3(4), pp. 203–212, 2010. [CrossRef]

[84] Mahapatra, "Prediction of daily ground-level ozone concentration maxima over New Delhi," Environmental Monitoring and Assessment, Vol. 170, pp. 159–170, 2010. [CrossRef]

[85] Prakash, U. Kumar, K. Kumar, and V. K. Jain, "A wavelet-based neural network model to predict ambient air pollutants' concentration," Environmental Modeling & Assessment, Vol. 16, pp. 503–517, 2011. [CrossRef]

[86] S. Chattopadhyay, and G. Chattopadhyay, "Modeling and prediction of monthly total ozone concentrations by use of an artificial neural network based on principal component analysis," Pure and Applied Geophysics, Vol. 169(10) pp. 1891–1908, 2012. [CrossRef]

[87] K. P. Singh, S. Gupta, A. Kumar, and S. P. Shukla, "Linear and nonlinear modeling approaches for urban air quality prediction," Science of the Total Environment, Vol. 426, pp. 244–255, 2012. [CrossRef]

[88] R. Bhardwaj, D. Pruthi, "Development of model for sustainable nitrogen dioxide prediction using neuronal networks," International Journal of Environmental Science and Technology, Vol. 17, pp. 2783–2792, 2020. [CrossRef]