Discover Artificial Intelligence

Research

Artificial intelligence in environmental monitoring: in-depth analysis

Emran Alotaibi[1,2,3] · Nadia Nassif[2]

Received: 4 August 2024 / Accepted: 7 November 2024
Published online: 18 November 2024

Abstract
This study provides a comprehensive bibliometric and in-depth analysis of artificial intelligence (AI) and machine learning (ML) applications in environmental monitoring, based on 4762 publications from 1991 to 2024. The research highlights a notable increase in publications and citations since 2010, with China, the United States, and India emerging as leading contributors. Key areas of research include air and water quality monitoring, climate change modeling, biodiversity assessment, and disaster management. The integration of AI with emerging technologies, such as the Internet of Things (IoT) and remote sensing, has significantly expanded real-time environmental monitoring capabilities and data-driven decision-making. In-depth analysis reveals advancements in AI/ML methodologies, including novel algorithms for soil mapping, land-cover classification, flood susceptibility modeling, and remote sensing image analysis. Notable applications include enhanced air quality predictions, water quality assessments, climate impact forecasting, and automated wildlife monitoring using AI-driven image recognition. Challenges such as the "black-box" nature of AI models, the need for high-quality data in resource-constrained regions, and the complexity of real-time disaster management are also addressed. The study highlights ongoing efforts to develop explainable AI (XAI) models, which aim to improve model transparency and trust in critical environmental applications. Future research directions emphasize improving data quality and availability, fostering interdisciplinary collaborations across environmental and computer sciences, and addressing ethical considerations in AI-driven environmental management. These findings underscore the transformative potential of AI and ML technologies for sustainable environmental management, offering valuable insights for researchers and policymakers in addressing global environmental challenges.

Keywords Deep learning models · Big data analytics · IoT-enabled monitoring · Explainable AI · Remote sensing technologies · Sustainable ecosystem management

1 Introduction

1.1 Background and rationale

In recent years, the confluence of artificial intelligence (AI) and machine learning (ML) with environmental sciences has led to significant advancements in environmental monitoring and data analysis. These technologies offer powerful tools to address complex environmental challenges by enabling more accurate predictions, real-time monitoring, and the ability to analyze vast datasets that traditional methods cannot manage efficiently.

✉ Emran Alotaibi, emran.alotaibi@ku.ac.ae; Nadia Nassif, u19200817@sharjah.ac.ae | [1]Department of Civil and Environmental Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates. [2]Department of Civil and Environmental Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. [3]Sustainable Civil Infrastructure Group, Research Institute of Sciences and Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates.

Environmental challenges such as climate change, deforestation, ocean pollution, and biodiversity loss represent significant threats to global ecosystems and human health. For instance, AI and ML have been employed to model climate change impacts, such as the rise in global temperatures, increased frequency of extreme weather events, and sea level rise, which threaten coastal populations and ecosystems ([50]). Deforestation—a major driver of biodiversity loss and carbon emissions—has also been extensively monitored using AI models that analyze satellite imagery and detect illegal logging activities in real-time [54]. Ocean pollution, particularly from plastic waste, poses severe risks to marine ecosystems, and AI-driven technologies have been used to detect and track pollution in coastal waters and the open sea, identifying patterns and sources of pollution more effectively [33]. These environmental challenges highlight the urgent need for innovative solutions, and AI/ML techniques provide automated, efficient, and scalable solutions. For example, AI models can process large volumes of data from sources such as satellite imagery, sensor networks, and historical datasets, offering insights that were previously unattainable. Additionally, biodiversity loss is being addressed through AI applications in automated species identification and ecosystem monitoring, providing crucial data for conservation efforts [30, 44].

The rationale for this study stems from the growing body of literature demonstrating the transformative potential of AI and ML in addressing these complex environmental challenges. For instance, machine learning algorithms have been used to predict air pollution levels, identify changes in land use, and model climate change impacts with high precision. Similarly, AI-driven tools have enhanced the detection and management of environmental hazards such as oil spills and forest fires by providing timely and actionable information [24, 55].

Given the rapid advancements and the increasing volume of research in this area, it is crucial to systematically analyze the existing literature to identify trends, gaps, and future directions. Bibliometric analysis, which involves the quantitative assessment of academic literature, provides a robust methodology to achieve this objective. By examining publication patterns, citation networks, and keyword trends, we can gain a comprehensive understanding of the research landscape and the contributions of different researchers, institutions, and countries.

1.2 Objectives of the study

The primary objective of this study is to conduct a bibliometric analysis of the research on AI and ML techniques in environmental monitoring and data analysis. This involves several specific aims:

1. To identify and analyze the main trends in the publication of research articles related to AI and ML in environmental monitoring. This includes examining the growth in the number of publications over time and the distribution of research across different journals and disciplines.
2. To assess the geographical distribution of research outputs, identifying leading countries and regions contributing to this field.
3. To map collaboration networks to understand the dynamics of research partnerships patterns.
4. To perform a keyword analysis to identify the most frequently used terms and emerging topics in the literature. This will help in understanding the focus areas and the evolution of research themes over time.
5. To conduct a citation analysis to identify highly cited papers and influential researchers. This will provide insights into the impact and recognition of various studies within the scientific community. In addition to H-index evaluation and it's relation to the publications in the field.
6. To discuss the implications of the findings for future research and practice in environmental monitoring. This includes identifying research gaps and suggesting potential areas for further investigation.

2 Background

2.1 Overview of machine learning and artificial intelligence

ML and AI are rapidly evolving fields that have transformed various sectors, including environmental monitoring and data analysis. AI is a broad discipline that encompasses the creation of systems capable of performing tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation. Machine learning, a subset of AI, focuses on developing algorithms that enable computers to learn from and make predictions or decisions based on data [13, 59, 60].

In the context of environmental monitoring, ML algorithms can be broadly categorized into supervised, unsupervised, and reinforcement learning. Supervised learning involves training a model on a labeled dataset, which means that each training example is paired with an output label. This approach is particularly useful for tasks like predicting air quality levels or classifying types of vegetation in satellite images. Common algorithms in supervised learning include decision trees, support vector machines (SVM), and neural networks [36, 57, 58].

Unsupervised learning, on the other hand, deals with data that has no labels. The goal is to infer the natural structure present within a set of data points. This can be useful for clustering similar pollution sources or detecting anomalies in environmental data. Algorithms such as k-means clustering and principal component analysis (PCA) are typical examples of unsupervised learning techniques [16].

Reinforcement learning is a type of ML where an agent learns to make decisions by performing certain actions and receiving rewards or penalties. This approach can be applied to optimize environmental management strategies, such as developing policies for reducing emissions or managing natural resources sustainably [40].

AI techniques also include deep learning, a subset of machine learning characterized by neural networks with many layers (deep neural networks). These techniques are particularly powerful in handling large datasets and complex patterns, making them suitable for tasks such as image recognition in satellite imagery or time-series prediction in climate data [22] and environmentalremote sensing [48].

2.2 Applications in environmental monitoring and data analysis

The application of AI and ML in environmental monitoring has seen significant growth and diversification. These technologies have been deployed in various domains to address critical environmental issues, demonstrating their versatility and effectiveness.

2.2.1 Air quality monitoring

AI and ML techniques are used to predict and monitor air quality levels. For instance, ML models can analyze historical air pollution data and meteorological variables to forecast future pollution levels. This information is crucial for public health planning and policy-making. Studies have shown that neural networks and ensemble learning methods can achieve high accuracy in air quality predictions [14, 45].

2.2.2 Water quality monitoring

AI-driven models have been employed to assess and predict water quality parameters such as pH, dissolved oxygen, and contaminant levels. These models can process data from sensor networks and remote sensing technologies, providing real-time assessments that are essential for managing water resources and ensuring useful cleaner water [1, 8,23, 27, 33].

2.2.3 Climate change modeling

Machine learning algorithms are instrumental in climate modeling, where they are used to simulate and predict climate patterns based on large-scale environmental data. These models help in understanding the impacts of climate change and in developing strategies for mitigation and adaptation [34].

2.2.4 Biodiversity and ecosystem monitoring

AI techniques, particularly those involving image recognition and processing, are used to monitor biodiversity and assess the health of ecosystems. For example, automated systems using deep learning can identify and classify species from camera trap images, helping to track wildlife populations and biodiversity changes over time [30, 44].

2.2.5 Disaster management

AI and ML are crucial in disaster management and response. Predictive models can forecast natural disasters such as floods, hurricanes, and wildfires, enabling timely interventions and resource allocation. Additionally, AI can assist in analyzing social media data to assess the impact of disasters and coordinate response efforts [24].

2.3 Previous bibliometric studies

Bibliometric studies provide valuable insights into the evolution and impact of research fields by analyzing publication patterns, citation networks, and other metrics. Several bibliometric analyses have been conducted to explore the landscape of AI and ML research in various domains, but fewer studies have focused specifically on their application in environmental monitoring.

One notable study analyzed the global trends in AI research, highlighting the exponential growth in publications and the increasing diversity of application areas [5]. Another study focused on the bibliometric analysis of climate change research, identifying key themes, influential authors, and major research collaborations [17]. These studies highlighted the value of bibliometric analyses in understanding research dynamics and guiding future investigations.

In the context of environmental monitoring, a few bibliometric studies have explored specific subfields. For example, bibliometric reviews of remote sensing applications in environmental monitoring highlighted the leading countries, institutions, and key research areas [46, 52]. Other studies examined the use of AI in water resource management [20], environmental pollution [25] and change detection in remote sensing [38], identifying researchtrends and future prospects.

However, a comprehensive bibliometric analysis that integrates AI and ML applications across the broader spectrum of environmental monitoring is still lacking. This gap presents an opportunity to systematically assess the current state of research, identify emerging trends, and highlight areas for future exploration.

3 Methodology

3.1 Data source and search strategy

For this bibliometric analysis, Scopus was selected as the primary data source due to its comprehensive coverage of peer-reviewed literature across various disciplines, including environmental science, computer science, and engineering. Scopus is renowned for its extensive database, which indexes a wide range of journals, conference proceedings, and other scholarly materials, making it ideal for a detailed bibliometric study [4].

To ensure a comprehensive review of the literature, we utilized a search query designed to capture papers in both AI/ML and environmental monitoring/data analysis domains. The query was constructed to include papers that mentioned either artificial intelligence, machine learning, or related terms, as well as papers focused on environmental monitoring or data analysis, without requiring both topics to appear together in the title, abstract, or keywords. The search query used was:

(TITLE-ABS-KEY("machine learning" OR "artificial intelligence" OR "AI" OR "ML") AND TITLE-ABS-KEY("environmental monitoring" OR "environmental data analysis" OR "environmental data" OR "ecological monitoring" OR "environmental assessment")).

In plain English, this query was structured to find papers that discussed any aspect of AI or ML, as well as those related to environmental monitoring, even if the two areas were not mentioned together. This ensures the search was not overly restrictive and allowed for a broad capture of relevant research.

To ensure that "AI" and "ML" specifically referred to artificial intelligence and machine learning in this study, a thorough manual review was conducted on the titles and abstracts of all retrieved papers to verify that these terms were used in the context of computational techniques relevant to environmental monitoring and relevant aspects, excluding any instances where they referred to unrelated fields.

3.2 Inclusion and exclusion criteria in bibliometrics

The search was conducted in August of 2024, covering all years up to the search date. To ensure the relevance and quality of the selected studies, specific inclusion and exclusion criteria were established. The inclusion criteria followed is listed below:

1. Articles must explicitly focus on the application of AI and ML techniques in environmental monitoring and related aspects.
2. Only peer-reviewed journal articles, review papers, and conference proceedings were included to ensure the reliability and scholarly impact of the findings [56].
3. Publications must be written in English.
4. Up to this stage, studies were evaluated on their title, abstracts and keywords. Full text exploration will be carried out in in-depth analysis Sect. 3.4.

While exclusion criteria was:

1. Studies focusing on AI and ML applications outside environmental monitoring, such as purely medical applications or financial forecasting, were excluded.
2. Editorials, opinion pieces, book chapters, and grey literature were not considered.
3. Duplicate entries were identified and removed to avoid redundancy.

3.3 Data extraction and analysis methods

Following the identification of relevant publications, a structured data extraction process was employed to gather key information from each study. The extracted data included:

1. Title, authors, publication year, journal name, volume, and issue.
2. Abstracts and keywords were used to understand the focus and scope of each study.
3. Number of citations each paper received which is indicative of its impact within the scientific community.
4. Author affiliations and countries to map the geographical distribution of research contributions.

Data extraction was performed using bibliometric tools available in Scopus and complemented by manual verification to ensure accuracy and completeness. The extracted data was then imported into a bibliometric analysis open source software i.e., Bibliometrix [2] and VOSviewer [43].

3.3.1 Analysis methods

1. Analysis of the annual growth in the number of publications to identify trends over time.
2. Identification of collaboration networks. Co-authorship networks were visualized to understand the structure of research collaborations.
3. Mapping of research output by country and region to highlight global research contributions.
4. Examination of the most frequently occurring keywords and their co-occurrence patterns to identify major research themes and emerging topics.

3.4 In-depth analysis methods

After completing the bibliometric analysis and initial collection of relevant literature, this section transitions to a detailed examination of the selected studies. The in-depth analysis focuses on key advancements, methodologies, and findings in the application of AI and machine learning (ML) for environmental monitoring. It aims to identify specific contributions and trends that have emerged in recent years, offering insights that go beyond publication metrics. By closely reviewing individual studies, this section provides a more technical and contextual understanding of how AI and ML are transforming environmental data analysis and monitoring practices.

3.4.1 Criteria for literature selection

The process of selecting literature for this in-depth review was based on a systematic approach to ensure that the most relevant, recent, and impactful studies were included. The focus was on papers that not only demonstrated significant advancements in the application of AI and ML techniques to environmental monitoring but also represented a diverse range of environmental issues, including air and water quality, climate change, biodiversity monitoring, and disaster management. Papers were selected based on the following criteria:

1. The study must explicitly apply AI or ML techniques in environmental monitoring contexts. This includes predictive models, data analysis methods, and decision-support systems tailored to environmental challenges.
2. Preference was given to papers published after 2020 to reflect the latest trends and technological advancements. Older papers were included only if they represent foundational work in this field. This approach was informed by the reviewer's comment regarding the need to include more recent research and papers that align with the latest developments in the field.
3. Prioritized papers that introduced novel methodologies or demonstrated significant improvements over existing techniques, as methodological advances often have far-reaching implications for both academic research and practical applications.
4. Ensured that the selected papers covered a wide range of environmental domains. This includes traditional areas such as air and water quality, as well as emerging fields like biodiversity monitoring, disaster management, and the use of big data in environmental science.
5. In addition to research impact, the papers needed to demonstrate practical relevance, particularly how their findings could be implemented in real-world environmental monitoring systems or influence policy-making. Studies were included if they highlighted applications in decision-support systems for environmental management or contributed to solving pressing environmental challenges.

3.4.2 Literature selection process

The process for selecting these papers involved three main stages, screening, followed by full text reviews and finally categorization:

1. Screening: After collecting the 4762 set of papers, an initial screening was performed based on titles and abstracts to remove papers that did not meet the inclusion criteria.
2. Full-text review: Then a full-text review of the remaining papers, focusing on their methodologies, contributions to the field, and relevance to environmental monitoring. This stage also involved evaluating the robustness of the studies, their dataset sizes, AI techniques used, and the scalability of their findings to real-world applications.
3. Categorization: Finally, the papers were categorized based on their contributions to specific environmental issues and AI/ML methodologies. This organization helped in identifying emerging trends and ensuring that the review provided a meaningful and insightful narrative.

The papers included in this section are thus representative of both foundational and cutting-edge research, and they collectively provide a comprehensive overview of the current state of AI/ML applications in environmental monitoring.

Fig. 1 General publication trends in AI and ML applications in environmental monitoring and data analysis (1991–2024)

4 Results

4.1 General publication trends

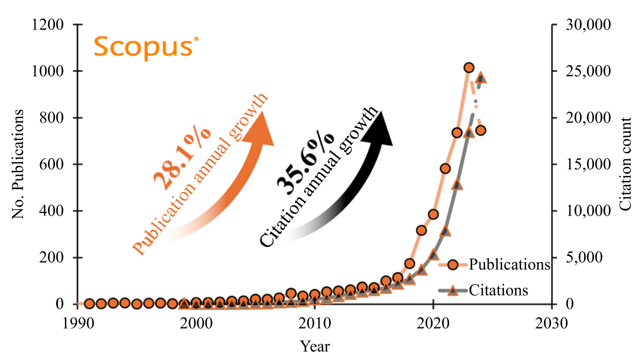
The general publication trends from 1991 to 2024 are summarized in Fig. 1. The analysis includes a total of 4762 documents sourced from 1,580 different sources. These documents were authored by 16,844 researchers, with an average of 5.08 co-authors per document. This high level of collaboration is further emphasized by the 30.24% of documents featuring international co-authors. The referenced works within these documents amount to 233,528, indicating extensive literature engagement and comprehensive research methodologies. On average, each document is cited 21.51 times, reflecting the impactful nature of the research in this field. The average age of the documents is 3.92 years, suggesting a relatively recent and ongoing focus on AI and ML applications in environmental monitoring.

4.1.1 Annual publication and citation growth

Figure 2 illustrates the annual growth in the number of publications and citations related to AI and ML applications in environmental monitoring from 1990 to 2024. The data, sourced from Scopus, reveals a substantial increase in both publications and citations over the past two decades [50].

The publication growth rate is marked at 28.1% annually, indicating a rising interest and output in this research area. This trend began to accelerate significantly around 2010 and continued to rise sharply through the 2020s. Similarly, the citation count has grown at an annual rate of 35.6%, reflecting the increasing impact and recognition of these studies within the scientific community. The rapid rise in citations, especially noticeable from 2015 onwards, suggests that the research produced in this field is highly valued and frequently referenced.

Fig. 2 Annual growth in the number of publications and citations for AI and ML applications in environmental monitoring and data analysis (1990–2024), based on data from Scopus



The substantial increase in both publications and citations over the past two decades reflects a growing recognition of AI and ML technologies as critical tools in addressing environmental challenges. This trend signals the scientific community's heightened focus on leveraging these advanced computational methods to solve complex, large-scale environmental problems. Moreover, the sharp rise in publications since 2010, and particularly in the last decade, shows that the integration of AI/ML into environmental monitoring is rapidly gaining momentum. This growth emphasizes the need for continuous development in AI/ML applications and methodologies to keep pace with the pressing environmental issues faced globally.

4.1.2 Distribution by journals and publishers

Figure 3 shows the distribution of publications related to AI and ML in environmental monitoring across different publishers. Elsevier leads significantly with 1,259 publications, followed by Springer with 675 publications. Institute of Electrical and Electronics Engineers (IEEE) is another major contributor with 510 publications. Other notable publishers include Multidisciplinary Digital Publishing Institute (MDPI) (381 publications), Academic Press (225 publications), and the American Chemical Society (115 publications). Smaller contributions come from John Wiley and Sons Inc., Public Library of Science, Nature Research, and Taylor and Francis Ltd., each contributing fewer than 100 publications.

Figure 4a details the distribution of these publications across specific journals. "Science of the Total Environment" (Elsevier) has the highest number of publications, totaling 307. "Environmental Monitoring and Assessment" (Springer) follows with 274 publications. Other prominent journals include "Environmental Science and Pollution Research" (Springer), "Environmental Research" (Elsevier), and "Environmental Pollution" (Elsevier). Journals from IEEE such as "IEEE Access" and "IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing" also feature significantly. The variety of journals listed reflects the interdisciplinary nature of AI and ML applications in environmental monitoring, covering environmental science, engineering, remote sensing, and pollution research.

The dominance of high-impact environmental science journals, such as "Science of the Total Environment" and "Environmental Monitoring and Assessment," in publishing AI/ML research highlights the growing mainstream integration of these technologies into environmental sciences. However, significant contributions from computer science and engineering journals, such as those from IEEE, indicate the interdisciplinary nature of this field. This pattern suggests that researchers in environmental monitoring should consider publishing across diverse academic domains to promote cross-disciplinary innovation. The presence of AI/ML research in a variety of high-impact journals also reflects the scientific community's acknowledgment of the transformative potential of these technologies for advancing environmental science.

4.1.3 Distribution by field

Figure 4b illustrates the distribution of AI and ML research in environmental monitoring across various academic fields. Environmental Science leads with 25% of the publications, indicating a strong focus on applying these technologies to environmental challenges. Computer Science follows with 16%, reflecting the technical and algorithmic development that supports these applications. Engineering accounts for 12% of the publications, showcasing the role of engineering solutions in environmental monitoring. Other significant fields include Earth and

Fig. 3 Distribution of AI and ML publications in environmental monitoring and data analysis by different publishers

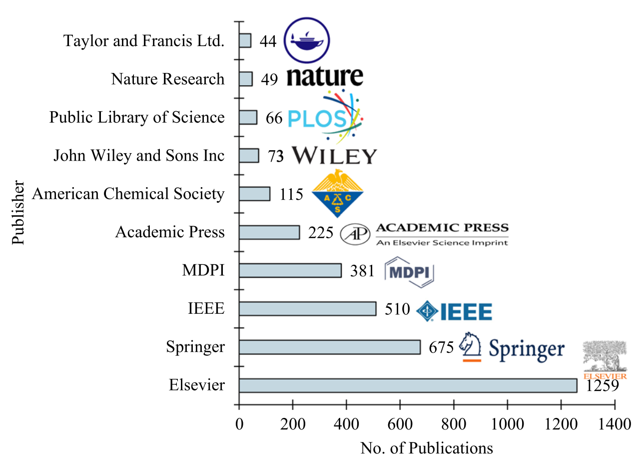
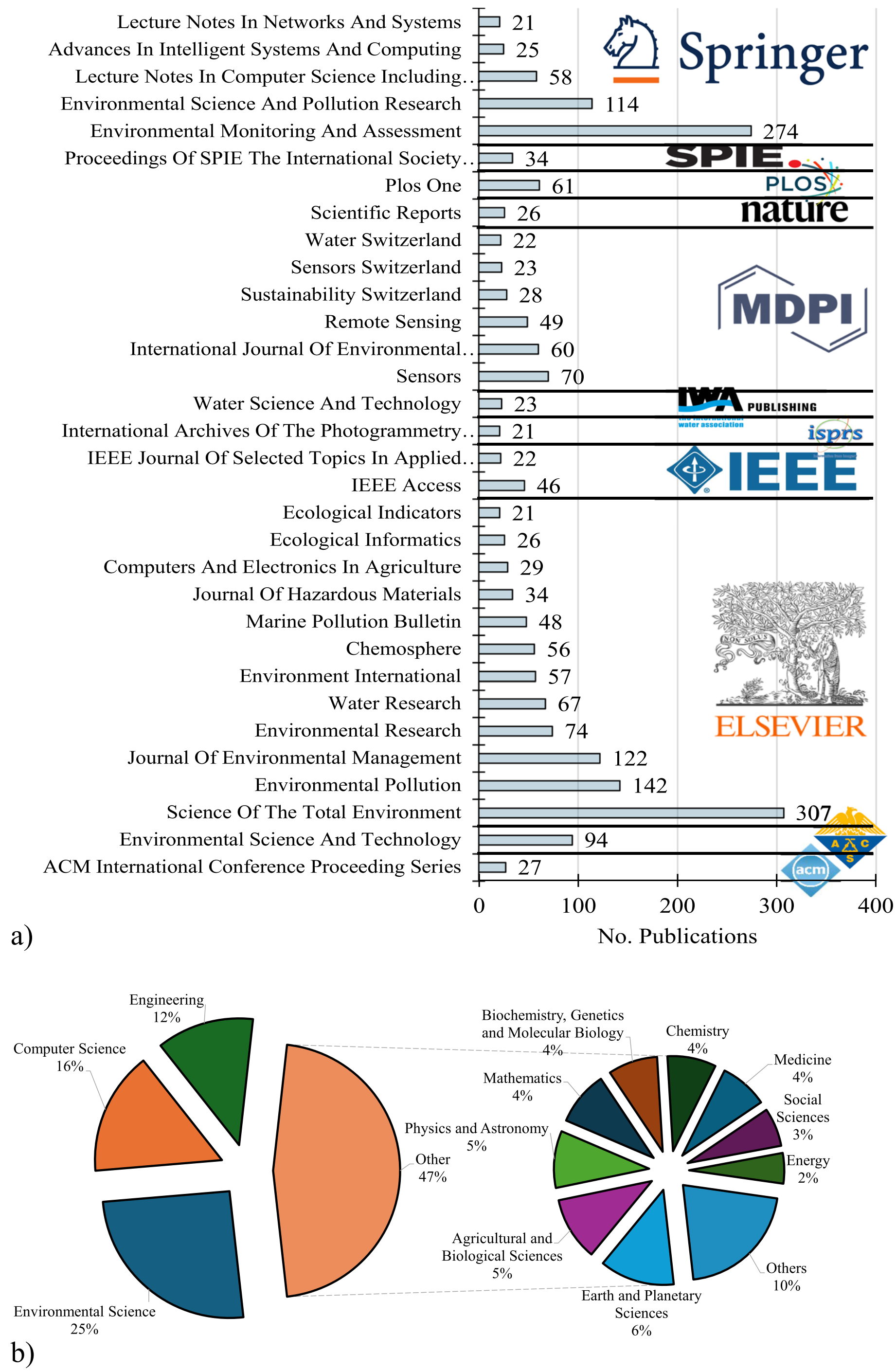
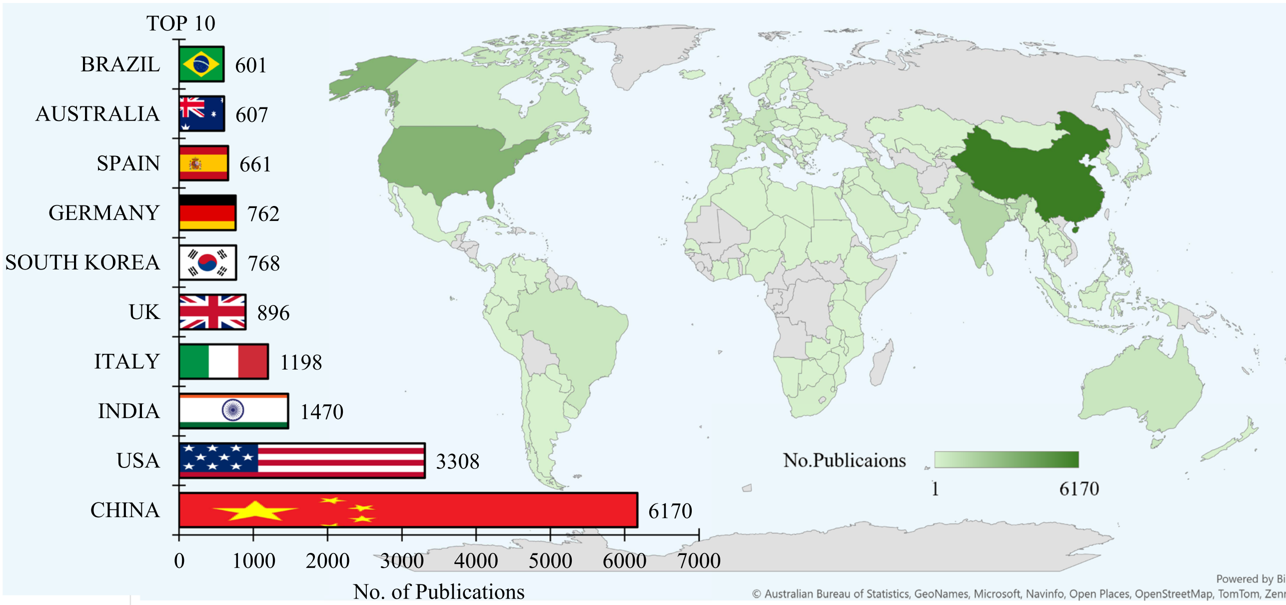


Fig. 4 Distribution of AI and ML publications in environmental monitoring and data analysis across. **a** Various journals; and **b** different academic fields

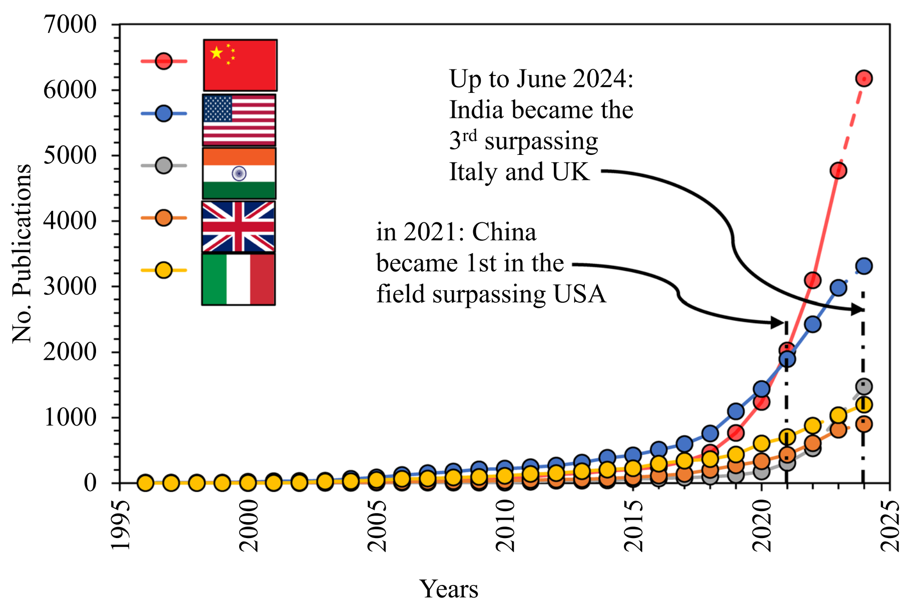


Planetary Sciences (6%), Agricultural and Biological Sciences (5%), Biochemistry, Genetics, and Molecular Biology (4%), Chemistry (4%), Mathematics (4%), and Medicine (4%). The remaining publications are spread across Physics and Astronomy, Social Sciences, and Energy, each contributing between 2 and 3%.

A considerable portion of the research falls under the 'Other' category (47%), highlighting the interdisciplinary nature of this field and the integration of AI and ML into various other scientific domains.



a)



b)

Fig. 5 Research contributions in AI and ML environmental monitoring and data analysis research publications: **a** Geographical distribution, **b** Temporal evolution of top five countries

4.2 Geographical distribution of research

Figure 5a presents the geographical distribution of research publications related to AI and ML in environmental monitoring. The map highlights the countries with the highest number of publications, the left side of the figure lists the top 10 contributing countries. China is leading significantly with 6170 publications. The United States follows with 3308 publications. India, the United Kingdom, and Italy also make substantial contributions with 1,470, 836, and 1198 publications, respectively. Other notable contributors include South Korea (768 publications), Germany (762 publications), Spain (661 publications), Australia (607 publications), and Brazil (601 publications). It is worth noting that these values are without factoring in co-authorships; for instance, if authors from both the USA and China are in one paper, it is counted as 1 for each country.

Figure 5b illustrates the contributions of the top five countries—China, the United States, India, the United Kingdom, and Italy—over time. This demonstrates the dynamic growth and shifting leadership in AI and ML research for environmental monitoring and data analysis, with significant increases in publication numbers particularly evident from 2015 onwards. China's research output surged past that of the United States in 2021, becoming the leading contributor in this field. By August 2024, India had risen to the third position, surpassing both Italy and the United Kingdom. The geographical distribution of research reveals a clear dominance by countries such as China and the United States, both of which have robust technological infrastructures and significant investments in AI and environmental research. However, there is a noticeable gap in research contributions from many developing nations, particularly in Africa and parts of Asia. This disparity highlights the need for global capacity-building initiatives to foster AI and ML applications in environmental monitoring across diverse ecological and socio-economic contexts. Collaborative efforts between leading countries and underrepresented regions could enhance the global applicability of AI/ML solutions and ensure these technologies are inclusive and beneficial to all regions, especially those facing the most severe environmental challenges.

4.3 Collaboration networks

The collaboration networks among countries in the field of AI and ML applications for environmental monitoring is shown in Fig. 6. The network visualization shows a web of connections, with nodes representing countries and edges indicating collaborative links between them.

In Fig. 6a, the network is color-coded, with red and blue indicating different clusters of closely collaborating countries. This clustering reveals regional patterns of collaboration, with countries in the same cluster often sharing geographical proximity or similar research interests. The United States and China emerge as central hubs with extensive collaboration networks. Both countries show a high degree of interconnectedness with other nations, highlighted by numerous connections and larger node sizes. Other significant nodes include Germany, the United Kingdom, Canada, and Australia, each showing strong collaborative ties within the network.

Moreover, Fig. 6b illustrates the collaboration networks with an emphasis on the average number of publications per year. The nodes are sized according to the average number of publications, and the color gradient from blue to green to yellow represents the timeline of collaborations, spanning from 2019 to 2023. This visualization highlights the dynamic nature of research collaborations over time, with significant contributions from countries like the United States, China, India, and the United Kingdom. The temporal evolution indicates growing international cooperation and an increasing number of joint publications in recent years. Notably, the United Arab Emirates, Morocco, Ecuador, and Kenya are the newest countries to join this cluster, as indicated by their yellow nodes.

4.4 Keyword analysis

4.4.1 Most common keywords

Figure 7a presents a word cloud depicting the most common keywords found in research publications on AI and ML applications in environmental monitoring and data analysis. The size of each keyword indicates its frequency of occurrence in the literature. The term "machine learning" is prominently featured, reflecting its central role in this research area. Other significant keywords include "artificial intelligence," "remote sensing," "prediction," "climate change," "air pollution," and "deep learning." These terms highlight the main themes and applications being explored within the field. Additional frequently occurring keywords include "internet of things," "random forest," "water quality," "monitoring," "air quality," "groundwater," and "big data." This indicates a broad range of topics being investigated, from specific techniques like random forests to various applications such as air quality and water quality monitoring. Terms like "particulate matter," "classification," "GIS," "sensors," and "agriculture" also appear frequently, emphasizing the diverse application areas and methodologies utilized in environmental monitoring research and data analysis.

4.4.2 Keyword co-occurrence network

Figure 7b illustrates the keyword co-occurrence network, showing how frequently different keywords appear together in publications. The nodes represent keywords, and the edges represent co-occurrences. The size of the nodes reflects

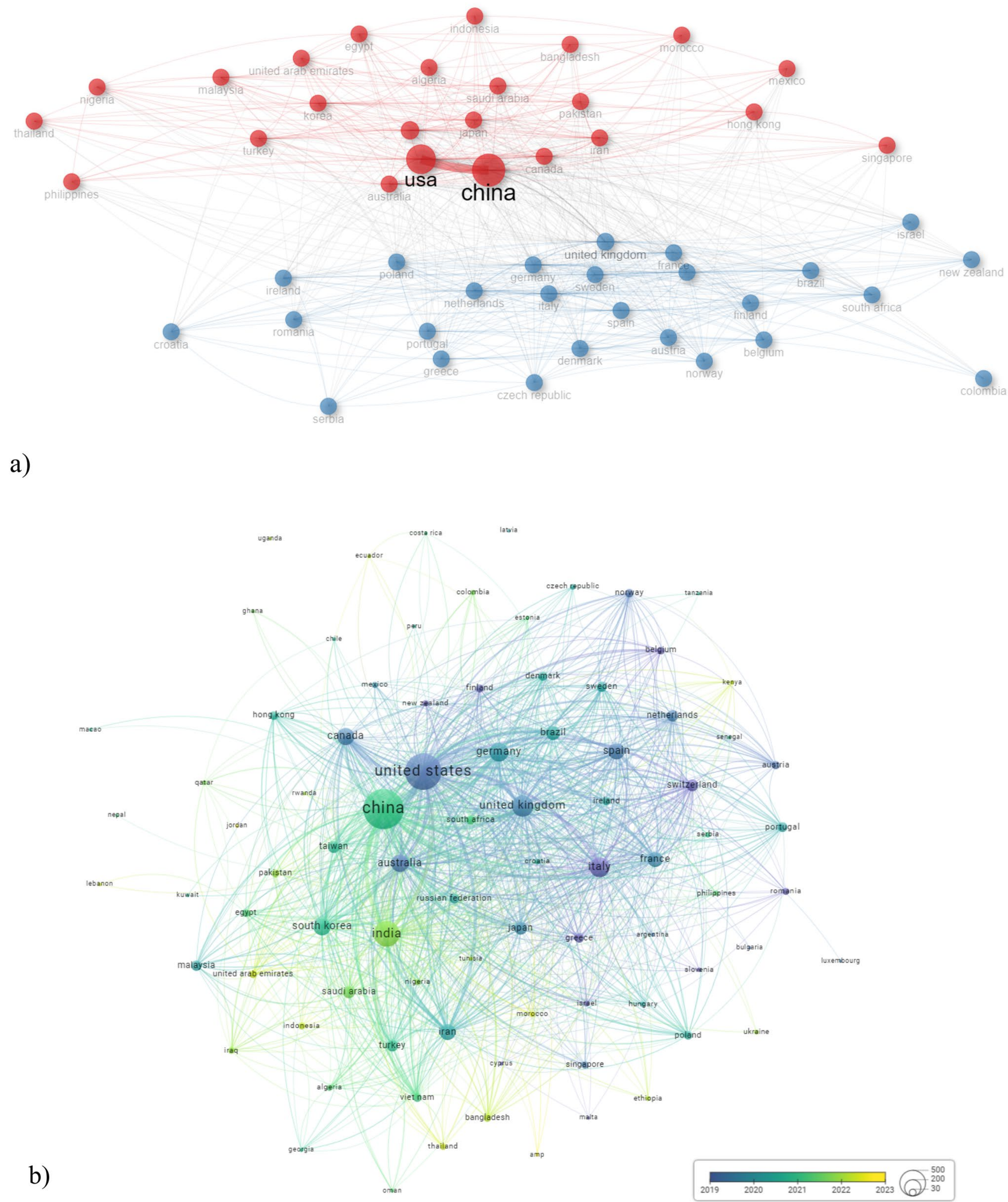


Fig. 6 International collaboration networks. **a** Clustered, and **b** with average publications/year in AI and ML research for environmental monitoring and data analysis. The threshold minimum of 5 publications/country which excluded 78 out of 163 countries (remaining 85)

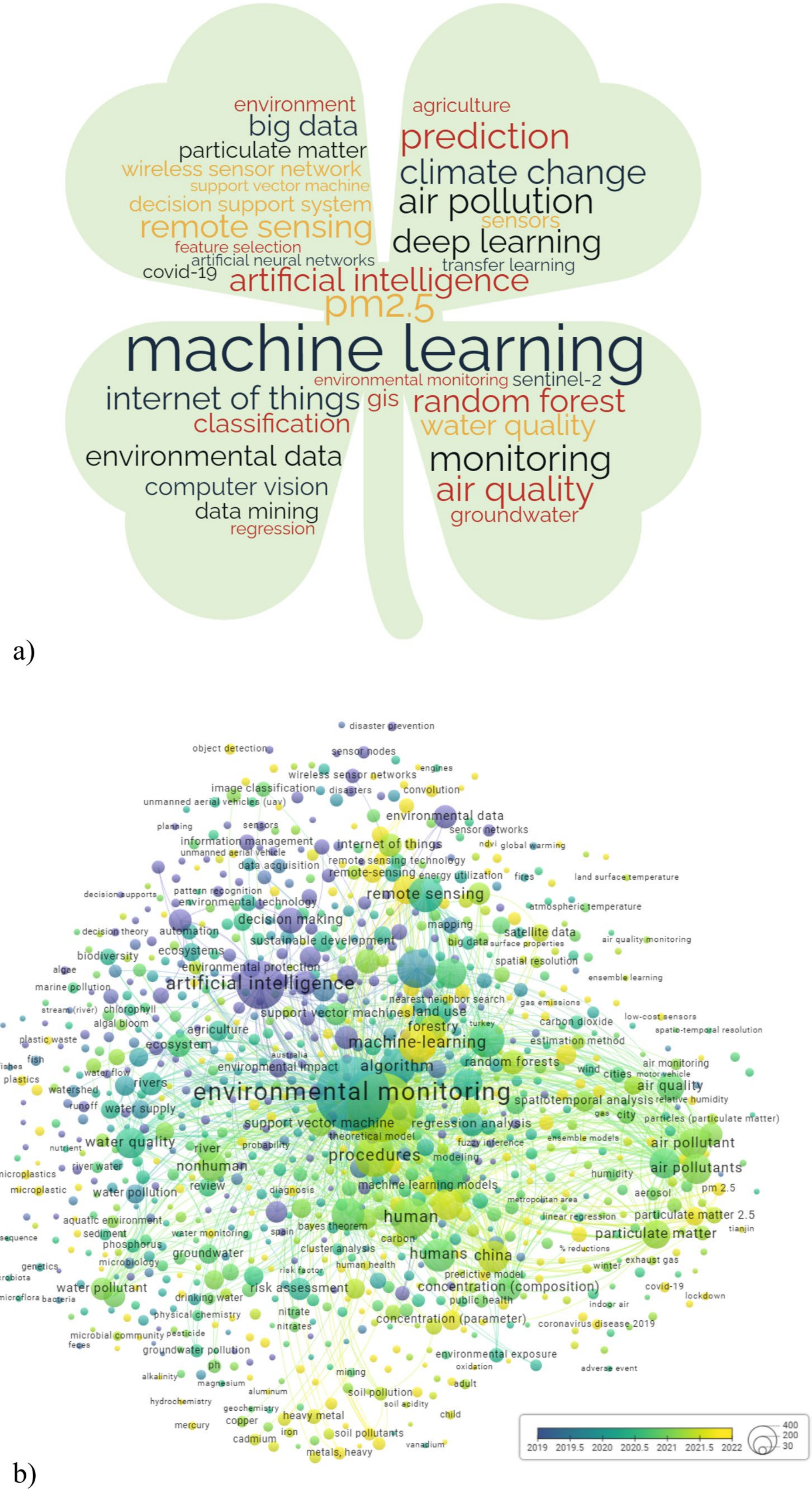


Fig. 7 **a** Word cloud and **b** keyword co-occurrence with publication/year AI and ML research for environmental monitoring and data analysis. Full counting with author keywords was used with minimum occurrence of 5 keywords was selected, eliminating 1659 words out of 1983 (remaining 324)

the frequency of the keywords, while the color gradient from blue to green to yellow represents the timeline from 2019 to 2023. This visualization provides perceptions into the interconnections between different research themes and the evolution of focus areas over time.

The prominence of terms related to "climate change," "air quality," "water quality," and "biodiversity" highlights the significant environmental challenges being addressed through AI and ML technologies. The inclusion of emerging technologies such as the "Internet of Things" and "deep learning" highlights the ongoing innovation in data collection and analysis methods.

A notable trend in Fig. 7b is the shift towards the yellow-colored nodes, representing keywords that have gained prominence in the most recent years (2022–2023). These include terms like "particulate matter," "water pollutants," "soil pollutants," "groundwater," and "heavy metals," indicating a growing research focus on contemporary environmental issues and their intersection with public health concerns. This trend reflects the research community's responsiveness to urgent environmental and societal challenges, leveraging AI and ML technologies for timely and impactful studies that has practical implications rather than theoretical concepts which demonstrates the adaptability and relevance of AI and ML applications in addressing emerging environmental issues. This trend is beneficial as it ensures that research efforts remain aligned with current global priorities, facilitating the development of innovative solutions for contemporary challenges in environmental monitoring and data analysis.

4.4.3 Thematic trends

Figure 8 provides a thematic map of emerging trends in AI and ML applications for environmental monitoring and data analysis, categorized into four quadrants: Motor Themes, Niche Themes, Basic Themes, and Emerging or Declining Themes.

The Motor Themes quadrant includes terms such as "internet of things" (IoT), "air pollution," "air quality," and "PM2.5." These themes are characterized by both high centrality and high density, indicating that they are well-developed and crucial to the research field. The integration of IoT with environmental monitoring allows for real-time data collection and analysis, enhancing the capability to monitor air quality and pollution levels efficiently. The

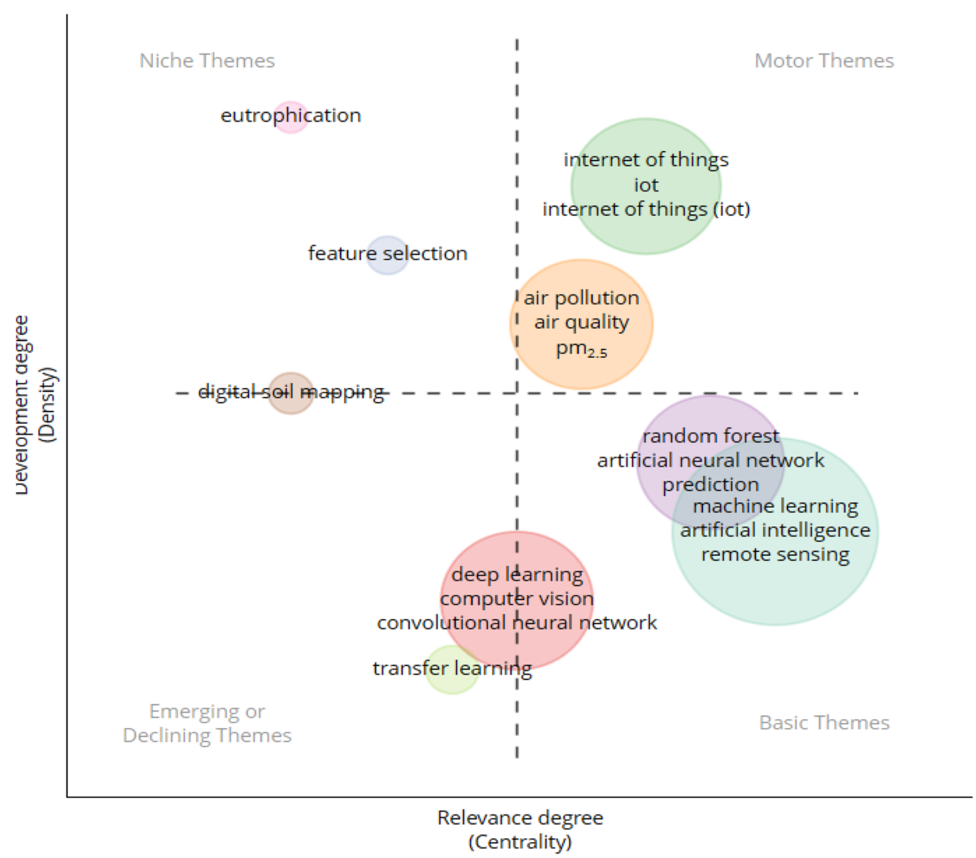


Fig. 8 Thematic map of emerging trends in AI and ML research for environmental monitoring and data analysis

frequent occurrence of terms related to air quality highlights the significant attention given to addressing pollution and its impacts on health and the environment.

In the Niche Themes quadrant, terms like "eutrophication" and "feature selection" are found. These themes have high density but low centrality, suggesting that they are specialized and well-developed within their niche but not as widely connected to other themes. Eutrophication is a specific environmental issue related to water bodies, indicating focused research efforts in this area. Feature selection, a crucial step in building effective AI models, shows the importance of refining data inputs to enhance model performance.

The Basic Themes quadrant contains terms such as "random forest," "artificial neural network," "prediction," "machine learning," "artificial intelligence," and "remote sensing." These themes have high centrality but lower density, signifying that they are fundamental to the field and widely applicable across various research areas. These basic AI and ML techniques form the backbone of numerous studies, facilitating diverse applications from prediction models to remote sensing data analysis.

In the Emerging or Declining Themes quadrant, terms like "deep learning," "computer vision," "convolutional neural network," "transfer learning," and "digital soil mapping" are present. These themes have low density and low centrality, indicating that they are either emerging or potentially declining areas of research. Deep learning and computer vision are gaining traction for their ability to process complex data types, such as images and video. Transfer learning is an emerging approach that leverages knowledge from one domain to improve model performance in another, highlighting innovative strategies in AI research. Digital soil mapping's position may suggest a more specialized focus or evolving interest in this area.

4.5 H-Index analysis

To assess the impact of journals that publish research on AI and ML in environmental monitoring, we conducted an h-index analysis. The h-index was collected from the Scimago Journal Rank (SJR) database, which is a widely recognized source for journal impact metrics [28]. The h-index reflects both the productivity and citation impact of the articles published in a journal.

Figure 9 shows the relationship between the number of published articles on AI for environmental monitoring and the overall h-index of the journals that published these articles. As shown, there is a positive trend between the number of articles and the h-index, indicating that journals with a higher volume of AI-related environmental publications tend to have higher impact metrics. The coefficient of determination $R^2 = 0.23$ suggests a moderate correlation between the number of articles and the journal's h-index.This relationship highlights that well-established and reputable journals are actively participating in the dissemination of research on AI for environmental monitoring, further underscoring the importance and growing influence of this field. Journals with higher h-index scores tend to publish work that receives greater attention and recognition, demonstrating the increasing relevance of AI and ML technologies in addressing critical environmental challenges.

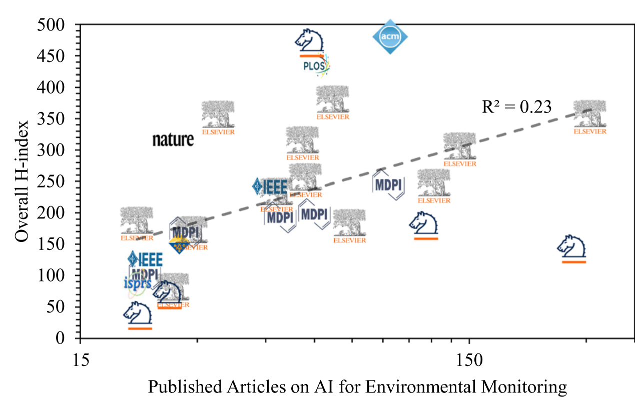


Fig. 9 Overall-H index of journals labeled with their publisher with published articles on AI for environmental monitoring

Table 1 Thematic categories of AI and ML applications in environmental monitoring

Categories	Areas covered
Methodological advances in AI/ML for environmental monitoring	Papers that focus on improving AI and ML techniques, introducing new algorithms, or enhancing the accuracy and interpretability of models
Applications in air quality monitoring	Studies applying AI and ML to the prediction and management of air quality, including models that integrate sensor data, satellite imagery, and real-time monitoring systems
Applications in water quality monitoring	Papers focusing on the use of AI/ML to predict and manage water quality, including models that address issues like pollution, contamination, and resource management
Climate change modeling and prediction	Studies applying AI/ML to the modeling of climate change impacts, predictive modeling of environmental patterns, and the development of decision-support tools for climate adaptation and mitigation
Biodiversity and ecosystem monitoring	Papers utilizing AI and ML to monitor biodiversity, assess ecosystem health, and predict changes in species distributions or habitat conditions
Disaster management and prediction	Studies focused on using AI/ML for predicting and managing environmental disasters, including floods, wildfires, and hurricanes
Emerging technologies and interdisciplinary applications	Studies that integrate AI/ML with other emerging technologies, such as the Internet of Things (IoT), unmanned aerial vehicles (UAVs), and big data platforms for enhanced environmental monitoring

5 In-depth analysis of key literature

5.1 Organization of the in-depth analysis

Following the methodology described previously for the in-depth analysis, the selected studies were organized into thematic key categories as shown in Table 1.

5.2 Methodological advances in AI/ML for environmental monitoring

Several studies in the reviewed literature have made significant contributions to advancing AI/ML methodologies specifically tailored for environmental monitoring. These methodological advances range from the development of novel algorithms to improving the interpretability and scalability of existing models.

One early important study by Zhu et al. [51] introduced the Soil-Land Inference Model (SoLIM), which combines geographic information systems (GIS), expert knowledge, and fuzzy logic to improve soil mapping. The integration of domain expertise with AI techniques exemplifies how combining AI with traditional environmental sciences can yield more accurate and actionable models. This study highlighted the potential for AI to enhance soil and land management practices by improving the resolution and accuracy of soil maps, which are critical for sustainable agriculture and land use planning [18].

Khatami et al. [21] conducted a comprehensive meta-analysis on land-cover classification methods, focusing on how the inclusion of texture data can improve classification accuracy. The study demonstrated that support vector machines outperform other classification algorithms when texture data is incorporated. This methodological advance is particularly relevant for remote sensing applications, where land-cover classification is essential for monitoring environmental changes over time.

Another key study by Chapi et al. [6] introduced the Bagging-LMT model, which combines the bagging ensemble method and the Logistic Model Tree (LMT) algorithm for flood susceptibility mapping. Their results showed that the hybrid model outperformed traditional models, such as logistic regression and random forest, by improving prediction accuracy and reducing model bias. This methodological improvement is crucial for disaster management, where accurate and timely predictions can help in resource allocation and risk mitigation.

In a more recent study, Ren et al. [35] proposed a dual-attention multiscale network (DAMFANet) for remote sensing image change detection. Their model integrates feature fusion and semantic information integration, which significantly improves the performance of change detection algorithms. This model addresses the common challenge of handling large-scale, heterogeneous data in environmental monitoring, offering a scalable and efficient solution for processing remote sensing images. The methodological innovation of combining attention mechanisms with multiscale feature learning opens new avenues for AI applications in environmental science, particularly in areas such as deforestation monitoring and land-use change detection.

These studies collectively represent significant advancements in AI/ML methodologies for environmental monitoring. By improving the accuracy, scalability, and interpretability of models, these methodological innovations are helping to bridge the gap between theoretical research and practical applications.

5.3 Applications in air quality monitoring

Air quality monitoring is a critical area where AI and ML have shown significant promise. Traditional air quality monitoring systems rely on fixed sensors that provide data at specific locations, but these systems often lack the spatial and temporal resolution needed for comprehensive environmental assessments [37]. AI/ML techniques offer the potential to enhance air quality monitoring by integrating data from various sources, such as satellite imagery, sensor networks, and meteorological data, to provide more accurate and timely predictions.

Cao et al. [55] applied a Bayesian optimization algorithm to predict pollutant levels in urban environments. Their study demonstrates that ensemble models, coupled with real-time sensor data, can improve the accuracy of air quality predictions, particularly in densely populated urban areas. The Bayesian optimization approach allowed for the fine-tuning of model parameters, resulting in improved predictive performance and model interpretability.

Chen et al. [7] developed a model using random forest algorithms to estimate particulate matter (PM2.5) concentrations across China from 2005 to 2016. Their model explained 83% of the variability in PM2.5 concentrations, highlighting the potential of AI/ML techniques to accurately predict air quality over large geographic areas.

Di et al. [9] developed an ensemble model that integrated multiple machine learning algorithms to estimate PM2.5 concentrations in the United States. The model achieved high accuracy with a tenfold cross-validated R2 of 0.86–0.89. Their ensemble approach demonstrates how combining multiple modelscan improve the robustness and generalizability of air quality predictions, particularly in regions with varying climatic conditions. In a more recent global PM2.5 scale study, Yu et al. [48] employed a deep ensemble machine learning (DEML) model to predict daily PM2.5 concentrations worldwide from 2000 to 2019. This approach combined ground-based measurements from 5,446 monitoring stations across 65 countries with chemical transport model outputs and meteorological data, achieving a high spatial resolution of 0.1° x 0.1°. The DEML model explained 91% of the observed variability in PM2.5 concentrations, with a root mean square error of 7.86 $\mu g/m^3$, providing detailed insights into global exposure patterns and spatiotemporal changes over two decades. The study identified significant regional disparities, with consistently high PM2.5 levels in South and East Asia, where more than 90% of days exceeded the WHO daily PM2.5 limit of 15 $\mu g/m^3$ in 2019, while regions such as Europe and North America saw reductions in population-weighted PM2.5 exposure over time. These studies are particularly relevant for policymakers and environmental managers, as they offer scalable approaches to air quality monitoring that can be applied across diverse geographic regions.

While these studies highlight the significant potential of AI and ML in air quality monitoring, they also point to several challenges. For instance, many models still struggle to predict extreme pollution events or account for the complexities of urban microclimates. Addressing these challenges will require future research to focus on integrating multi-scale data sources, such as combining satellite data with ground-based sensors, to improve the spatial and temporal resolution of air quality predictions. Additionally, developing more interpretable AI models could provide valuable insights into the causal drivers of air pollution, which would better inform policymakers and urban planners in their decision-making processes. Interpretable models would allow researchers to not only predict pollution levels but also to understand the underlying causes, which is crucial for designing more effective interventions to improve air quality.

5.4 Applications in water quality monitoring

Water quality monitoring is another crucial environmental domain where AI and ML have demonstrated significant value. Traditional water quality assessment relies on manual sampling and laboratory analysis, which can be time-consuming, expensive, and logistically challenging, particularly in remote areas. AI/ML techniques offer the potential to automate and enhance these processes by utilizing data from sensor networks, remote sensing platforms, and historical datasets to predict water quality parameters in real-time.

Palani et al. [31] applied artificial neural networks (ANNs) to forecast water quality in Singapore's coastal waters. The model achieved high predictive accuracy, with Nash–Sutcliffe efficiency coefficients ranging from 0.8 to 0.9, which demonstrates the effectiveness of AI in predicting key water quality parameters, such as nutrient concentrations and dissolved oxygen levels. This study is particularly relevant for managing coastal ecosystems, where water quality fluctuations can have significant ecological and economic impacts.

A more recent study by Lu and Ma [26] introduced a hybrid model that integrates the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method and extreme gradient boosting (XGBoost) for water quality prediction in the Tualatin River. The hybrid model outperformed traditional models, highlighting the advantages of combining signal processing techniques with machine learning algorithms to improve the accuracy of environmental predictions. Their approach of using empirical mode decomposition to preprocess noisy water quality data before applying machine learning algorithms is a methodological innovation that could be applied to other environmental monitoring contexts with noisy or incomplete datasets.

Naghibi et al. [29] employed a range of machine learning models, including boosted regression trees (BRT) and random forest (RF), for groundwater potential mapping in Iran. Their results showed that the BRT model outperformed other models in terms of predictive accuracy, suggesting that ensemble learning techniques are well-suited for environmental applications that require the integration of heterogeneous datasets. Groundwater potential mapping is essential for sustainable water resource management, particularly in arid and semi-arid regions where water is a scarce resource [39].

In another study, Hill and Minsker [19] developed a real-time anomaly detection system for environmental data streams using autoregressive modeling and multilayer perceptron networks. Their system was able to detect erroneous data in water quality sensors and flag anomalies in real-time, demonstrating the potential of AI/ML techniques to enhance the reliability of sensor-based monitoring systems.

While the studies reviewed here demonstrate the potential of AI and ML to improve water quality monitoring, they also highlight several challenges. For example, many models rely on high-quality, real-time data, which may not always be available in resource-constrained regions. Future research should focus on developing robust AI models that can handle incomplete or noisy datasets, as well as exploring the potential of data fusion techniques to integrate data from multiple sources, such as satellite imagery, sensor networks, and historical datasets. Additionally, the development of explainable AI models would provide greater transparency and trust in AI-driven water management systems, particularly in regulatory and policy-making contexts.

5.5 Climate change modeling and prediction

AI and ML techniques have been increasingly applied in climate change modeling and prediction, where traditional simulation models often struggle with the complexity and scale of environmental data. Machine learning offers a powerful tool for uncovering patterns and making predictions based on vast amounts of historical climate data, enabling more accurate forecasts of climate change impacts and better-informed decision-making for mitigation and adaptation strategies.

Reichstein et al. [34] demonstrated the application of deep learning techniques to simulate climate dynamics and predict long-term environmental changes. Their study highlighted the potential of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to model time series data and capture complex spatio-temporal relationships in climate systems. By leveraging high-resolution satellite data and integrating it with ground-based measurements, this study offers a scalable approach to climate modeling that could be applied to various regions and environmental contexts.

Yu et al. [47] developed a climate change impact assessment model using a combination of neural networks and decision trees. Their model focused on predicting the effects of climate change on agricultural yields, which is critical for food security planning in the face of changing climate patterns. The study showed that machine learning models could predict crop yields with high accuracy, offering a valuable decision-support tool for farmers and policymakers in regions vulnerable to climate change.

Another important contribution to this field comes from Forzieri et al. [10], who integrated satellite-based vegetation indices with random forest models to assess the resilience of forests to climate change. Their model was able to identify vulnerable ecosystems and predict how different climate scenarios would affect forest health over time. This research is particularly relevant for conservation efforts and biodiversity management, as it provides a scalable approach to monitoring the impacts of climate change on ecosystems.

AI/ML techniques are also being used to model the impacts of extreme weather events, such as floods, droughts, and hurricanes. For example, Chapi et al. [6] applied machine learning algorithms to predict flood susceptibility in Iran, demonstrating the potential of these models to identify at-risk areas and inform disaster preparedness efforts. Similarly, Gibb et al. [11] used AI to analyze social media data and satellite images to assess the real-time impacts of hurricanes and provide decision-support for emergency response teams.

Despite the significant advancements in AI/ML-based climate modeling, there are still challenges that need to be addressed. One of the main challenges is the "black-box" nature of many machine learning models, which makes it difficult to interpret the results and understand the underlying mechanisms driving the predictions. Future research should focus on developing explainable AI models that provide insights into the drivers of climate change and its impacts. Additionally, there is a need for more research on the integration of AI models with physical climate models to ensure that predictions are grounded in established scientific principles and are more generalizable across different regions and scenarios.

5.6 Biodiversity and ecosystem monitoring

Biodiversity monitoring is another critical area where AI and ML have been applied with increasing success. AI/ML techniques, particularly those involving image recognition and data analysis, are being used to monitor biodiversity, assess ecosystem health, and track changes in species distributions and habitat conditions. These techniques offer a scalable and efficient alternative to traditional monitoring methods, which often rely on manual data collection and are limited in scope and coverage.

Phillips et al. [32] developed the Maxent algorithm, a machine learning technique for species distribution modeling, which has since become a widely used tool in biodiversity research. The study demonstrated that Maxent outperforms other algorithms, such as the Genetic Algorithm for Rule-set Production (GARP), in predicting species ranges based on environmental variables. This research has significant implications for conservation planning and biodiversity management, as it allows researchers to predict how species distributions will shift in response to environmental changes, such as habitat loss and climate change.

Norouzzadeh et al. [30] utilized deep learning algorithms to analyze camera trap images for wildlife monitoring. Their model was able to automatically identify and classify animal species with high accuracy, significantly reducing the time and effort required for manual image analysis. This approach is particularly relevant for monitoring wildlife in remote or inaccessible areas, where traditional methods may be impractical.

In a similar vein, Gonzalez et al. [12] combined unmanned aerial vehicles (UAVs) with AI techniques, such as object tracking and thermal imaging, to monitor wildlife populations in forest ecosystems. By integrating AI with UAVs, the researchers were able to track species in real-time and identify population trends with greater accuracy than traditional ground-based methods. This interdisciplinary approach demonstrates the potential of AI to enhance biodiversity monitoring by providing real-time data on species distributions and habitat conditions.

The use of AI in biodiversity monitoring is not limited to wildlife tracking. Zhang et al. [49] explored the application of machine learning in enhancing fluorescence-based detection methods for identifying environmental contaminants that affect biodiversity, such as aluminum and fluoride ions. Their model demonstrated superior accuracy and sensitivity compared to traditional methods, highlighting the potential of AI to improve environmental monitoring techniques in a variety of contexts.

While these studies highlight the significant contributions of AI to biodiversity monitoring, there are still challenges to overcome. One of the main challenges is the need for large, high-quality datasets to train machine learning models. Many regions, particularly in the Global South, lack comprehensive biodiversity data, which limits the applicability of AI/ML techniques in these areas. Future research should focus on developing models that can work with limited or imbalanced datasets, as well as exploring the potential of data augmentation techniques to improve model performance in data-scarce regions.

5.7 Disaster management and prediction

AI and ML have also been applied in disaster management and prediction, where they offer significant improvements over traditional approaches. By analyzing large amounts of data from various sources, such as satellite imagery, social media, and sensor networks, AI/ML models can provide real-time predictions of natural disasters, such as floods, hurricanes, and wildfires, and offer decision-support for emergency response teams.

Chapi et al. [6] applied the Bagging-LMT model to predict flood susceptibility in Iran. Their model was able to identify high-risk areas with greater accuracy than traditional methods, offering a valuable tool for disaster preparedness and risk management. The study demonstrates how AI/ML techniques can be used to enhance the accuracy of disaster prediction models and provide decision-makers with actionable insights for resource allocation and emergency planning.

In another study, Linardos et al. [24] developed a machine learning model that integrates social media data with satellite images to assess the real-time impacts of hurricanes. Their model was able to analyze the geographic distribution of social media posts to identify affected areas and estimate the severity of the damage. This approach demonstrates the potential of AI/ML to augment traditional disaster management tools by providing real-time, crowdsourced data that can be used to inform emergency response strategies. The integration of social media data with satellite imagery offers a more comprehensive view of the situation on the ground, enabling faster and more accurate assessments of disaster impacts.

Uddin et al. [42] applied Monte Carlo simulations and Gaussian Process Regression for uncertainty analysis in water quality index (WQI) models used during natural disasters, such as floods and tsunamis. By integrating probabilistic methods with machine learning, their model accounted for the uncertainties inherent in environmental data, which is often noisy and incomplete during disaster events. This approach offers a robust framework for making informed decisions during emergencies, where time and accuracy are of the essence.

Another recent study by Haridasan et al. [15] proposed an automated system for detecting and classifying natural hazards, such as landslides and earthquakes, using convolutional neural networks (CNNs) and principal component analysis (PCA). Their system demonstrated high predictive accuracy and real-time processing capabilities, which are essential for disaster preparedness and response. By leveraging AI techniques, their model was able to process large datasets in real-time, providing emergency responders with early warnings and actionable insights.

These studies illustrate the potential of AI and ML to revolutionize disaster management by improving the speed and accuracy of predictions and enhancing decision-making during emergencies. However, several challenges remain. One of the main challenges is the scalability of AI/ML systems for disaster prediction in different geographic regions. Many models are trained on region-specific data, which may not generalize well to other areas with different environmental and socio-economic conditions. Future research should focus on developing more adaptable models that can be applied across diverse regions and disaster types. Additionally, there is a need for more research on the ethical implications of using AI in disaster management, particularly regarding issues of data privacy and the potential for bias in decision-making processes.

5.8 Emerging technologies and interdisciplinary applications

The integration of AI/ML with emerging technologies, such as the Internet of Things (IoT), unmanned aerial vehicles (UAVs), and big data platforms, has opened up new possibilities for environmental monitoring and disaster management. These interdisciplinary applications combine the strengths of AI/ML with advanced data collection and processing technologies, enabling more comprehensive and real-time environmental assessments.

Asha et al. [3] developed the IoT-based Environmental Toxicology for Air Pollution Monitoring with Artificial Intelligence Techniques (ETAPM-AIT) model, which combines IoT sensors with AI algorithms to monitor air quality and detect toxic pollutants. Their system was designed to provide real-time alerts to industrial facilities and government agencies, enabling faster responses to air pollution events. By integrating IoT with AI, their model represents a significant advancement in real-time environmental monitoring, offering scalable solutions for cities and industries looking to mitigate the impacts of air pollution.

In another interdisciplinary study, Gonzalez et al. [12] used unmanned aerial vehicles (UAVs) equipped with AI-based object detection algorithms to monitor wildlife populations in remote forest areas. The combination of UAVs with AI allowed for real-time data collection and analysis, providing a more efficient and accurate way to track species distributions and monitor ecosystem health. This approach demonstrates the potential of AI to enhance traditional environmental monitoring methods by providing real-time insights that can inform conservation and land management strategies.

Tamiminia et al. [41] explored the use of the Google Earth Engine (GEE) platform for big data analysis in remote sensing applications. By leveraging AI/ML techniques within the GEE platform, the study was able to process large-scale satellite imagery datasets for land cover classification and climate analysis. Their work highlights the growing role of big data platforms in environmental monitoring, where AI/ML models can be deployed at scale to analyze vast amounts of geospatial data in real-time.

The use of AI in environmental monitoring is also expanding into novel areas, such as biosensors and public health monitoring. Zhang et al. [49] demonstrated the application of machine learning in enhancing fluorescence-based detection methods for environmental contaminants, such as aluminum and fluoride ions. Their AI-enhanced biosensor showed superior accuracy and sensitivity compared to traditional detection methods, highlighting the potential of AI to improve environmental health monitoring and public safety.

While these interdisciplinary applications represent significant advancements in the field, there are still challenges to be addressed. One of the main challenges is the integration of AI/ML models with emerging technologies in a way that ensures scalability and robustness. Many AI-based systems require large amounts of high-quality data, which may not always be available, particularly in resource-constrained regions. Future research should focus on developing more robust and adaptable AI models that can function effectively in low-resource environments. Additionally, there is a need for more research on the ethical and regulatory challenges associated with the deployment of AI-based systems for environmental monitoring, particularly regarding issues of data security, privacy, and the potential for algorithmic bias.

5.9 Enhancing the categorization of research

As highlighted in this review, AI and ML applications in environmental monitoring can be categorized more meaningfully based on their methodological contributions, application areas, and interdisciplinary approaches. Moving beyond simple categorizations by domain (e.g., air quality, water quality), this review proposes a more nuanced classification that considers the following dimensions:

1. Methodology: Studies that focus on the development of novel algorithms, such as deep learning, ensemble models, and hybrid approaches, or that enhance the interpretability and scalability of existing models.
2. Real-time: Research that integrates AI/ML with real-time data collection systems, such as IoT networks, sensor arrays, and UAVs, to enable continuous monitoring of environmental parameters.
3. Big data and remote sensing: Papers that leverage big data platforms and remote sensing technologies to analyze large-scale environmental data, particularly for applications like land use monitoring, climate change modeling, and disaster management.
4. Interdisciplinary: Studies that explore the intersection of AI/ML with other fields, such as biosensors, public health, and wildlife conservation, to address complex environmental challenges in novel ways.

By organizing the literature along these dimensions, it is expected to gain a deeper understanding of the contributions and limitations of each study. For instance, papers that focus on real-time monitoring systems can be evaluated based on their scalability, data accuracy, and ability to provide actionable insights in real-time. Similarly, research on AI/ML methodologies can be assessed in terms of its novelty, generalizability, and applicability to different environmental contexts.

This refined categorization also allows for the identification of gaps in the literature. For example, while there has been significant progress in applying AI to air and water quality monitoring, there is still relatively little research on integrating AI with emerging technologies, such as quantum computing or blockchain, for environmental data management. Future research should focus on exploring these interdisciplinary areas to unlock new possibilities for environmental monitoring and management.

6 Study limitations

Despite the comprehensive nature of this study, several limitations should be acknowledged. The bibliometric analysis conducted in this study primarily used the Scopus database, which, while comprehensive, may not cover all relevant literature, especially publications in non-English languages or those from regional journals. This limitation may result in the omission of important contributions from less prominent research regions or non-English-speaking countries. Future research could expand the analysis by including additional databases such as Web of Science, IEEE Xplore, and Google Scholar to provide a more exhaustive review of the literature. It is worth noting that the bibliometric analyses, including this one, often rely on citation counts as a measure of a study's impact. However, citation practices can vary widely between disciplines and regions, and highly cited papers may not always represent the most innovative or influential research. Moreover, newer studies may be underrepresented due to the time lag in citation accumulation. Incorporating alternative metrics, such as Altmetrics, could provide a more balanced view of research influence. Moreover, the analysis identified certain regions, such as China, the United States, and Europe, as major contributors to AI/ML research in environmental monitoring. However, there is limited representation of research from developing regions, which are often the most vulnerable to environmental challenges. This imbalance highlights the need for global capacity-building initiatives and collaboration between leading research nations and underrepresented regions.

7 Conclusion

7.1 Summary of key findings

This bibliometric analysis has provided a comprehensive overview of the research landscape in the application of artificial intelligence (AI) and machine learning (ML) techniques for environmental monitoring and data analysis. The study has revealed several key findings:

- The number of publications and citations related to AI and ML in environmental monitoring has experienced substantial growth, particularly since the late 2010s. This trend reflects the rising interest and recognition of the transformative potential of these technologies in addressing complex environmental challenges.
- The research contributions in this field are globally distributed, with China, the United States, India, the United Kingdom, and Italy emerging as the leading countries. International collaboration networks have played a significant role in advancing knowledge sharing and fostering interdisciplinary research.
- AI and ML techniques have been applied across various domains of environmental monitoring, including air quality monitoring, water quality assessment, climate change modeling, biodiversity monitoring, and disaster management. Researchers have developed and refined various methodologies, such as ensemble models, hybrid approaches, and advanced computer vision techniques, to enhance the accuracy and efficiency of environmental monitoring and data analysis.
- The integration of AI and ML with emerging technologies, such as the Internet of Things (IoT), unmanned aerial vehicles (UAVs), and passive acoustic monitoring, has opened new avenues for environmental monitoring. Additionally, the application of these techniques has expanded into interdisciplinary domains, including biosensors, disease detection, and public health monitoring.
- The advent of platforms like Google Earth Engine (GEE) and the increasing availability of big data have facilitated large-scale environmental data analysis and enabled the exploration of complex patterns and relationships. Deep learning and neural network techniques have demonstrated their effectiveness in handling these vast datasets and extracting valuable insights.

7.2 Recommendations for future research

Based on the findings of this bibliometric analysis, the following recommendations are proposed for future research in the field of AI and ML applications for environmental monitoring and data analysis:

- To address the complex challenges of environmental monitoring, it is critical to foster interdisciplinary collaborations among researchers from environmental science, computer science, engineering, and public health. These partnerships will lead to more innovative and holistic approaches, ensuring that AI and ML applications are tailored to meet real-world environmental challenges from multiple perspectives. For instance, the study by Asha et al. [3] highlights the successful integration of IoT sensors with AI models to monitor industrial air pollution, demonstrating how interdisciplinary efforts can result in real-time monitoring solutions. Similarly, the work by Gonzalez et al. [12] combining UAVs with AI-based wildlife monitoring systems reinforces the potential of cross-disciplinary approaches in addressing complex environmental issues such as biodiversity loss.
- There is a need to improve the quality and availability of environmental data, particularly in regions with limited resources. Many AI and ML models rely on large volumes of high-quality data to function effectively, but in areas with limited data availability, such as rural or developing regions, these models may underperform. Future research should focus on developing robust AI and ML techniques that can handle noisy, incomplete, or biased data. This includes advanced data fusion methods, which combine data from multiple sources, and transfer learning techniques that adapt models for data-scarce regions. The study by Athanasoulias et al. [53], which explores the development of the Plegma dataset for enhancing environmental simulations, supports this recommendation by illustrating how the development of open-access, high-quality datasets can significantly advance AI-driven environmental monitoring in underserved regions. Moreover, Caesary et al. [54] emphasizes the importance of improving data quality in cost-effective environmental applications, providing a foundation for addressing data limitations in resource-scarce areas.

- AI and ML models, despite their high predictive power, often face criticism for their "black-box" nature. This lack of transparency can limit the trust and adoption of AI systems in environmental monitoring, where decisions can have significant social and ecological consequences. Continued research on explainable AI (XAI) techniques is essential to make these models more transparent and interpretable. Enhancing the understanding of model decision-making processes will improve trust among stakeholders, including policymakers, researchers, and the general public, facilitating broader adoption of these technologies in environmental decision-making. Studies such as Lu and Ma [26], which integrate explainable machine learning methods into water quality prediction models, showcase the potential of XAI in environmental applications, making the AI decision-making process more understandable and actionable for end-users.
- Combining data-driven AI approaches with domain-specific knowledge, such as environmental principles and physical models, can significantly improve model accuracy and generalizability. Hybrid models that integrate physics-based principles into AI/ML systems hold promise for more comprehensive and interpretable environmental simulations. This integration will enhance the robustness of models, particularly in complex environmental phenomena such as climate change and ecosystem dynamics. The early study by Zhu et al. [51], which combines GIS with expert knowledge and fuzzy logic in soil mapping, exemplifies the potential of hybrid models in environmental science, leading to more accurate and context-sensitive simulations. Chapi et al. [6] applied hybrid AI models for flood prediction, demonstrating how combining machine learning with hydrological models can improve predictive accuracy and robustness in disaster management.
- Continuous exploration of emerging technologies, such as advanced sensors, quantum computing, and novel AI algorithms, offers new capabilities for environmental monitoring. These technologies can enable real-time, large-scale data analysis and unlock new potentials for environmental prediction and management. For example, the study by Anees et al. (2024) on fracking dynamics illustrates the growing need for scalable AI models that can process large volumes of data in real-time to manage environmental hazards effectively. In addition, the integration of AI with IoT, as explored by Asha et al. [3], highlights the potential for AI-enabled sensor networks to deliver scalable, real-time environmental monitoring systems that can be deployed in a variety of contexts, from industrial pollution management to urban air quality monitoring.
- As AI techniques become integral to environmental monitoring, it is essential to develop responsible AI frameworks that prioritize transparency, fairness, and accountability. Researchers should address potential biases in AI systems and ensure that these technologies do not reinforce existing inequalities. Ethical AI practices must be integrated into environmental monitoring, particularly in applications affecting vulnerable communities or ecosystems. The work by Phillips et al. [32] on species distribution models using Maxent, while highly impactful, also highlights the ethical concerns of relying on AI models that may lack transparency, particularly when used in conservation decision-making. Future research must ensure that AI models used in these contexts are not only accurate but also accountable and fair. In regions with limited resources or data availability, efforts should be made to avoid bias in AI systems that could result in skewed environmental predictions or inequitable resource management decisions.
- Global collaboration and data sharing are essential for advancing the field of AI and ML in environmental monitoring. By making environmental datasets more accessible and encouraging international partnerships, researchers can develop models that are more generalizable and applicable across diverse regions and environmental contexts. The study by Cao et al. [55], which addresses the global nature of air quality monitoring through cross-regional data integration, underscores the need for improved data-sharing practices. Similarly, Tamiminia et al. [41] emphasized the potential of global platforms like Google Earth Engine (GEE) to facilitate large-scale environmental data analysis, enabling collaboration across countries and disciplines.
- The growing frequency of natural disasters due to climate change calls for AI models that are not only accurate but also adaptable to changing environmental conditions. Future research should focus on building AI systems that can adjust to new data inputs and environmental changes in real-time. The application of AI in disaster management, as seen in Haridasan et al. [15] for landslide detection and Uddin et al. [42] in flood prediction, demonstrates the potential of these systems to provide early warnings and help mitigate the impacts of natural disasters. However, there is a need for models that can adapt to new types of disasters or unforeseen environmental changes, ensuring their long-term relevance and effectiveness.

7.3 Final thoughts

This bibliometric analysis has demonstrated the remarkable progress and transformative potential of AI and ML techniques in advancing environmental monitoring and data analysis. The findings highlight the multidisciplinary nature of this field and the collaborative efforts of researchers worldwide in addressing critical environmental challenges.

As the world confronts pressing issues such as climate change, biodiversity loss, and resource depletion, the integration of AI and ML technologies will play a crucial role in providing accurate and timely information for decision-making, enabling effective mitigation strategies, and promoting sustainable environmental management practices.

Although significant advancements have been made, the field of AI and ML in environmental monitoring remains dynamic and constantly evolving. Continuous research, innovation, and collaboration will be essential to unlock the full potential of these technologies and contribute to the creation of a more sustainable and environmentally resilient future.

Author contributions All authors wrote and reviewed the manuscript text. N.N. prepared figures. E.A conceptualized the idea and collected the data.

Funding Authors confirm that they did not receive funding for this article.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Amani M, Ghorbanian A, Ahmadi SA, Kakooei M, Moghimi A, Mirmazloumi SM, Moghaddam SHA, Mahdavi S, Ghahremanloo M, Parsian S, Wu Q, Brisco B. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J Selected Top Appl Earth Obs Remote Sens. 2020;13:5326–50. https://doi.org/10.1109/JSTARS.2020.3021052.
2. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. J Informet. 2017;11(4):959–75. https://doi.org/10.1016/J.JOI.2017.08.007.
3. Asha P, Natrayan L, Geetha BT, Beulah JR, Sumathy R, Varalakshmi G, Neelakandan S. IoT enabled environmental toxicology for air pollution monitoring using AI techniques. Environ Res. 2022. https://doi.org/10.1016/j.envres.2021.112574.
4. Baas J, Schotten M, Plume A, Côté G, Karimi R. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quant Sci Stud. 2020;1(1):377–86. https://doi.org/10.1162/QSS_A_00019.
5. Bracarense N, Bawack RE, Wamba SF, Carillo KDA. Artificial intelligence and sustainability: a bibliometric analysis and future research directions. Pacific Asia J Assoc Inform Syst. 2022;14(2):9. https://doi.org/10.17705/1pais.14209.
6. Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ Model Softw. 2017;95:229–45. https://doi.org/10.1016/j.envsoft.2017.06.012.
7. Chen G, Li S, Knibbs LD, Hamm NAS, Cao W, Li T, Guo J, Ren H, Abramson MJ, Guo Y. A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Sci Total Environ. 2018;636:52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251.
8. Chen K, Chen H, Zhou C, Huang Y, Qi X, Shen R, Liu F, Zuo M, Zou X, Wang J, Zhang Y, Chen D, Chen X, Deng Y, Ren H. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Res. 2020. https://doi.org/10.1016/j.watres.2019.115454.
9. Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, Sabath MB, Choirat C, Koutrakis P, Lyapustin A, Wang Y, Mickley LJ, Schwartz J. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 2019. https://doi.org/10.1016/j.envint.2019.104909.

10. Forzieri G, Dakos V, McDowell NG, Ramdane A, Cescatti A. Emerging signals of declining forest resilience under climate change. Nature. 2022;608(7923):534–9. https://doi.org/10.1038/s41586-022-04959-9.
11. Gibb R, Browning E, Glover-Kapfer P, Jones KE. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol Evol. 2019;10(2):169–85. https://doi.org/10.1111/2041-210X.13101.
12. Gonzalez LF, Montes GA, Puig E, Johnson S, Mengersen K, Gaston KJ. Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. Sensors (Switzerland). 2016. https://doi.org/10.3390/s16010097.
13. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press; 2016.
14. Grange SK, Lewis AC, Carslaw DC. Source apportionment advances using polar plots of bivariate correlation and regression statistics. Atmos Environ. 2016;145:128–34. https://doi.org/10.1016/J.ATMOSENV.2016.09.016.
15. Haridasan A, Thomas J, Raj ED. Deep learning system for paddy plant disease detection and classification. Environ Monit Assess. 2023. https://doi.org/10.1007/s10661-022-10656-x.
16. Hastie T, Tibshirani R, Friedman J. The elements of statistical. Learning. 2009. https://doi.org/10.1007/978-0-387-84858-7.
17. Haunschild R, Bornmann L, Marx W. Climate change research in view of bibliometrics. PLoS ONE. 2016;11(7): e0160393. https://doi.org/10.1371/JOURNAL.PONE.0160393.
18. Hengl T, De Jesus JM, Heuvelink GBM, Gonzalez MR, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B. SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE. 2017. https://doi.org/10.1371/journal.pone.0169748.
19. Hill DJ, Minsker BS. Anomaly detection in streaming environmental sensor data: a data-driven modeling approach. Environ Model Softw. 2010;25(9):1014–22. https://doi.org/10.1016/j.envsoft.2009.08.010.
20. Kamyab H, Khademi T, Chelliapan S, SaberiKamarposhti M, Rezania S, Yusuf M, Farajnezhad M, Abbas M, Hun Jeon B, Ahn Y. The latest innovative avenues for the utilization of artificial Intelligence and big data analytics in water resource management. Results Eng. 2023;20: 101566. https://doi.org/10.1016/J.RINENG.2023.101566.
21. Khatami R, Mountrakis G, Stehman SV. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens Environ. 2016;177:89–100. https://doi.org/10.1016/j.rse.2016.02.028.
22. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.
23. Li X, Su J, Wang H, Boczkaj G, Mahlknecht J, Singh SV, Wang C. Bibliometric analysis of artificial intelligence in wastewater treatment: current status, research progress, and future prospects. J Environ Chem Eng. 2024;12(4): 113152. https://doi.org/10.1016/J.JECE.2024.113152.
24. Linardos V, Drakaki M, Tzionas P, Karnavas YL. Machine learning in disaster management: recent developments in methods and applications. Mach Learn Knowl Extract. 2022;4(2):446–73. https://doi.org/10.3390/MAKE4020020.
25. Liu X, Lu D, Zhang A, Liu Q, Jiang G. Data-driven machine learning in environmental pollution: gains and problems. Environ Sci Technol. 2022;56(4):2124–33. https://doi.org/10.1021/acs.est.1c06157.
26. Lu H, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere. 2020. https://doi.org/10.1016/j.chemosphere.2020.126169.
27. Maier HR, Kapelan Z, Kasprzyk J, Kollat J, Matott LS, Cunha MC, Dandy GC, Gibbs MS, Keedwell E, Marchi A, Ostfeld A, Savic D, Solomatine DP, Vrugt JA, Zecchin AC, Minsker BS, Barbour EJ, Kuczera G, Pasha F, Reed PM. Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. Environ Model Softw. 2014;62:271–99. https://doi.org/10.1016/J.ENVSOFT.2014.09.013.
28. Mañana-Rodríguez J. A critical review of SCImago journal and country rank. Res Eval. 2015;24(4):343–54. https://doi.org/10.1093/RESEVAL/RVU008.
29. Naghibi SA, Pourghasemi HR, Dixon B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess. 2016;188(1):1–27. https://doi.org/10.1007/s10661-015-5049-6.
30. Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci USA. 2018;115(25):E5716–25. https://doi.org/10.1073/PNAS.1719367115.
31. Palani S, Liong SY, Tkalich P. An ANN application for water quality forecasting. Mar Pollut Bull. 2008;56(9):1586–97. https://doi.org/10.1016/j.marpolbul.2008.05.021.
32. Phillips SB, Aneja VP, Kang D, Arya SP. Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. Int J Glob Environ Issues. 2006;6(2–3):231–52. https://doi.org/10.1016/j.ecolmodel.2005.03.026.
33. Rana R, Kalia A, Boora A, Alfaisal FM, Alharbi RS, Berwal P, Alam S, Khan MA, Qamar O. Artificial intelligence for surface water quality evaluation, monitoring and assessment. Water. 2023;15(22):3919. https://doi.org/10.3390/W15223919.
34. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat N. Deep learning and process understanding for data-driven Earth system science. Nature. 2019;566(7743):195–204. https://doi.org/10.1038/s41586-019-0912-1.
35. Ren H, Xia M, Weng L, Hu K, Lin H. Dual-attention-guided multiscale feature aggregation network for remote sensing image change detection. IEEE J Sel Top Appl Earth Obs Remote Sensing. 2024;17:4899–916. https://doi.org/10.1109/JSTARS.2024.3362370.
36. Russell S, Norvig P. Artificial intelligence: a modern approach. 4th ed. Hoboken: Pearson; 2021.
37. Scott SM, James D, Ali Z. Data analysis for electronic nose systems. Microchim Acta. 2006;156(3–4):183–207. https://doi.org/10.1007/s00604-006-0623-9.
38. Shi, W., Zhang, M., Zhang, R., Chen, S., & Zhan, Z. (2020). Change detection based on artificial intelligence: State-of-the-art and challenges. In *Remote Sensing* (Vol. 12, Issue 10). MDPI AG. https://doi.org/10.3390/rs12101688
39. Subbarayan S, Thiyagarajan S, Karuppannan S, Panneerselvam B. Enhancing groundwater vulnerability assessment: comparative study of three machine learning models and five classification schemes for Cuddalore district. Environ Res. 2024. https://doi.org/10.1016/j.envres.2023.117769.
40. Sutton RS, Barto AG. Reinforcement learning : an introduction. 2nd ed. Cambridge: MIT Press; 2018.

41. Tamiminia H, Salehi B, Mahdianpari M, Quackenbush L, Adeli S, Brisco B. Google earth engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J Photogramm Remote Sens. 2020;164:152–70. https://doi.org/10.1016/j.isprsjprs.2020.04.001.
42. Uddin MG, Nash S, Rahman A, Olbert AI. A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. Water Res. 2023. https://doi.org/10.1016/j.watres.2022.119422.
43. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523–38. https://doi.org/10.1007/S11192-009-0146-3/FIGURES/7.
44. Villa F, Bagstad KJ, Voigt B, Johnson GW, Portela R, Honzák M, Batker D. A methodology for adaptable and robust ecosystem services assessment. PLoS ONE. 2014. https://doi.org/10.1371/journal.pone.0091001.
45. Vitolo C, Scutari M, Ghalaieny M, Tucker A, Russell A. Modeling air pollution, climate, and health data using bayesian networks: a case study of the english regions. Earth Space Sci. 2018;5(4):76–88. https://doi.org/10.1002/2017EA000326.
46. Wang L, Zhang G, Wang Z, Liu J, Shang J, Liang L. Bibliometric analysis of remote sensing research trend in crop growth monitoring: a case study in China. Remote Sens. 2019;11(7):809. https://doi.org/10.3390/RS11070809.
47. Yu W, Ye T, Zhang Y, Xu R, Lei Y, Chen Z, Yang Z, Zhang Y, Song J, Yue X, Li S. Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study. Lancet Planet Health. 2023;7: e209.
48. Yuan Q, Shen H, Li T, Li Z, Li S, Jiang Y, Xu H, Tan W, Yang Q, Wang J, Gao J, Zhang L. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens Environ. 2020. https://doi.org/10.1016/j.rse.2020.111716.
49. Zhang Q, Li X, Yu L, Wang L, Wen Z, Su P, Sun Z, Wang S. Machine learning-assisted fluorescence visualization for sequential quantitative detection of aluminum and fluoride ions. J Environ Sci (China). 2025;149:68–78. https://doi.org/10.1016/j.jes.2024.01.023.
50. Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, Ma X, Marrone BL, Ren ZJ, Schrier J, Shi W, Tan H, Wang T, Wang X, Wong BM, Xiao X, Yu X, Zhu JJ, Zhang H. Machine learning: new ideas and tools in environmental science and engineering. Environ Sci Technol. 2021;55(19):12741–54. https://doi.org/10.1021/ACS.EST.1C01339/ASSET/IMAGES/MEDIUM/ES1C01339_0005.GIF.
51. Zhu AX, Hudson B, Burt J, Lubich K, Simonson D. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci Soc Amer J. 2001. https://doi.org/10.2136/sssaj2001.6551463x.
52. Zhuang Y, Liu X, Nguyen T, He Q, Hong S. Global remote sensing research trends during 1991–2010: a bibliometric analysis. Scientometrics. 2013;96(1):203–19. https://doi.org/10.1007/S11192-012-0918-Z/METRICS.
53. Athanasoulias S, Guasselli F, Doulamis N, Doulamis A, Ipiotis N, Katsari A, Stankovic L, Stankovic V. The Plegma dataset: Domestic appliance-level and aggregate electricity demand with metadata from Greece, Scientific Data, 2024. https://doi.org/10.1038/s41597-024-03208-0.
54. Caesary D, Kim H, Nam MJ. Cost effectiveness of carbon capture and storage based on probability estimation of social cost of carbon. Applied Energy. 2025;377:124542. https://doi.org/10.1016/j.apenergy.2024.124542.
55. Cao H, Han L, Liu M, Li L. Spatial differentiation of carbon emissions from energy consumption based on machine learning algorithm: A case study during 2015–2020 in Shaanxi, China. Journal of Environmental Sciences. 2025;149:358–73. https://doi.org/10.1016/j.jes.2023.08.007.
56. Philippe, Mongeon Adèle, Paul-Hus. The journal coverage of Web of Science and Scopus: a comparative analysis Scientometrics 2016;106(1);213–228. https://doi.org/10.1007/s11192-015-1765-5.
57. Omar M, Arab MG, Alotaibi E, Alshibli KA, Shanableh A, Elmehdi H, Malkawi DA, Tahmaz A. Natural soils' shear strength prediction: A morphological data-centric approach. Soils and Foundations. 2024;64(6):101527.
58. Omar M, Alotaibi E, Arab MG, Shanableh A, Malkawi DAH, Elmehdi H, Tahmaz A. Harnessing natureinspired soft computing for reinforced soil bearing capacity prediction: a neuro-nomograph approach for efficient design. International Journal of Geosynthetics and Ground Engineering, 2023;9(4) p.53.
59. Nassif N, Junaid MT, Altoubat S, Maalej M, Barakat S. Data-Driven Prediction of The Bond Coefficient Between Fibre- Reinforced Polymer (FRP) Bars and Concrete. Special Publication. 2024;360:106–21.
60. Nassif N, Talha Junaid M, Hamad K, Al-Sadoon Z, Altoubat S, Maalej M. Performance-Based Prediction of Shear and Flexural Strengths in Fiber-Reinforced Concrete Beams via Machine Learning. Structural Engineering International. 2024:1–6.