Journal of Anatolian Environmental and Animal Sciences
(*Anadolu Çevre ve Hayvancılık Bilimleri Dergisi*)
https://doi.org/10.35229/jaes.1813761

JAES
Year: 11, 2026 (1-22)
AÇEH
Yıl: 11, 2026 (1-22)

ARAŞTIRMA MAKALESİ **RESEARCH PAPER**

The Development of Research on Machine Learning-Based Water Quality Index (WQI) Prediction: A Bibliometric Analysis

Selda PALABIYIK *
Giresun University Şebinkarahisar Health Services Vocational School, Giresun, Türkiye

Received: *31.10.2025* ***Published:*** *19.02.2026*

How to cite: **Palabıyık, S. (2026).** The Development of Research on Machine Learning-Based Water Quality Index (WQI) Prediction: A Bibliometric Analysis. *J. Anatolian Envion. Anim. Sci., **11***, 1-22. https://doi.org/10.35229/jaes.1813761

Atıf yapmak için: **Palabıyık, S. (2026).** Makine Öğrenimi Tabanlı Su Kalitesi Endeksi (WQI) Tahminine İlişkin Araştırmaların Gelişimi: Bibliyometrik Analiz. *Anadolu Çev. Hay. Bil. Derg., **11**,* 1-22. https://doi.org/10.35229/jaes.1813761

Abstract: This paper showcases the evolution of research in machine learning (ML)-based water quality index and water quality forecasting through a multi-database bibliometric and content analytical framework for the period of 2010 to 2025. The data were integrated in R through the use of bibliometrix as well as Biblioshiny, and duplicate records were removed to obtain a comprehensive data set suitable for citation and network analysis. The descriptive study combined traditional indicators, annual scientific output, author/source impact indexes, collaboration networks, conceptual mapping, and thematic evolution to identify the intellectual pillars of the field and its emerging topics. The results show a highly collaborative and growing research environment with increasing methodological complexity. There is distinctly apparent shift in the methods of ML toward deep and ensemble techniques over the recent few years. Further results of text mining and content analysis show that parameter complexity is linking very closely to model selection, which supports the use of both highly nonlinear, biochemical sophisticated architectures variables and more interpretable methods for stable physical indicators. This paper finds that explainability has become of emerging importance together with reproducibility and decision-oriented modeling in long-term water quality management strategies, into which the domain is rapidly integrating.

***Corresponding author's:**
Selda PALABIYIK
Giresun University Şebinkarahisar Health Services Vocational School, Giresun, Türkiye
✉: selda.palabiyik@giresun.edu.tr

Keywords: Water quality ındex (WQI), bibliometrics, machine learning (ML), artificial ıntelligence (AI).

Makine Öğrenimi Tabanlı Su Kalitesi Endeksi (WQI) Tahminine İlişkin Araştırmaların Gelişimi: Bibliyometrik Analiz

Öz: Bu makale, 2010-2025 dönemi için çoklu veritabanı bibliyometrik ve içerik analitik çerçevesi aracılığıyla makine öğrenimi (MÖ) tabanlı su kalitesi endeksi ve su kalitesi tahminine yönelik araştırmaların gelişimini göstermektedir. Veriler, bibliometrix ve Biblioshiny kullanılarak R'ye entegre edilmiş ve alıntı ve ağ analizi için uygun kapsamlı bir veri seti elde etmek amacıyla yinelenen kayıtlar kaldırılmıştır. Tanımlayıcı çalışma, geleneksel göstergeler, yıllık bilimsel çıktı, yazar/kaynak etki endeksleri, işbirliği ağları, kavramsal haritalama ve tematik evrimi bir araya getirerek, bu alanın entelektüel temellerini ve ortaya çıkan konularını belirlemiştir. Sonuçlar, metodolojik karmaşıklığın arttığı, işbirliğine dayalı ve büyüyen bir araştırma ortamını göstermektedir. Son birkaç yılda, MÖ yöntemlerinde derin ve toplu tekniklere doğru belirgin bir geçiş olduğu görülmektedir. Metin madenciliği ve içerik analizinin diğer sonuçları, parametre karmaşıklığının model seçimiyle çok yakından bağlantılı olduğunu göstermektedir. Bu da, hem son derece doğrusal olmayan biyokimyasal sofistike mimari değişkenlerin hem de daha yorumlanabilir yöntemlerin istikrarlı fiziksel göstergeler için kullanılmasını desteklemektedir. Bu makale, uzun vadeli su kalitesi yönetim stratejisinde açıklanabilirliğin, tekrarlanabilirlik ve karar odaklı modelleme ile birlikte giderek önem kazandığını ve bu alanın hızla entegre olduğunu ortaya koymaktadır.

***Sorumlu yazar:**
Selda PALABIYIK
Giresun Üniversitesi, Şebinkarahisar Sağlık Hizmetleri Meslek Yüksekokulu, Giresun, Türkiye
✉: selda.palabiyik@giresun.edu.tr

Anahtar kelimeler: Su kalitesi indeksi (SKİ), bibliyometri, makine öğrenmesi (MÖ), yapay zeka (YZ).

INTRODUCTION

Water is important in our daily lives and economic activities; therefore effective water resource management remains a key responsibility. Scientifically, effective water resource management falls within the discipline that emphasizes efficient supply management to respond to community demands as well as addressing disasters related to water. This includes several activities that take advantage of present technological advancements (Iqbal et al., 2023). Surface water bodies, rivers in particular, happen to be the most sensitive freshwater resource to human-induced pressures such as urbanisation and suburbanisation. Compared to lakes, reservoirs and coastal waters these systems have a greater capacity for carrying or assimilating municipal, industrial and agricultural waste. A high transport capacity means a greater risk of loss of ecological integrity due to pollutant loads (Shrestha and Kazama, 2007; Chaudhary et al., 2024; Chaudhary et al., 2025).

Water plays the major role in sustainable development. The most feasible way of assessing the health status of any water body is indexing its quality (Dash and Kalamdhad, 2021). Water quality means physical, chemical, biological and radiological characteristics of water; it is the set of parameters which can be used to predict any water quality variable or Water Quality Index (WQI) (Shalby et al., 2020). The non-linear relationships between the spatial and temporal variability typical of a catchment and water quality parameters make very difficult the prediction at any point within the catchment of those parameters that define water quality. However, predictive modeling of water quality is fundamentally important to public health and environmental protection, for sustaining resource management, and in meeting legal obligations (Sangwan and Bhardwaj, 2024). Surface water systems are stressed with pressure from climate change, rapid urbanization or changes in land use and increasing water demand for various purposes. This is the opposite mode of intended purpose hydrological models operate-in classical, non-dynamic hydrological model (Gacu et al., 2025). The environmental monitoring tools involve satellite technologies. They include large databases on hydrology providing new opportunities for the application of artificial intelligence (AI) in river basin management through machine learning (ML), since environmental monitoring tools also converge with different technologies and even more advanced nowadays, thus widening the scope of ML application in river management (Kan et al., 2020; Rozos et al., 2022; Heintzman et al., 2024).

AI precipitates a revolutionary approach to the above problems (Olawade et al., 2024). The AI applications in water resource management as earlier mentioned depict composite challenges due to climate change, unrelenting demand for water, and environmental vagaries. Therefore, AI frameworks of systems become an imperative accuracy, efficiency, and sustainability enhancer in practices applied toward managing the concept (Sharma et al., 2022; Sun et al., 2022). Recently, AI models have become a trend for assessing and monitoring water quality. The study found that trends can be identified in large volumes of data by AI models-which are capable of handling the data much more efficiently than traditional models allow because they struggle to handle such big volumes of information (Rana et al., 2023). There are different branches of computer science in which AI can be defined as the automation of human-specific processes through predictive modeling. Applications include artificial neural network (ANN) methods, support vector machines (SVMs), and hybrid models that have very effectively enhanced the management of water resources in particular river flow and water availability prediction under scant information about any hydrological regime. AI basically entered climatic hydrology formally when a neural network was developed to imitate complex systems. Haider et al. (2024) acknowledge the growing importance given of SVM and fuzzy logic for dealing with climatic variables such as rainfall and soil moisture.

Lately, ML and deep learning (DL) algorithms for data-driven modeling have been broadly applied because advances in monitoring techniques ensure a large availability of data. This represents the maximum technological advance in modeling the dynamic properties of water (Lau et al., 2019; Reichstein et al., 2019; Nong et al., 2025). A data-driven approach to ML, being the most popular area of research (Zhan et al., 2024). This method was also common in the water and environmental sciences. Where it has been used for modelling and assessment of water quality (Işık and Akkan, 2025), prediction of safe water availability (Işık et al., 2024), and modelling pH variability in marine systems (Işık et al., 2025). The same method has also been applied to predict sea surface temperatures (Akkan et al., 2022), calculating fish condition factors (Mutlu and Akkan, 2025), and in ML validation of acidification process studies (Alver et al., 2025). A few very recent works can be found that proposed new hybrid prediction models by combining fuzzy logic with Gaussian process regression (Yucesoy et al., 2025). ML is widely practiced and has now also gained wide acceptance as a useful approach in areas such as environmental modeling and water resources forecasting, marine science, and aquatic biology.

Unlike standard statistical methodologies, ML is capable of recognizing patterns inside complex and composite data. Also, as an added advantage which nowadays seems to be extremely important when compared with the old traditional physics-based models it consumes less computational resource cost (Li et al., 2021). The article therefore aims to provide a bibliometric analysis of the water quality index and ML techniques to show the trend in scientific research over the last decade. This study adds to existing knowledge by carrying out a bibliometric analysis on scientific literature that integrates ML with studies on water quality. Zhu et al. (2022) have conducted bibliometric studies that provide an in-depth and holistic view for researchers by combining these two areas. It has emphasized possible themes for future research, collaboration, and concentration. This paper sheds more light on the evolution and present context of employing ML in water quality management. A bibliometric analysis was first carried out to determine the existing state of research, highlight knowledge gaps, and frame recommendations.

Recent progress in environmental sensing, remote sensing, open data platforms and cloud computing have massively added to both the volume and diversity of water quality information. This has created new opportunities for data-driven modelling as well as real time decision support. It also creates burning questions on which algorithms and parameters are emerging most prominent , how collaboration network structures influence the diffusion of methodological approaches besides if an increasing body of evidence develops transparently with respect to policy relevance.

The present study develops its methodological base to conduct a wide bibliometric analysis accounting for the intellectual, thematic, and collaborative development of ML -based studies on WQI prediction from 2010 to 2025. A bibliographic dataset was built by merging records from two leading databases: Web of Science Core Collection (WoS) and Scopus to reduce single-source dependence and increase disciplinary coverage through a standardized search strategy restricted only to certain fields. It thus becomes imperative to differentiate this study from earlier works so as to highlight its scientific importance. Most earlier works related to ML in hydrology have either been based on a single database or contain an insufficient depth analysis regarding the evolution of algorithms. Table 1 provides a comparative summary between some selected representative previous works with that addressed clearly demonstrated gap.

Table 1. Comparison of this study with existing bibliometric and review studies in the literatüre.

Study	Scope	Database(s)	Techniques
Muñoz-Alegría et al. (2025)	ML and DL applications for freshwater (surface and groundwater) quality prediction, with emphasis on research trends and future directions.	Scopus (2000-2024).	Bibliometric-assisted systematic literature review (B-SLR); topic modeling for document clustering; network-based science mapping.
Li et al. (2024)	AI in wastewater treatment, covering monitoring, control, optimization, and treatment performance rather than water-quality prediction per se.	Web of Science Core Collection (2000-2023).	Bibliometric analysis using CiteSpace, VOSviewer, and Bibliometrix; keyword, country, and institutional networks.
Li et al. (2025)	ML for water quality prediction (field-level evolution and hotspots).	Web of Science Core Collection (literature spanning 1993-2023; large-scale dataset reported).	Visual bibliometric analysis using CiteSpace and R/Bibliometrix to map research fronts, keywords, and disciplinary evolution.
Gao et al. (2023)	Climate change–water quality nexus, mapping the intellectual structure (not ML-methodology focused).	Web of Science Core Collection (1990-2022).	Bibliometric mapping (VOSviewer/Bibliometrix/CiteSpace); co-citation and keyword evolution.
This study (ours)	Focuses specifically on ML-driven water-quality prediction and assessment, structuring the literature by target variables/indices, water-system context, and data modalities, rather than trends alone.	Employs a multi-database retrieval strategy (e.g., Scopus and Web of Science) with fully reproducible search strings, screening criteria, and deduplication procedures.	This study provides a reproducible summary review of the main AI model families applied to WQI and water-quality studies, the indicators emphasized, and most frequently reported evaluation metrics. Dual-database summaries include conceptual structure and co-citation analysis together with an overlay content analysis addressing titles, abstracts, or keywords-supplementing results of dual-database bibliometric mapping that comprises thematic evolution (trends) analysis-from Scopus and Web of Science databases.

The easiest way to underline the scientific importance of this study is to differentiate it from previous research findings. Previous works spoke about the application of ML in hydrology, or probably water quality studies, but they were limited either by a single database or did not provide an elaborate discussion on the development of algorithms. Table 1 provides a clear comparative analysis that explicitly shows the gap targeted by this study with selected studies from relevant literature.

The dataset was processed using R (bibliometrix) and analyzed in Biblioshiny, with additional science mapping analysis methods such as annual scientific production, source dynamics, author and country collaboration networks, keyword co-occurrence network, thematic map and development, and co-citation-based intellectual structure analysis. The study substantially extends traditional bibliometric mapping by conducting an in-depth content analysis of titles, abstracts and keywords to determine the frequency and temporal distribution of major families of AI models, such as neural networks, DL architectures, ensemble learning, boosting-as well as their association with commonly modelled water quality parameters-validation metrics. This is a multi-database bibliometric study implemented together with method-oriented content analysis to develop a reproducible summary on how algorithmic choices-data aspects-collaboration patterns have evolved together; finding research gaps related explainability-standardization-translating AI based WQI modeling into reliable decision making tool for water quality management.

METHODOLOGY

Data Collection: The dataset for this study was compiled from WoS, the two most prestigious sources in the field due to their advanced data infrastructure for bibliometric analyses and comprehensiveness they offer in the literature (Alshater et al., 2021).

The bibliographic corpus of this study was compiled from the two leading scientific databases, Web of Science (WoS) and Scopus. The technically most refined example for that reason is WoS, and it has been selected because from citation matching algorithms, WoS contains a very well structured index on the rationale most of the influential literature worldwide. This ensures methodological consistency with other studies in the domain (Serter and Gumusburun Ayalp, 2024; Işık and Palabıyık, 2025). The richness of bibliometric indicators in the Web of Science, together with validated citation networks, also make it a preferred source for consistent data in scientific mapping studies. These features have established WoS a baseline reference source for high-quality research used for interdisciplinary analysis across the globe (Işık and Palabıyık, 2025; Işık, 2025; Odabaş Alver, 2025; Tekin, 2025). A bibliometric review is defined as the application and cross-application of mathematics, information sciences, computing, and visual analytics to scientific output. The best characterization of a bibliometric review contains an element that quantitatively measures scholarly literature at the interface between ML and water quality. However, this can also be construed to mean any study laying emphasis on quantitative assessment of academic literature at said interface (Md Shahri et al., 2024).

A dual database design was chosen to increase the disciplinary scope and decrease the potential source bias found in single database bibliometric studies. Records were retrieved and exported on 03 January 2026.

Search Strategy: A restricted field search was applied in both WoS and Scopus for studies explicitly dealing with WQI or water quality prediction using AI methods. To make the retrieval more topically accurate and reduce conceptually unrelated retrieval, records were required with WQI related terms in the Title field and AI terms and prediction intent terms in the Abstract field. Core Collection of the Web of Science. The search query was structured as: TI = ("water quality index" OR "WQI" OR "water quality") AND AB = ("machine learning" OR "deep learning" OR "artificial intelligence") AND AB = (predict* OR forecast* OR estimat*) resulting in an initial search yield of 619 records. Through the interface in WoS, further refinements have been made on the records to ensure their relevance and comparability: years of publication between 2010-2025, document type article, language English, and web of science index SCI-EXPANDED.

Web of Science Categories were used as a filter to make the search relevant to the disciplines. These categories include Environmental Sciences, Engineering Environmental, Water Resources, and Remote Sensing. Geosciences Multidisciplinary; Green Sustainable Science Technology, Computer Science Interdisciplinary Applications. Computer Science Information Systems; Computer Science Artificial Intelligence and Engineering Multidisciplinary. Multidisciplinary Sciences, Environmental Studies, Public Environmental Occupational Health, Marine Freshwater Biology, Ecology, Statistics Probability, and Mathematics Interdisciplinary Applications are also included as categories. The WoS dataset has 598 records after all these filters have been applied.

To make Scopus results eligible under WoS criteria, the following parameters were applied: publication years 2010-2025, document type Articles, language English, and subject areas Environment, Engineering, Computer Science, Agriculture and Earth and Environmental Sciences Mathematics Decision Sciences Multidisciplinary. The unified search query was:(TITLE ("water quality index" OR "WQI" OR "water quality") AND ABS ("machine learning" OR "deep learning" OR "artificial intelligence") AND ABS (predict* OR forecast* OR estimat*). Initial retrieval from Scopus counted 1,338 records. After refinement the Scopus export contains 822 records.

Eligibility Criteria and Filters: To make the databases comparable, we applied the same inclusion principles. We limited the time period from 2010 to 2025 and only included peer-reviewed journal articles in English. We restricted the subjects to keep them environmentally and water related relevant as well as AI oriented methodological domain relevant.

In WoS, records were additionally restricted to SCI-EXPANDED and to a predefined set of relevant WoS categories. In Scopus, subject areas were restricted to the specified Scopus subject classifications Environmental Science (ENVI), Engineering (ENGI), Computer Science (COMP), Agricultural and Biological Sciences (AGRI), Earth and Planetary Sciences (EART), Mathematics (MATH), Decision Sciences (DECI), and Multidisciplinary (MULT). The dataset was considered topically relevant with no need for any additional manual screening because the search strategy required explicit WQI related terminology in the Title field, thereby providing high query specificity and substantially reducing the likelihood of off-topic retrieval.

Data Export Procedures: WoS records were exported in Plain Text format with both Full Record and Cited References to preserve cited reference data necessary for any reference-based analyses such as co-citation mapping. The WoS records were downloaded in two batches because of export limits. Scopus bibliographic data, abstracts, and keywords were exported in BibTeX format. Character encoding normalization was implemented on import to avoid parsing problems resulting from non-ASCII characters in author names.

Data Preprocessing and Integration in R: Records from WoS Core Collection were analyzed using Bibliometrix R package (version 2025.05.0) and its Biblioshiny interface. All preprocessing and integration steps took place in RStudio with the bibliometrix package. First, two WoS export files were transformed into bibliographic data frames and merged after harmonization of common metadata fields. Second, the Scopus BibTeX file was transformed into a bibliographic data frame after encoding normalization. Third, the WoS and Scopus datasets were integrated by mergeDbSources() which does cross-database matching based on identifiers such as DOI, title, authors, publication year etc., and removes duplicates. After deduplication there were 841 unique peer reviewed journal articles published between 2010 to 2025 in final corpus.

Preservation of Cited References and Quality Assurance: Cited references were preserved by exporting WOS records with cited references and checking if the field for cited reference was still available after merging. A completeness check at the dataset level in Biblioshiny, where reference data accessibility for any reference-based analyses was checked, ensured that all layers of data contained fully accessible reference information.

Bibliometric Analysis and Visualization: Bibliometrix is a specialized R package that allows advanced quantitative bibliometric and scientometric analyses to be carried out within a single tool (Aria and Cuccurullo, 2017). R is very powerful, flexible and free software for statistical analysis, data manipulation and graphical display (Rodríguez-Soler et al., 2020). Bibliometric analysis was run on the unified dataset in bibliometrix and explored interactively through Biblioshiny. The workflow steps comprise: descriptive metrics, annual scientific production, source dynamics, author productivity, co-authorship network at different levels (authors-institutions-countries), keyword co-occurrence network analysis with thematic evolution and conceptual structure mapping accompanied by a comprehensive depiction of the field's intellectual and social structure through co-citation analysis.

Data Mining and Content Analysis: A content analysis was carried out to provide a methodological insight into the supplementation of bibliometric mapping, identifying AI algorithm families and frequently modeled water quality parameters. Titles, abstracts, and author keywords were extracted using Python-based tools as well as a rule-based dictionary that contains the names of algorithms together with common variants. The study focused on model families that were selected as targets and are now widely accepted for use, such as ANNs, DL architectures (CNN, LSTM, RNN), Random Forests (RF), SVM, Decision Trees (DT), Reinforcement techniques (XGBoost, AdaBoost), and Generalized Linear Models. In other words , target models belonged to the most popular classes of ML methods. Co-occurrence analysis was later used to check the associations between model families and water quality indicators (pH, dissolved oxygen, turbidity). Model performance metrics (R^2, RMSE, MAE) were also co-grouped to expose any trend in reporting model performances by different studies. The workflow systematically depicts bibliographic record retrieval from WOS and Scopus for ML–based water quality research, their harmonization, integration accompanied by a quality assurance step before content analysis as shown in Figure 1.

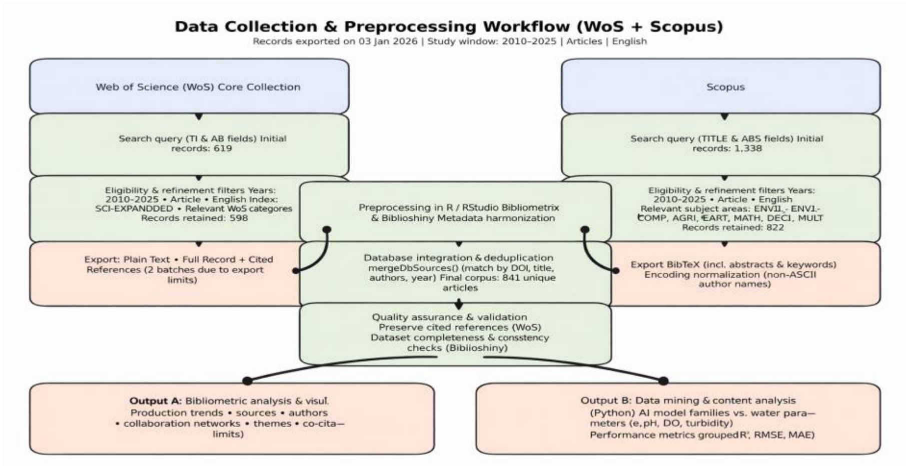

Figure 1. Workflow of data collection, preprocessing, and analysis using WoS and Scopus.

RESULTS

Overview of the Bibliographic Dataset (2010–2025)

This bibliographic dataset was built by merging the records of WOS and Scopus from 2010 to 2025, a period marking both an extremely productive research landscape and one rapidly consolidating in collaboration (Figure 2). The corpus contains 841 documents published in 293 unique venues. Wide coverage increases the probability that the sample is representative. The annual growth rate is 39.34%. The average age of the documents is only 2.69 years. There has obviously been some very recent hot topic among scholars. Meanwhile, the academic influence reflected by an average number of citations per document equal to 23.21 is quite impressive for a field where, generally because it is so dynamically new, there tends to be a time lag in accumulating citations. The number of 24.355 references speaks to a very sound intellectual base and intense connectivity with the rest of the literature. The similarity in numbers between Keywords Plus (2.280) and Authors' Keywords (2.151) may reflect similarity between the indexing mechanisms and conceptual frameworks of the authors.

There are 2,826 identified authors. Only 42 papers are single-authored. The mean value of co-authors per document is 4.91. Apparently, knowledge production is mainly collaborative. Share of publications with international co-authorship is 27.82%. This proves the intensive process of globalization and cross-border scholarly cooperation. Mainly the dataset contains research articles (811). However, it also comprises a small number of documents indexed as early access publications (17). Both reflect a fast publication process and an active developing research agenda.

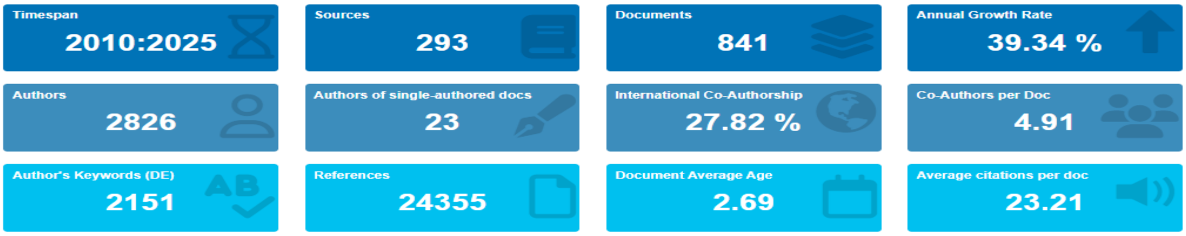

Figure 2. Basic descriptive statistics of the bibliographic collection from the combined WOS-Scopus dataset (2010-2025).

Growth Pattern and Forecast of Scientific Production: A logistic growth curve can best fit the scientific output over time within this dataset, showing an initial low activity phase, entering a marked high growth rate period to eventually predicted slowdown as the field becomes saturated (Figure 3). The carrying capacity (K) is estimated at 3.665 publications; thus, the cumulative number of publications will remain almost steady at that value. The maximum annual publication rate-per-year is attained around 2027-2028 with 472 publications per year at the inflection point of maximum growth.

Accuracy measurements (R^2 = 0.994, RMSE = 6.83) strongly support the model's robustness in representing observed publication trends. By 2025, a total of 841 publications will have been recorded in the field, representing only 22.9% of the estimated saturation level. The value falls well within the rapid growth region. The projections show very high growth in the next few years, which slowly starts decreasing as cumulative output nears its supposed maximum limit-at which point it is considered to have shifted from an expansion mode into a consolidation one.

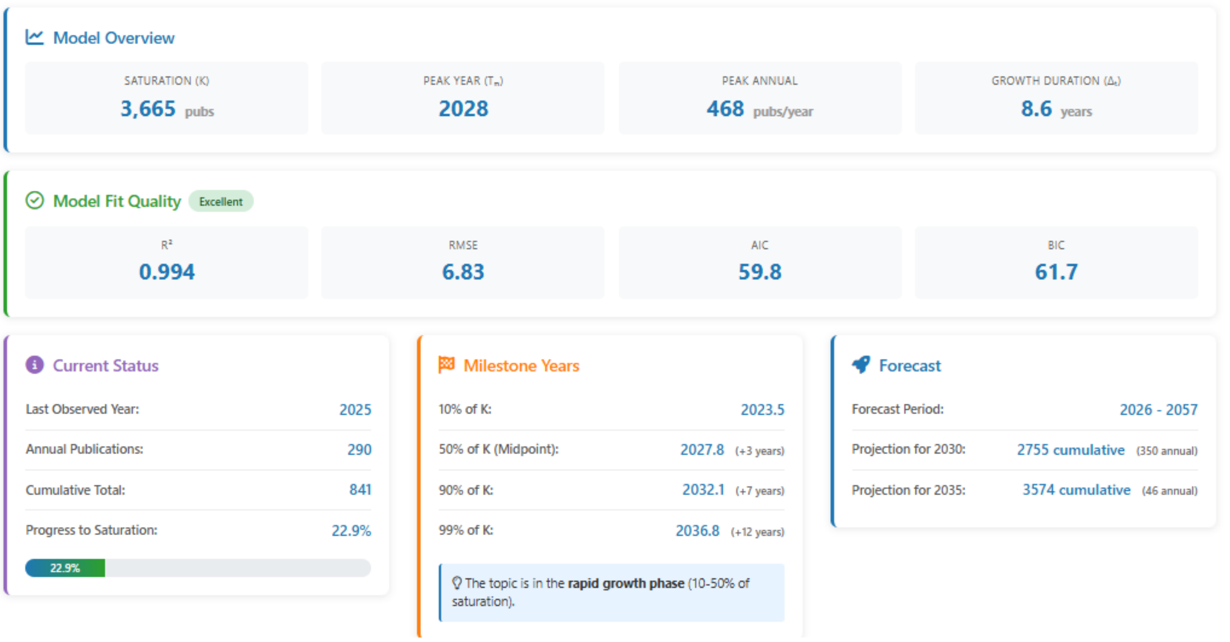

Figure 3. Logistic growth model of scientific production and forecast.

Three-Field Relationship between Keywords, Countries, and Sources: The three-field plot presents a highly organized but cross-linked map of research between leading author keywords, major countries, and preferred journals or sources of publication (Figure 4). Key thematic elements occupy top positions within the keyword field. These include ML, WQ, and DL. They are strongly linked with a small group of very prolific countries-namely China, India, and the United States. In addition to its multiple strong links through several high-frequency keywords, China appears to be pursuing an intensive and well-focused methodological line of research.

The country-to-source relationship analysis further discloses that these top contributing countries largely channel their output through a fairly steady group of core journals, including Journal of Environmental Management, Journal of Hydrology, Water Research and Science of the Total Environment. This fact emphasizes the already well-entrenched publication channels in the discipline. In the absence of links between keywords and journals, their mutual connection with countries suggests national research profiles within which certain methodological themes are systematically associated with preferred publication outlets. The flows emphasize the increasing integration of thematic specialization, geographic concentration, and top journals indicating an advanced but steadily growing state of the research ecosystem.

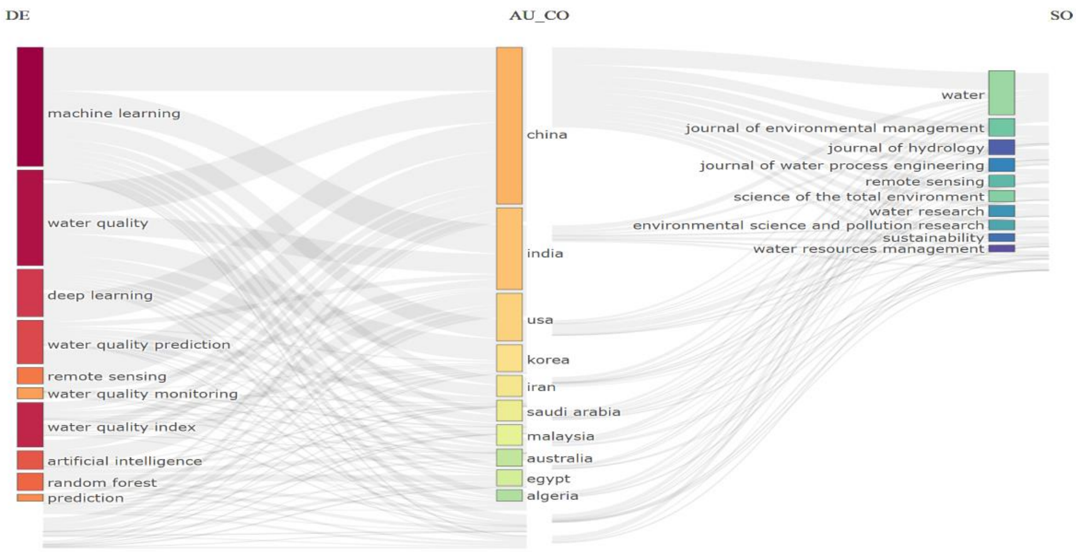

Figure 4. Three-field plot of keywords, countries, and sources.

Conceptual Structure of the Research Field: The thematic map describes the conceptual structure of the research domain by positioning major themes at levels of central and developed maturity, thus also indicating emerging and predominant knowledge (Figure 5). Motor themes are identified in clusters such as ML, water quality, water quality index, and water quality monitoring. These appear in the upper-right quadrant of the map which means that they are well discussed both within themselves and with other topics in general academic discussion. Linking them to recent blockbuster papers shows that ML based assessment and monitoring frameworks are the current intellectual core of the field. On the other hand, themes such as DL, water quality prediction, and Long Short-Term Memor (LSTM) networks appear to be basic themes with high centrality but significantly lower internal density; these represent nascent yet rapidly growing research threads.

Niche themes are tightly bound inside themselves and extend with loose links into the larger thematic network. The subdomains that have been detected are mainly methodological: WQI based ML models, turbidity focused studies, and explainability oriented approaches (for example SHAP) alongside algorithms such as CatBoost and LightGBM. Remote sensing, water quality parameters, and ML algorithms lie either at the periphery or occupy a transitional zone between emerging content and core content-this is a research avenue of substantial future potential that has not yet consolidated into any central structure within the literature.

Keyword Co-occurrence Network Structure: The keyword co-occurrence network is well organized with themes forming layers beginning from the methodologies and ending at the application domains, connecting intricately linked methodological approaches and application domains (Figure 6). Two highly central nodes 'water quality' and 'ML' anchor this network to serve as main integrative hubs between several thematic communities. The largest cluster emphasizes the use of ML techniques in water quality assessment, management, and pollution control, thus reflecting an applied orientation of the discipline.

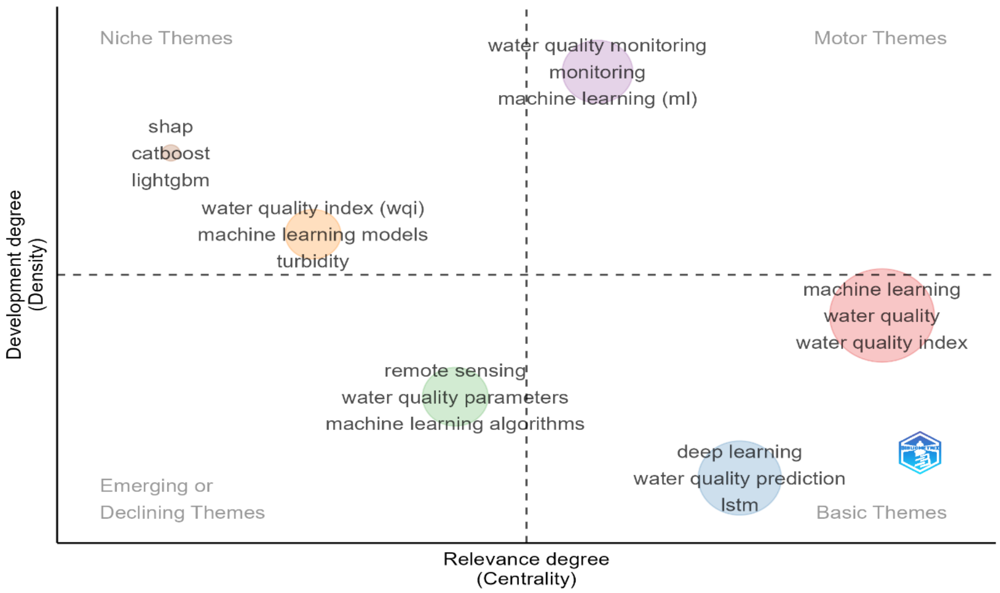

Figure 5. Thematic map of conceptual structure.

Keyword Co-occurrence Network Structure: The keyword co-occurrence network is well organized with themes forming layers beginning from the methodologies and ending at the application domains, connecting intricately linked methodological approaches and application domains (Figure 6). Two highly central nodes 'water quality' and 'ML' anchor this network to serve as main integrative hubs between several thematic communities. The largest cluster emphasizes the use of ML techniques in water quality assessment, management, and pollution control, thus reflecting an applied orientation of the discipline.

A clear methodological theme appears in the continued work on developing and fine- tuning models, clustered around regression, classification, algorithms, and model performance. Forming a strong subnetwork with themes of monitoring and remote sensing are parameter-specific themes chlorophyll-a, phosphorus, nitrogen, turbidity in which physically interpretable water quality indicators remain relevant to current research. Meanwhile, there is a small but very well-defined cluster concentrating on DL, LSTM and water quality prediction. This reflects advanced time-series and non-linear modeling techniques gradually being taken up. The heavy density of links between clusters shows strong cross-fertilization between new methods and specific areas of application. In this connection, the field has developed an integrated research framework as opposed to maintaining isolated thematic silos.

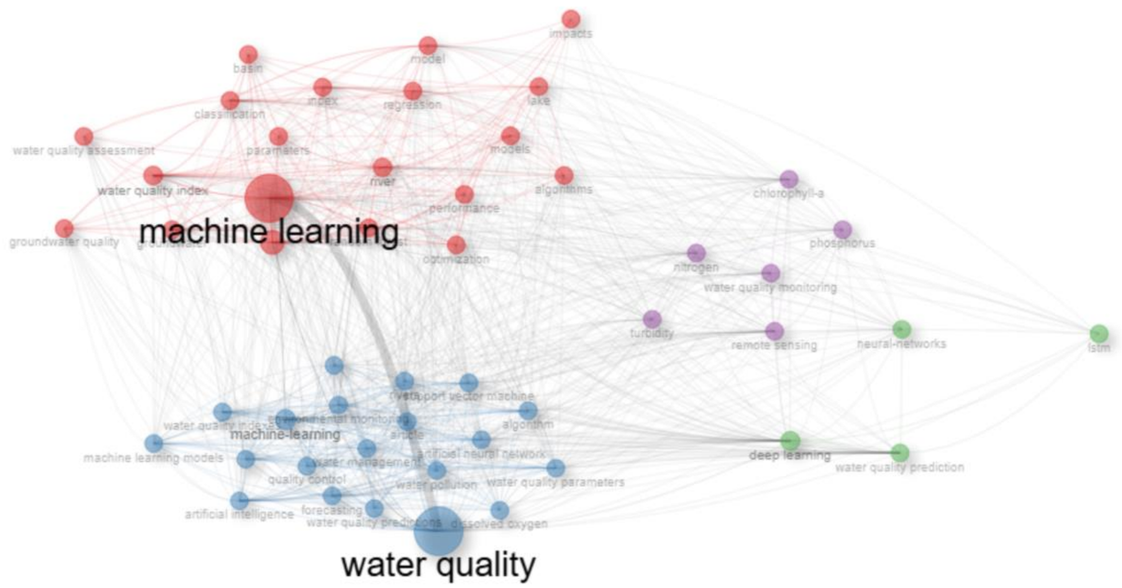

Figure 6. Keyword co-occurrence network of the research field.

Intellectual Structure of the Research Field: The paper-level co-citation network displays a modular intellectual structure, indicating that the research field is organized around clearly defined and coherent knowledge bases (Figure 7). A major founding cluster is provided by Breiman (2001), emphasizing the lasting influence of classical ML approaches in the literature-decision trees and ensemble learning, among others. A dense, recent cluster with connections to other clusters lies in publications between 2019 and 2022. These include Lu et al. (2021), Barzegar, (2020) and Ahmed et al. (2019). This particular cluster provides an analysis of how applied ML frameworks specifically designed for water quality assessment and environmental modeling are consolidated. At the same time, there is a clearly defined and extremely coherent cluster led by Uddin et al. 2022 (2022–2023) and related studies which represent a new, hot research front with strong recent citation activity characterized by continuing methodological development. Strong links between clusters indicate that modern applied research still invokes both classical ML methods as well as "intermediate" innovations in methodology, thereby demonstrating an additive-and-integrative evolution at the intellectual core of this discipline.

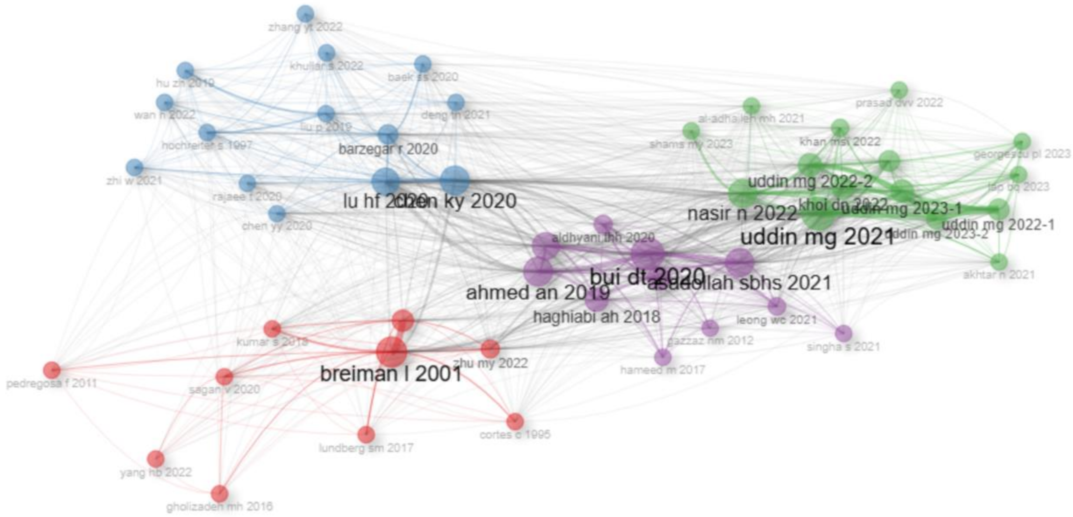

Figure 7. Paper-level co-citation network.

Global Collaboration Structure of the Research Field: International collaboration of a globally networked research community with a few major hubs and several regionally cohesive communities can be described (Figure 8). In terms of research output and strength of collaborative ties, the United States (USA) and China are most influential. They play core roles in sustaining scientific exchange across continents.

Strong bilateral ties between the USA and China, along with thick connections binding these countries to Western Europe, Canada, Australia and East Asia point to a heavily transatlantic and transpacific partnership influencing knowledge production in this field. Outside of these main hubs, India appears as an important collaborative hub within a Middle Eastern-South Asian (and partly African) network signaling emerging regional leadership. Community detection methods reveal a China-centric cluster with strong linkages to the United Kingdom and other countries, a USA-centric cluster that incorporates partners from various geographic locations, smaller European and Asia-Pacific clusters. This gives an integrated yet hierarchical organization of collaboration at the global level, wherein leading scientific powers assume the role of anchor connecting established as well as newly emerging newcoming research communities.

Distribution of AI Algorithms: The distribution of algorithms used in WQI prediction is given in Figure 9. ANN appear to be the most popular approach, appearing in 284 studies. This is followed by DL methods, represented in 261 studies, and RF algorithms, featured in 202 studies. A big share for ANN because hydrological processes have proven effective at capturing complicated non-linear behaviors within systems and between parameters inside a system.

Nonetheless, the substantial application of traditional algorithms such as RF and SVM, which together account for over 30% of the studies, is significant. This trend clearly reflects an obvious inclination among researchers to prioritize not only the predictive performances-but also other supplementary criteria in the selection of best-suited models for interpretation possibility and computational efficiency.

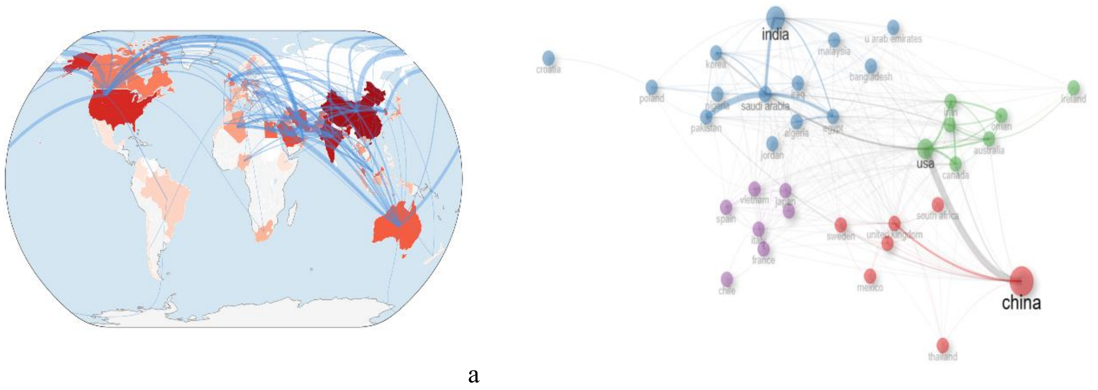

Figure 8. International collaboration network by country.

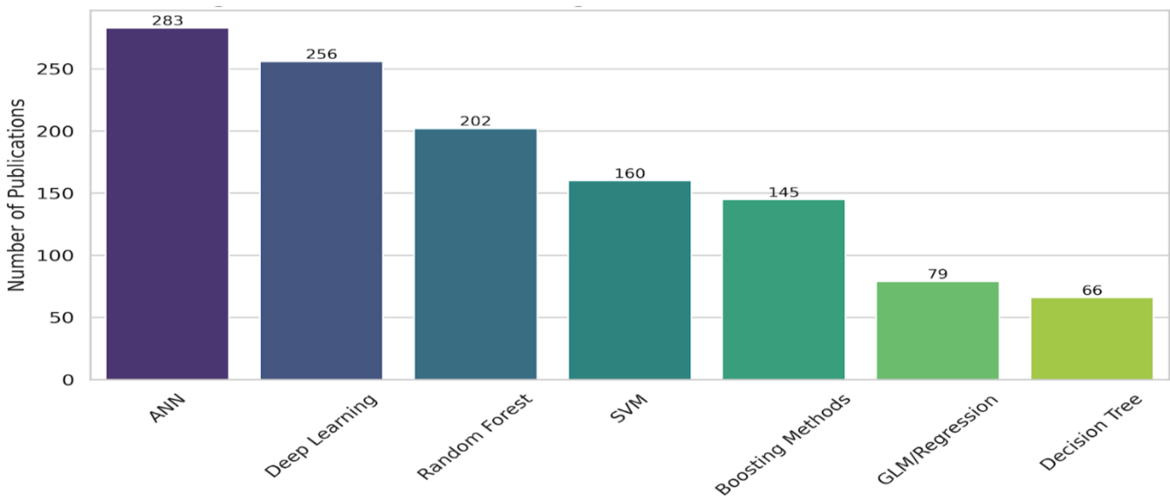

Figure 9. Distribution of the most frequently used AI algorithms in WQI prediction research (2010–2025).

The frequencies in Figure 9 show that ANN and DL techniques are most prominent within the literature of research on WQI estimation. RF comes next in frequency while SVM/Regression appears much less frequently, mostly as parts of comparison studies. This trend is apparent with current application literature-for instance, research developing ANN-based WQI frameworks using multiple input parameters demonstrates that ANN is directly applied as the main model in WQI contexts (Banda and Kumarasamy, 2024). Similar comparative assessments of LSTM with the classical methods, RF and SVM, within an identical experimental framework on time-series water quality data that report better performance in some cases also explain the sudden rise of DL for WQI estimation (Han et al., 2025). Meanwhile, RF's strong presence is also supported by its frequent use as a baseline method in multi-model comparisons; recent comparative studies on WQI prediction and classification consistently test RF alongside SVM and boosting techniques, with RF delivering competitive outcomes. Another fact making the results more comparable across different case studies is that the most frequently applied algorithms belong to three groups: SVMs, ensemble methods (boosting or bagging trees, mainly RF and Gradient Boosted Trees), and ANNs (Hridoy et al., 2025).

The explanation for this, as seen in Figure 9, is that SVM has mostly been used as a baseline method rather than being reported as the best performing method or being part of any hybrid models. For example, in hybrid approaches combining RF and SVM , this is explicitly described as a "hybrid" design (Sakaa et al., 2022). Overall, Figure 9 suggests that the field is currently expanding mainly along the ANN/DL path, with RF remaining a robust and widely used method, while SVM/SVR holds a strategic role primarily in comparative and hybrid model contexts. RF can be considered among the most popular and resilient methods meanwhile SVM/SVR hold more particular strategic roles mostly in comparative and hybrid model contexts.

Temporal Evolution and Technology Trends: Temporal trends of the algorithms are illustrated in Figure 10. ANN and SVM show a rather steady pattern from 2010 to 2018. There is an upward trend that starts to grow dramatically after 2019 in DL and Boosting techniques such as XGBoost implementations. This can be related to recent huge improvements in environmental sensing technology and big data approaches applied to water management that started around this same time. What is notable from 2024 and 2025 results, DL approaches are steadily outperforming the conventional algorithms to become the main approach. This transition of DL approaches that outperform the conventional ones is justified by increasingly large and complex water quality datasets.

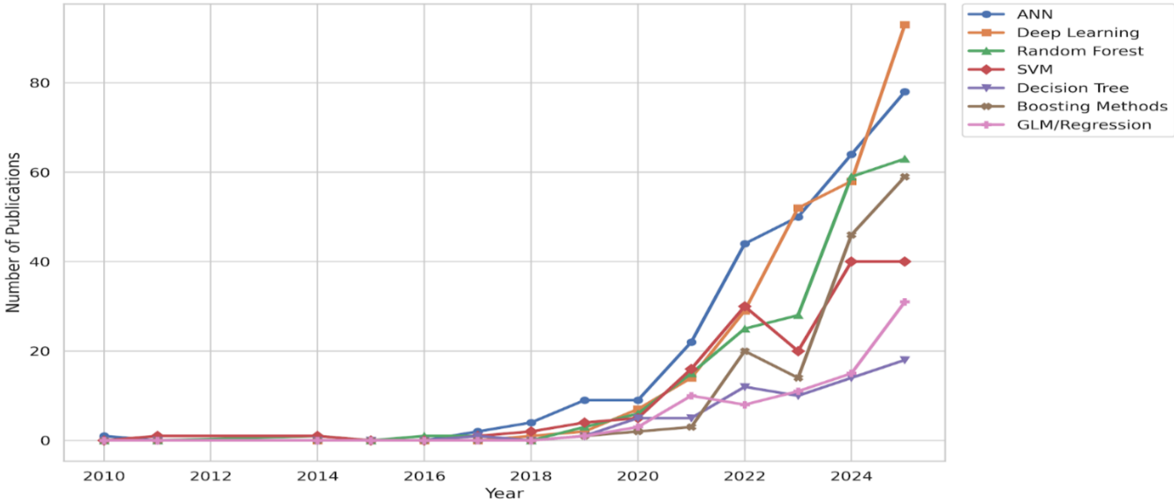

Figure 10. Temporal evolution of the top 5 ML models used in water quality studies from 2010 to 2025.

Parameter Complexity and Model Fit: The preferred algorithms for modeling various water quality parameters are displayed as a heatmap in Figure 11. There appears to be an association between the type of parameter and model choice: Physical parameters such as turbidity and temperature are mostly predicted by traditional models (RF, ANN), whereas DL is more frequently used for nitrogen- and phosphorus-related variables and fecal coliform–three parameters with complex biochemical interactions plus large uncertainty in measurement. Increasingly, researchers seem to be using even more advanced architectures such as LSTM or convolutional neural networks (CNN) to deal with complicated/noisy datasets.

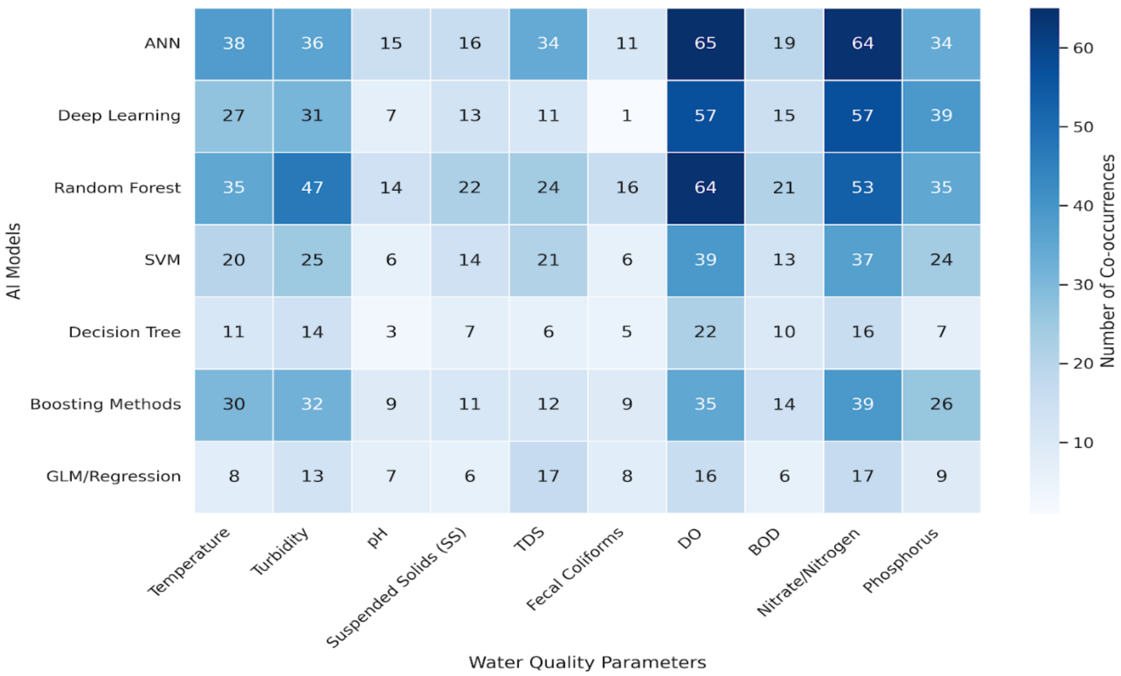

Figure 11. Co-occurrence heatmap showing the relationship between water quality parameters and AI models.

Figure 11 displays a co-occurrence analysis between model and parameter to find out which water quality parameters are most available in the literature for WQI estimation. Dissolved oxygen and nitrogen compounds (nitrate/total nitrogen) appear with very high frequencies among different AI models, followed by turbidity, phosphorus, and temperature. Parameters that appear with low frequencies in this review study are pH and fecal coliform. This is quite similar to recent research works of WQI modeling.

For example, in the study by Xu et al. (2024) on coastal urban areas, DO and nutrient salts, particularly phosphorus and nitrogen species, were found to be important predictors of WQI. Makumbura et al. (2024), through a SHAP-based explainability analysis result found that DO and biochemical oxygen demand (BOD) are the main contributors toward variability in WQI with turbidity and nitrate having secondary yet consistent influences.

Bibliometric results show that parameters indicative of oxygen dynamics and nutrient loading form the core elements in research on WQI estimation, supported by physical parameters such as turbidity and temperature, while pH and microbiological parameters receive much less attention, often very specific to individual study contexts.

Scientific Impact Analysis: A citation-based assessment of the methodological influence (Figure 12) provides a clear picture that impact and growth in publications do not always accompany each other. Though research studies related to DL are growing rapidly, the average citations per paper for these studies remain lower than those for already established techniques such as SVM and ANN. This trend mostly aligns with the shorter citation window of many recently published papers on DL.

Interestingly, Decision Tree-based studies show the highest average citations per paper; yet, this should be viewed with caution as the mean may be overstated by a small number of highly cited papers and by the frequent use of decision trees as a transparent baseline in comparative WQI and water quality modeling studies. However, boosting-based methods show high visibility in citations most probably due to their use as common competing learners in applied WQI tasks and general availability from standard software toolkits. While there is substantial attention on new approaches proposed, the literature still concentrates on a few well-recognized model classes that are dependable and simple to apply. Apparent leadership in average citation metrics should be construed with respect to the age of citations together with the uneven distribution of citations rather than being taken as a direct indicator of methodological development.

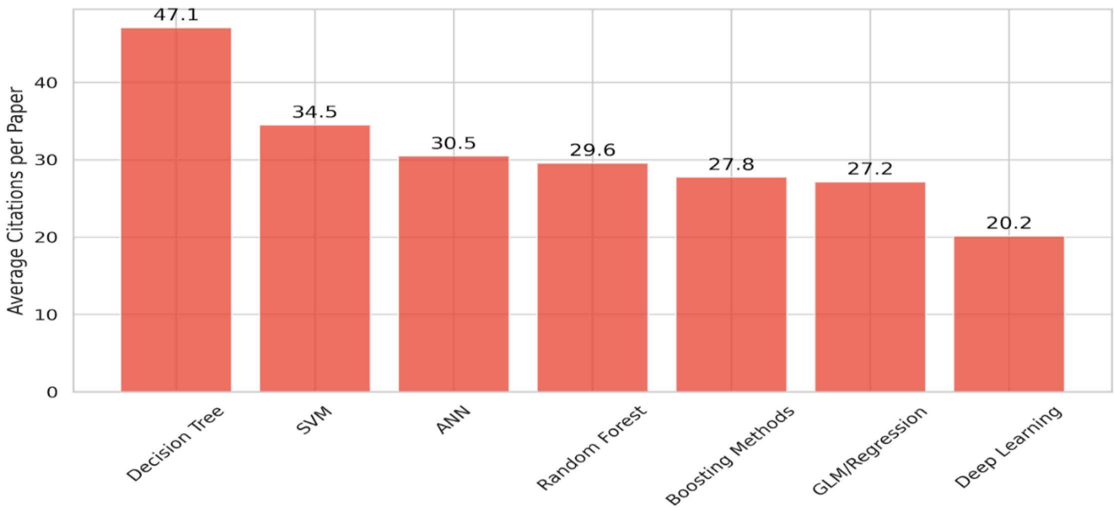

Figure 12. Comparison of the scientific impact of water quality prediction models based on average citation counts.

Country Strategies and Policy: The choice of AI model families by country reflects preferred methodologies rather than a single global standard (Figure 13). China leads in sheer volume of research with more than 240 publications and demonstrates an apparent preference for DL architectures. This is not manifested as a reliance on one single method, however, since there is strong parallel usage of ensemble methods such as RF and Boosting.

India is the high-capacity eclectic model, being the second-largest contributor while keeping a more diverse portfolio and allowing traditional algorithms such as SVM and Decision Trees to remain relatively active compared with other top countries. The United States presents perhaps the most interesting deviation; despite a lower total volume than China's, much of its research falls into this 'Other' category, apparently indicating that hybrid or non-standardized frameworks frequently experimented with by U.S. researchers instead of adhering strictly to mainstream DL models.

This pattern of a strong current DL wave mixing with a long "tail" of other methods applies in moderate contributor countries such as Iran and South Korea. In the final analysis, even though there can be no denying or doubting the fact that this field is moving towards deep and ensemble learning, contextualized at regional research settings across Asia, what the data speaks volumes about is significant methodological diversity rather than single-solution convergence.

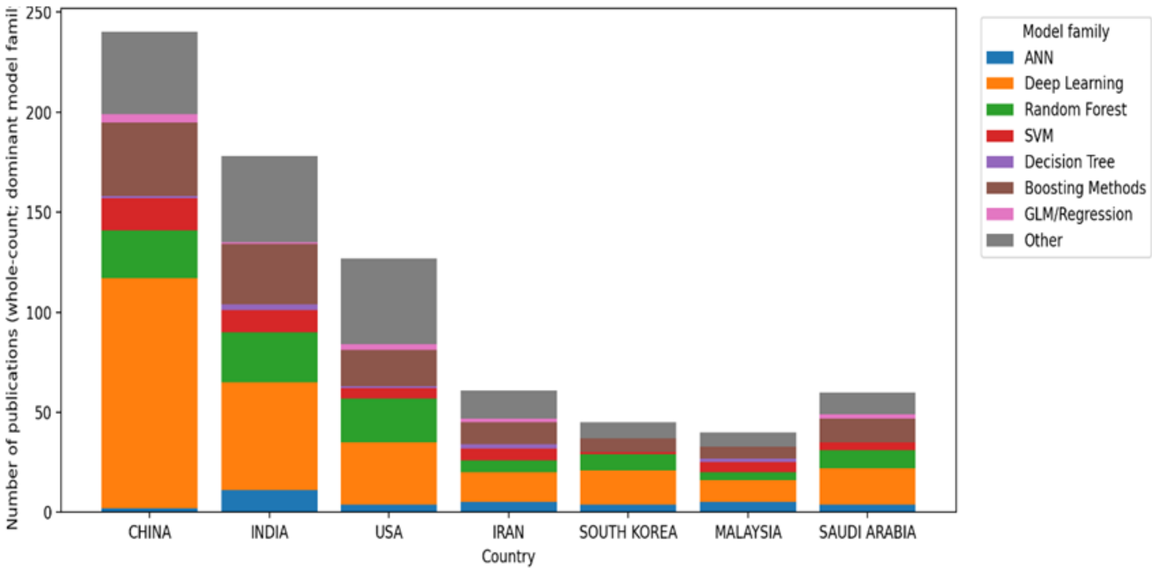

Figure 13. Country distribution of publications by dominant AI model family.

Validation Metrics: The statistical criteria that were used in most studies to check and validate the reliability of their model are given in Figure 14. The results show a clear preference for accuracy. In 342 research findings, most water quality assessments are considered classification tasks where WQI levels are categorized into different status levels. RMSE (n = 173) has been frequently reported with MAE (n = 106) in terms of performance evaluation. This is very consistent with common practice, which associates a highly sensitive metric normally penalizing large deviations with a more robust absolute error measure.

The relatively high count for R^2 (n = 155) displays its status as the most common and regular complementary indicator to any error metric, summarizing variance proportion explained by the model. MSE (n = 105) is reported much lesser in comparison with RMSE ,therefore suggesting a preference among authors to report it in its root transformed form so that it becomes interpretable in original units. Overall Figure 14 displays that model reliability has been predominantly conveyed by this research as a combination of classification performance metrics with regression error and fit statistics rather than through some single standardized evaluation method.

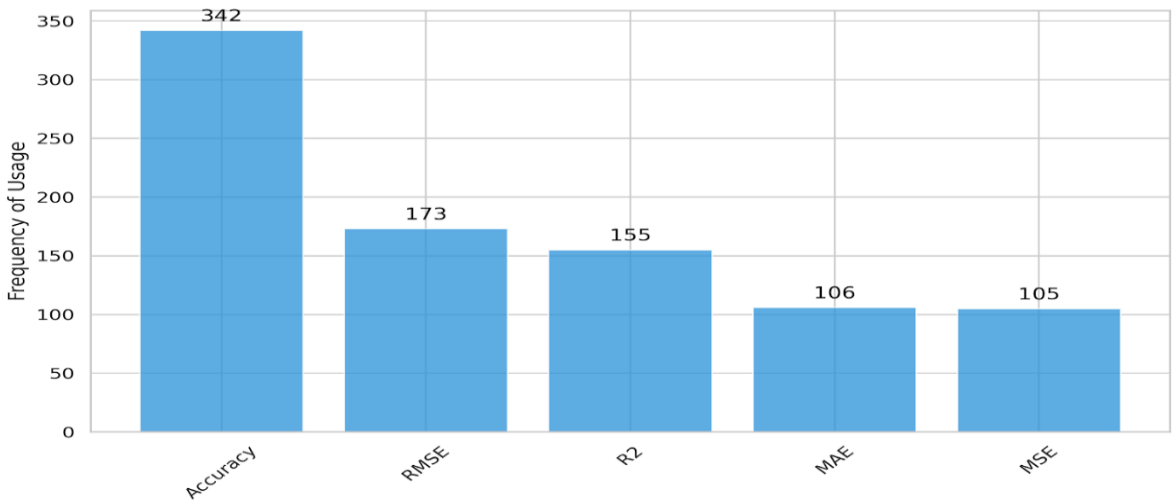

Figure 14. Frequency of evaluation metrics reported in AI based WQI and water quality prediction studies.

DISCUSSION

Current research status: The logistic growth curve provides insightful information about the stage of development within this area of research. The steep section of the curve places the domain beyond an embryonic state but not fully matured, a state aptly described by active formations of research topics and methods that also extend to collaboration structures.

The anticipated inflection point seems to be quite near, which implies a critical transition in the coming years. This transition will be from answering basic foundational questions toward more specific applied questions. The record is available for 2010-2025, during which increased remote sensing products and in situ monitoring become available open data initiatives and access to cloud-based computational infrastructures that support large-scale model development, environmental modeling using AI appears structurally mature as a discipline. A boom of publications after 2020 is evident when the global trend of accelerating adoption manifested itself also on the country level.

Studies have found a post-pandemic increase in the use of hybrid and ensemble ML methods for water-related prediction tasks, showing how improvements in data accessibility and computational capabilities have altered research paths in this field. The sentence is too factual to be changed much. Add sentences to that paragraph for me to reword with it. (Muñoz Alegría et al., 2025; Xia et al., 2024). The present bibliometric analysis takes a snapshot of the field at this significant stage and thus provides an excellent baseline from which incipient tendencies and configurations can be detected before they become consolidated. This is the interpretation supported by the dynamics of production to which further support is added in explicitly thematic and network-based analyses carried out within this study (for instance, thematic map and keyword co-occurrence structure).

The three-field framework strongly supports the fact that this area of research has developed systematically and not randomly, around a few major methodological themes, regionally concentrated expertise, and publication outlets. No explicit results are provided to determine whether advanced data-driven methods have become a national rather than an international topical specialization; however, such results can be implied from the strong association between certain countries and advanced data-driven methods.

At the same time, publications are also clustered within a limited number of journals. Knowledge consolidation appears to be happening through accepted disciplinary channels. From the perspective of bibliometric analysis, this would support an interpretation that the field is moving from a stage of explosive growth to one characterized by thematic crystallization in which incremental developments take place within increasingly clearly circumscribed methodological and institutional settings.

Access to top outlets and the concentration of publications in high-visibility journals (such as Water and Science of the Total Environment) further speaks for the institutionalisation of publication channels and reinforced targeted dissemination; this observation matches recent findings suggesting that while the average citation metric may sometimes be an overestimate in specific situations, environmental and water research is becoming more genuinely influential (Duarte et al., 2025). The geographical separation implied by the distinction between scope country and publication country together with the SCP/MCP split adds to a general tendency that internationally co-authored papers get higher visibility and more citations-perhaps different with outcomes depending on particular subfields and network connections (Vélez-Estévez et al., 2022).

The observed thematic progression corresponds to the quick expansion phase which has been recognized in production dynamics thus, methodological innovation has proven one of the main entry points for recent growth within this field. There exist traditional concepts on water quality together with new techniques and methods based on AI and interpretability, indicating a process of accretion whereby new methodologies are added rather than completely substituted.

This trend signifies a mature area of research in which traditional environmental assessment models are being added to more advanced computational methods for better predictions, explanations, and decision support. From the perspective of bibliometrics, explainable AI methods appear to be rising because researchers have explicitly started to recognize the importance of transparent and robust results when water quality research is applied in policies or management practices. This shift towards model-oriented topics can be observed throughout the whole study period by both changes within frameworks over time as well as thematic and co-occurrence analysis results that highlight their increasing importance.

The conceptual structure of the thematic map reflects a stage in the discipline at which core methodological paradigms are firmly set and supplementary, more specific approaches continue to develop around these central frameworks. It raises an important theme about monitoring and assessment based on ML that speaks volumes about how data-driven paradigms have become standard consolidated methodologies within the domain of water quality research. The study also found DL to be one of the baseline themes; hence, application-oriented development may still be experimental but is catching up fast.

Niche clustering around explainability and specific indices is indicative of a trend toward research on transparent models, robust models, and contextually refined models. The significance of this trend is best highlighted by the fact that scientific knowledge can be severely distorted due to limited interpretability of outcomes at a very aggregate level without any strong linkage to operational decision-making contexts. It means that research results fail to display the actual situation (Scheel et al., 2021). Also, the more productive high capacity models remain largely a black box adds to delays in widescale environmental application since lack of clarity leads to lack of trust, and consequently accountability and applicability over different settings (Reichstein et al., 2019). This is therefore evidence of an emerging mature research field that is increasingly oriented around methodological specialization and integration rather than wholly new thematic directions. Apart from methodological performance, there are hidden yet growing impacts on sustainability due to the ongoing “data-intensive revolution” in AI-based water quality research.

High-capacity DL architectures normally require large-scale computational infrastructures which in turn consume a lot of energy and add to carbon emissions (Strubell et al., 2020). Green AI is still largely perceived as a technical option, meanwhile becoming an important research area and governance issue for water management-a sustainability-relevant sector. The insufficient technical capacities, organizational barriers and lack of trust among various stakeholders and implementers translate greener modeling practices into operational decision contexts (Turan et al., 2025). Future advances of ML-based WQI prediction will reproducibility involve energy aware a compatible policy modeling pipeline besides an algorithmic improvement intellectual framework that the cocitation network analysis discloses.

ML-based water quality research has been developed knowledge in cumulative and accretive process rather than by sudden paradigmatic shifts. The applied investigations stay close to their foundational algorithmic contributions while more recent publication clusters show continued fine tuning, adjustment, and contextualization of methodologies to address domain specific issues. The tightly connected contemporary research fronts are indicative of increasing specialization and methodological fine-tuning in predictive modeling and decision-support systems. From the bibliometric point of view, such a configuration implies that the field has achieved an intellectual coherence while still spawning new ideas and innovations-placing it in a state where future dominance by incremental improvements in methods as well as applications oriented enhancements can be hypothesized. The collaboration framework exemplifies both the globalization and hierarchical organization of scientific output within the discipline. Research activities may be widely distributed on different continents, but a few countries play key roles as central coordination nodes which allow knowledge to flow and collaborations to spread widely.

The leading positions of the US and China highlight the importance of domestic research capacity, funding infrastructures and institutional frameworks in underpinning modes of international collaboration. At the same time, India’s emerging central bridging role points to the increasing influence that developing and transitional research systems exert on global networks. From a bibliometric standpoint, this structure suggests that future development of the field will be driven by multinational cooperation based around core scientific powers with an accelerating contribution from new peripheral areas being assimilated into the global research environment.

The results show that even though ANN have long been and still are the most popular method, there has recently been a technological shift to DL and Boosting methods such as XGBoost since 2019. Therefore, a methodological revolution can be described in the field not only by an increasing number of publications but also by improved predictive performance at the task level. A recent study provides evidence for this fact by showing DL models substantially outperformed traditional linear models in the estimation of two key parameters chlorophyll-a and turbidity (Anand et al., 2024). In addition, Recent studies on prediction of water quality clearly establish the fact that present day ML and DL techniques out perform the autoregressive integrated moving average ARIMA and conventional techniques. This further justifies a rapidly growing tendency for a data driven, nonlinear modeling framework (Chen et al., 2024). At the same time, present literature displays an incremental shift from fuzzy systems to DL and hybrid AI setups, in line with the general trend of integrating high-capability learning architectures for environmental modelling. An evolution highly related to the enlargement of Big Data analytics and application of high-frequency sensor networks within water resource management.

Meanwhile, increasing availability of water quality datasets in large volumes and diversities because of remotely sensed and in situ observations has made purely conventional modeling pipelines less appropriate. This acts as one more driving factor for the rapid ML and neural network based methods’ adoption to address spatiotemporal prediction tasks (Zhi et al., 2024). These models, ANN, SVR and RF are the three most popular models successfully applied not only for the estimation of water quality index but also for several other environmental purposes. These include phytoplankton density and indicators of eutrophication among others (Hafeez et al., 2019; Fang et al., 2025). This initiates computational scientific research based on a data-driven environmental science involving both computational methods and environmental concepts. The concepts comprise climate change with standard protocols on the assessment of water quality (Martinez et al., 2020; Zheng et al., 2025). A more complex nature of the data lies beyond a simple linear relationship, and conventional modeling becomes inadequate such that it calls for an architectural framework supporting higher levels of abstraction. Another important result is about emerging trends in "model-data fitting" phenomena. Analysis shows researchers deliberately choose DL techniques to model highly non-linear biochemical parameters (for example nitrogen, fecal coliforms) while choosing more interpretable models like RF for stable physical parameters (for example turbidity). The finding emphasized that the selection of a model has now become an informed strategic decision based on the characteristic features of a problem instead of being imposed as a random choice.

In addition, differences between China's AI-based national techno policies and the low-cost modeling strategies of the USA and India reflect a scholarly output of how techno policies can be set at a national level. This is also reflected in country-level differences in patterns of algorithmic adoption, as reported in results. Finally, the fact that more than 90% of the papers use R^2 and RMSE can also be interpreted as a sign that methodological maturity has been mostly achieved in this field. The absence-or rather limited presence-of classification metrics such as Accuracy, however, reveals an inner truth about WQI research: it is still greatly oriented toward developing regression-based prediction models instead of fully elaborated decision support systems. This paper provides ample material on which to base final remarks concerning the future direction of AI for water quality management.

Existing problems and future research: A perennial problem with water quality index based assessment is that different indices may classify the same water body, leaving unanswered the question as to which index is most appropriate and reliable (Simian et al., 2025). New classification systems and robust classifiers enhance class prediction in coastal systems but, scientific basis for classification is weakened by lack of consideration of temporal variation and restricts its use to the management purposes (Uddin et al., 2023). Most of the existing studies on coastal WQI using tree and ensemble methods leave a huge gap by not allowing for temporally resolved output, since non-temporally resolved results are much harder to integrate into practical monitoring and decision-making (Uddin et al., 2022). Data-driven reservoir WQI frameworks can decrease uncertainty and resolve well-documented aggregation issues, but their need for a large amount of data and site-specific information means that external validation across multiple reservoirs is a key area that needs further attention in the future (Zare et al., 2025).

Several of these methodological problems stem from the quality and scope of monitoring data. Limited station density, low sampling frequency, and restricted parameter scope, combined with inadequate real-time validation and the exclusion of certain crucial indicator groups, directly diminish the representativeness of WQI assessments (Das, 2025). In groundwater environments, budget limitations and insufficient sampling data exacerbate the long-standing issue of creating dependable irrigation-related assessments in conditions of limited data availability (Hussein et al., 2024). Working with datasets that are irregular and noisy makes prediction more complicated and also limits the interpretation of results when drivers of environmental factors and interactions among variables are not accurately depicted (Kouadri et al., 2021). Good performance under limited sampling and narrow predictor sets is not automatically transferable to other basins without specific testing (Sakaa et al., 2022). Generalization claims for biological oxygen demand (BOD) and dissolved oxygen (DO) prediction models developed from a single regional dataset can only be made after broader external validation (Al-Mukhtar et al., 2024).

Remote sensing clearly provides opportunities together with multisource data integration but also underlines the need for disciplined feature selection. Rather than always assuming that adding more satellite bands will improve results, selecting band combinations in which WQI is sensitive can be more consistent with both accurate and efficient (Nikoo et al., 2024). Remote sensing, meteorological and land use data show better results in prediction for urban and peri-urban lakes. However, the limitations appear more explicitly in fine-scale inference within highly complex environments such as modern cities where ground truth sampling is hugely imbalanced together with medium spatial resolution (Dawn et al., 2025). In addition, models based on small datasets do not represent actual real-world boundary conditions. The practical applicability of such models becomes totally unrealistic when important physical processes-for example, dilution in mixing zones-is not incorporated or accounted for (Siddiq et al., 2025).

On the modeling side, methods always perform well when aimed at capturing a spatio-temporal dependence but this has become the new emerging bottleneck as scales grow. The hybrids between spatio-temporal can achieve high accuracies but computational load and scalability become major challenges when dealing with high dimensional or large scale datasets (Wang et al., 2025). Ensemble approaches to multi-depth estimation can improve the prediction capability. However, a large number of sub-models have to be trained and sensitive hyperparameters need fine-tuning which becomes an immediate obstacle in achieving operational deployment quickly (Zare et al., 2024). Combining wavelet-based CNN-GRU architectures can provide better long horizon forecasting for nonstationary and nonlinear time series. Yet it remains to be seen how well such highly complex models, developed/trained on a single catchment/system, generalize/adapt when applied/deployed over different systems/catchments (Khosravi et al., 2025). In other associated hybrid DL contexts, the total absence of explicit consideration for different types of uncertainty-aleatoric, epistemic and model leaves trust unanswered even if headline accuracy is high (Zamani et al., 2024). Furthermore, single variable WQI time series models may offer limited ability to explain future management scenarios unless external factors like urbanization or industrial pressure are included in the modelling framework (Niknam et al., 2025).

The choice of hyperparameter search strategy in applied work is a methodological decision that significantly influences both predictive accuracy and computational cost, with optimization playing a crucial role (Uddin et al., 2025). Ensembles and tuning can enhance the accuracy for imbalanced WQI classification issues but dimensionality reduction, online learning, and real-time decision support integration continue to be significant areas of future research (Lawal et al., 2024). Interpretability limitations are a major factor hindering the translation of research prototypes into real-world use. A need exists for complementary and multi-method interpretability as some explanation techniques can be sensitive to model architecture and may introduce biased interpretations, rather than relying on a single tool (Majnooni et al., 2024). SHAP style analyses can enhance decision support by clarifying key determinants, however, larger datasets and systematic comparisons with other explainability techniques are still necessary to stabilise explanations and increase generalisability (Aldrees et al., 2024).

Across these studies, the most practical future plan comes together in expanding and balancing monitoring data, improving the combination of data from different locations and times, clearly defining uncertainty, and providing interpretable models that can be integrated with confidence into real-time water quality management systems.

LIMITATIONS

Despite the use of a multi-database approach, several limitations need to be recognised. Integrating Web of Science and Scopus into the corpus improves its coverage compared to single-source analyses, but it may still neglect studies found only in regional databases or published in non-English languages, thereby restricting the visibility of localised research findings. Bibliometric indicators primarily focus on the publication and citation patterns rather than the controlled predictive performance of individual models. The content-mining step identifies dominant algorithm families, parameters, and validation metrics, but it does not re-implement or benchmark models under controlled conditions. Furthermore, the significant increase in advanced AI techniques after 2019 may lead to a recency bias, as newer methods have had less time to amass citations and prove their long-term reliability. Country level patterns and policy relevant interpretations are examined as indicative associations since this study does not directly measure policy investments or operational adoption in real world water management settings.

CONCLUSION

This paper shall review, in a most systematic way, how the methods applied to AI for WQI have evolved by academic literature between 2010 and 2025. It found that the digitalization of water resources management moves from an earlier ANN-based method to powerful DL and boosting-based methods capable of handling large datasets. A “Model Data Fit” oriented selection appears to be emerging preferential toward physically interpretable parameter models and more complex biochemical variables using DL architectures, as explained through further results herein. The study also confirms the large influence of national technological infrastructure policy on academic research outcomes. Enhance explainability to make decision-makers confident in high-performing black box models through a rigorous XAI practice is recommended, together with the hybrid modeling pipeline standardization that integrates preprocessing and optimization to overcome the limitations of standalone learners. Further research should move from mostly static datasets towards developing operational real-time prediction systems integrated with IoT-based sensor networks.

Future research should take up a wider methodological issue in the harmonization of standards for the formulation and classification of Water Quality Index (WQI), to reduce inconsistencies, for example, between different hydrological sites. For strong generalization claims, an evaluation beyond a single case study is needed that applies independent verification across multiple basins under different climates concomitantly restructuring architectures for scalable spatio-temporal learning integrating remote sensing-meteorological-land use data-integration addressing monitoring gaps as yet unaddressed. The next enhancement AI brings into this domain shall be not only at some pinpoint level accuracy but rather Decision Support Systems explicit quantification harnessed towards transforming multifaceted output into executable management strategies through informed early warning even when confronted by incomplete noisy data.

ACKNOWLEDGEMENTS

Author's Contributions: The author contributed to the entire work on his own. Conceptualisation, Analysis; Writing-review and editing.

Conflict Of Interest: The authors declared that there are no conflicts of interest regarding the publication of this article.

Data availability: The author confirms that data supporting the findings of this study are available within the article.

Funding: None.

Ethical Approval: None.

REFERENCES

Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., ..., & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, ***578***, 124084. https://doi.org/10.1016/j.jhydrol.2019.124084

Akkan, T., Mutlu, T., & Baş, E. (2022). Forecasting sea surface temperature with feed-forward artificial networks in combating the global climate change: The sample of Rize, Türkiye. *Ege Journal of Fisheries and Aquatic Sciences,* **39**(4), 329-336. https://doi.org/10.12714/egejfas.39.4.06

Aldrees, A., Khan, M., Taha, A.T.B., & Ali, M. (2024). Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches. *Journal of Water Process Engineering*, ***58***, 104789. https://doi.org/10.1016/j.jwpe.2024.104789

Al-Mukhtar, M., Srivastava, A., Khadke, L., Al-Musawi, T., & Elbeltagi, A. (2024). Prediction of irrigation water quality indices using random committee, discretization regression, reptree, and additive regression. *Water Resources Management*, ***38***(1), 343-368. https://doi.org/10.1007/s11269-023-03674-y

Alshater, M.M., Hassan, M.K., Khan, A., & Saba, I. (2021). Influential and intellectual structure of Islamic finance: a bibliometric review. *International Journal of Islamic and Middle Eastern Finance and Management*, ***14***(2), 339-365. https://doi.org/10.1108/IMEFM-08-2020-0419

Alver, D.O., Işık, H., Palabıyık, S., Akkan, B.E., & Akkan, T. (2025). pH acidification in the Red Sea: A machine learning-based validation study. *Journal of Sea Research*, 102613. https://doi.org/10.1016/j.seares.2025.102613

Anand, V., Oinam, B., & Wieprecht, S. (2024). Machine learning approach for water quality predictions based on multispectral satellite imageries. *Ecological Informatics*, ***84***, 102868. https://doi.org/10.1016/j.ecoinf.2024.102868

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, ***11***(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Banda, T.D., & Kumarasamy, M. (2024). Artificial neural network (ANN)-based water quality index (WQI) for assessing spatiotemporal trends in surface water quality-A case study of South African river basins. *Water*, ***16***(11), 1485. https://doi.org/https://doi.org/10.3390/w16111485

Barzegar, R., Aalami, M.T., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, ***34***(2), 415-433**.** https://doi.org/10.1007/s00477-020-01776-2

Breiman, L. (2001). Random forests. *Machine learning*, ***45***(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chaudhary, S., Duraiyarasan, A., Gupta, D., &Mishra, V. K. (2024). Simulating and predicting surface water quality for drinking and bathing purposes through combined approach of PCA, entropy-based WQI, and stochastic models. *Stochastic Environmental Research and Risk Assessment,* ***38***(12), 5001-5017. https://doi.org/10.1007/s00477-024-02847-4

Chaudhary, S., Siwach, C., Anand, S., & Mishra, V.K. (2025). A bibliometric analysis of statistical methods used for river water quality assessment & monitoring during the last three decades (1992–2023). *Water and Environment Journal.* https://doi.org/10.1111/wej.12984

Chen, J., Li, Y., Zhang, H., & Wang, X. (2024). A Review of Machine Learning and Deep Learning Applications to Water Quality Prediction. *Applied Sciences,* ***14***(19), 8755. https://doi.org/10.3390/app14198755

Das, A. (2025). Prediction of Urban Surface Water Quality Scenarios Using Water Quality Index (WQI), Multivariate Techniques, and Machine Learning (ML) Models in Water Resources, in Baitarani River Basin, Odisha: Potential Benefits and Associated Challenges. *Earth Systems and Environment*, 1-37. https://doi.org/10.1007/s41748-025-00623-0

Dash, S., & Kalamdhad, A.S. (2021). Science mapping approach to critical reviewing of published literature on water quality indexing. *Ecological Indicators*, ***128***, 107862. https://doi.org/10.1016/j.ecolind.2021.107862

Dawn, A., Hinge, G., Kumar, A., Nikoo, M.R., & Hamouda, M.A. (2025). Assessment of water quality in urban lakes using multi-source data and modeling techniques. *Sustainability*, ***17***(16), 7258. https://doi.org/10.3390/su17167258

Duarte, I., Nunes, S., Pereira, L., & Silva, C. (2025). Clean water production–A content and bibliometric review of two decades of research on green membranes for desalination. *Heliyon, **11***(9), e28919. https://doi.org/10.1016/j.heliyon.2025.e28919

Fang, Y., Huang, R., Zhang, Y., Zhang, J., Xi, W., & Shi, X. (2025). Utilizing machine learning models to grasp water quality dynamic changes in lake eutrophication through phytoplankton parameters. *Frontiers of Environmental Science & Engineering*, ***19***(2), 14. https://doi.org/10.1007/s11783-025-1934-6

Gacu, J.G., Monjardin, C.E.F., Mangulabnan, R.G.T., Pugat, G.C.E., & Solmerin, J.G. (2025). Artificial Intelligence (AI) in Surface Water Management: A Comprehensive Review of Methods, Applications, and Challenges. *Water, **17***(11), 1707. https://doi.org/10.3390/w17111707

Gao, J., Zhu, S., Li, D., Jiang, H., Deng, G., Wen, Y., ..., & Cao, Y. (2023). Bibliometric analysis of climate change and water quality. *Hydrobiologia*, ***850***(16), 3441-3459. https://doi.org/10.1007/s10750-023-05270-y

Hafeez, S., Wong, M.S., Ho, H.C., Nazeer, M., Nichol, J., Abbas, S., ..., & Pun, L. (2019). Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: A case study of Hong Kong. *Remote sensing*, ***11***(6), 617. https://doi.org/10.3390/rs11060617

Haider, S., Rashid, M., Tariq, M.A.U.R., & Nadeem, A. (2024). The role of artificial intelligence (AI) and Chatgpt in water resources, including its potential benefits and associated challenges. *Discover Water*, ***4***(1), 113. https://doi.org/10.1007/s43832-024-00173-y

Han, Z., Zhang, S., & He, L. (2025). Predicting and investigating water quality index by robust machine learning methods. *Journal of Environmental Management*, *381*, 125156. https://doi.org/10.1016/j.jenvman.2025.125156

Heintzman, L.J., Ghaffari, Z., Awawdeh, A.R., Barrett, D.E., Yarbrough, L.D., Easson, G., ..., & Yasarer, H.I. (2024). Assessing Differences in Groundwater Hydrology Dynamics Between In Situ Measurements and GRACE-Derived Estimates via Machine Learning: A Test Case of Consequences for Agroecological Relationships Within the Yazoo–Mississippi Delta (USA). Hydrology, ***11***(11), 186. https://doi.org/10.3390/hydrology11110186

Hridoy, M.A.A.M., Shawkat, A.I., Bordin, C., Acharjee, M.R., Masood, A., Baki, A.O., & Al Mamun, M.A. (2025). Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management. *Science of the Total Environment*, ***1008***, 180965. https://doi.org/10.1016/j.scitotenv.2025.180965

Hussein, E.E., Derdour, A., Zerouali, B., Almaliki, A., Wong, Y.J., Ballesta-de los Santos, M., ..., & Elbeltagi, A. (2024). Groundwater quality assessment and irrigation water quality index prediction using machine learning algorithms. *Water*, ***16***(2), 264. https://doi.org/10.3390/w16020264

Iqbal, U., Riaz, M.Z.B., Barthelemy, J., Perez, P., & Idrees, M.B. (2023). The last two decades of computer vision technologies in water resource management: A bibliometric analysis. *Water and Environment Journal, **37***(3), 373-389. https://doi.org/10.1111/wej.12845

Işık, H., & Akkan, T. (2025). Water quality assessment with artificial neural network models: Performance comparison between SMN, MLP and PS-ANN methodologies. *Arabian Journal for Science and Engineering, **50***(1), 369-387. https://doi.org/10.1007/s13369-024-08402-w

Işık, H., & Palabıyık, S. (2025). Evde Bakım Hizmetlerinde Yapay Zekanın Rolü: Bibliyometrik Bir Analiz. *Doğu Karadeniz Sağlık Bilimleri Dergisi*, ***4***(2), 123-138. https://doi.org/10.59312/ebshealth.1674263

Işık, H., Baş, E., Egrioglu, E., & Akkan, T. (2024). A new single multiplicative neuron model artificial neural network based on black hole optimization algorithm: Forecasting the amounts of clean water given to metropolis. *Stochastic Environmental Research and Risk Assessment, **38***(11), 4259-4274. https://doi.org/10.1007/s00477-023-02552-3

Işık, H., Karahanlı, E., & Palabıyık, S. (2025). Implementation of a machine learning-based predictive model for assessing pH variability in coastal marine waters. *MEMBA Su Bilimleri Dergisi, **11***(3), 335-345. https://dergipark.org.tr/tr/pub/memba/issue/94927/1743888

Işık, H. (2025). Global Trends in Condition Factor and Length–Weight Relationships in Fisheries Science (2000-2025): A Bibliometric Analysis. *Journal of Anatolian Environmental and Animal Sciences*, ***10***(6), 889-901. https://doi.org/10.35229/jaes.1766015

Kan, G., Liang, K., Yu, H., Sun, B., Ding, L., Li, J., & Shen, C. (2020). Hybrid machine learning hydrological model for flood forecast purpose. *Open Geosciences*, ***12***(1), 813-820. https://doi.org/10.1515/geo-2020-0166

Khosravi, S., Nikoo, M. R., Zamani, M. G., Barzegar, R., Najafipour, P., & Gandomi, A. H. (2025). Enhancing long-term water quality forecasting with a hybrid deep-learning approach integrating MODWT, CNN, and GRU. *Hydrological Sciences Journal*, (just-accepted). https://doi.org/10.1080/02626667.2025.2581264

Kouadri, S., Elbeltagi, A., Islam, A.R.M.T., & Kateb, S. (2021). Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (*Algerian southeast*). *Applied Water Science*, ***11***(12), 190. https://doi.org/10.1007/s13201-021-01528-9

Lau, B.P.L., Marakkalage, S.H., Zhou, Y., Hassan, N.U., Yuen, C., Zhang, M., & Tan, U. X. (2019).

A survey of data fusion in smart city applications. *Information Fusion*, ***52***, 357-374. https://doi.org/10.1016/j.inffus.2019.05.004

Lawal, Z.K., Aldrees, A., Yassin, H., Dan'azumi, S., Naganna, S.R., Abba, S.I., & Sammen, S.S. (2024). Optimized Ensemble Methods for Classifying Imbalanced Water Quality Index Data. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3502361

Li, J., Lu, W., & Luo, J. (2021). Groundwater contamination sources identification based on the Long-Short Term Memory network. *Journal of Hydrology*, *601*, 126670. https://doi.org/10.1016/j.jhydrol.2021.126670

Li, X., Su, J., Wang, H., Boczkaj, G., Mahlknecht, J., Singh, S.V., & Wang, C. (2024). Bibliometric analysis of artificial intelligence in wastewater treatment: Current status, research progress, and future prospects. *Journal of Environmental Chemical Engineering*, ***12***(4), 113152. https://doi.org/10.3390/w17202994

Li, Y., Zhang, H., Yuan, W., Qi, H., Ding, B., & Zhang, B. (2025). Evolutionary Trends in Machine Learning for Water Quality Prediction: A Visual Analysis Using CiteSpace and RStudio Bibliometrix. *JAWRA Journal of the American Water Resources Association*, ***61***(6), e70061. https://doi.org/10.1111/1752-1688.70061

Lu, Q., Si, W., Wei, L., Li, Z., Xia, Z., Ye, S., & Xia, Y. (2021). Retrieval of water quality from UAV-borne hyperspectral imagery: A comparative study of machine learning algorithms. *Remote Sensing*, ***13***(19), 3928. https://doi.org/10.3390/rs13193928

Majnooni, S., Fooladi, M., Nikoo, M. R., Al-Rawas, G., Haghighi, A. T., Nazari, R., ... & Gandomi, A. H. (2024). Smarter water quality monitoring in reservoirs using interpretable deep learning models and feature importance analysis. *Journal of Water Process Engineering*, ***60***, 105187. https://doi.org/10.1016/j.jwpe.2024.105187

Makumbura, R.K., Mampitiya, L., Rathnayake, N., Meddage, D.P.P., Henna, S., Dang, T.L., ..., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results in Engineering*, ***23***, 102831. https://doi.org/10.1016/j.rineng.2024.102831

Martinez, E., Brini, A., Gorgues, T., Drumetz, L., Roussillon, J., Tandeo, P., ..., & Fablet, R. (2020). Neural network approaches to reconstruct phytoplankton time-series in the global ocean. *Remote Sensing*, ***12***(24), 4156. https://doi.org/10.3390/rs12244156

Md Shahri, N.H.N., & Mohd Ali, Z. (2024). Bibliometric perspectives in mapping the progress of river water quality index studies. *Water Quality Research Journal,m* ***59***(4), 232-267. https://doi.org/10.2166/wqrj.2024.022

Muñoz-Alegría, J.A., Núñez, J., Oyarzún, R., Chávez, C.A., Arumí, J.L., & Rodríguez-López, L. (2025). A Bibliometric-Systematic Literature Review (B-SLR) of Machine Learning-Based Water Quality Prediction: Trends, Gaps, and Future Directions. *Water*, ***17***(20), 2994. https://doi.org/10.3390/w17202994

Mutlu, C., & Akkan, B.E. (2025). An up-to-date approach using machine learning methods in fish condition factor estimation. *Turkish Journal of Fisheries and Aquatic Sciences,* ***25***(7). https://doi.org/10.4194/TRJFAS25498

Niknam, A. R. R., Sabaghzadeh, M., Barzkar, A., & Shishebori, D. (2025). Comparing ARIMA and various deep learning models for long-term water quality index forecasting in Dez River, Iran. *Environmental Science and Pollution Research*, ***32***(16), 10206-10222. https://doi.org/10.1007/s11356-024-32228-x

Nikoo, M.R., Zamani, M.G., Zadeh, M.M., Al-Rawas, G., Al-Wardy, M., & Gandomi, A.H. (2024). Mapping reservoir water quality from Sentinel-2 satellite data based on a new approach of weighted averaging: Application of Bayesian maximum entropy. *Scientific Reports*, ***14***(1), 16438. https://doi.org/10.1038/s41598-024-66699-2

Nong, X., He, Y., Chen, L., & Wei, J. (2025). Machine learning-based evolution of water quality prediction model: An integrated robust framework for comparative application on periodic return and jitter data. *Environmental Pollution,* ***369***, 125834. https://doi.org/10.1016/j.envpol.2025.125834

Odabaş Alver, D. (2025). Global and Turkish Research Trends in PM_{10} Prediction Using Machine Learning: A Bibliometric Perspective. *Journal of Anatolian Environmental and Animal Sciences,* ***10***(6), 954-971. https://doi.org/10.35229/jaes.1766267

Olawade, D.B., Wada, O.Z., Ige, A.O., Egbewole, B. I., Olojo, A., & Oladapo, B.I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. *Hygiene and Environmental Health Advances*, ***12***, 100114. https://doi.org/10.1016/j.heha.2024.100114

Rana, R., Kalia, A., Boora, A., Alfaisal, F.M., Alharbi, R.S., Berwal, P., & Qamar, O. (2023). Artificial intelligence for surface water quality evaluation, monitoring and assessment. *Water*, ***15***(22), 3919. https://doi.org/10.3390/w15223919

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature,* 566(7743), 195-204. https://doi.org/10.1038/s41586-019-0912-1

Rodríguez-Soler, R., Uribe-Toril, J., & Valenciano, J. D. P. (2020). Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. *Land Use*

Policy, ***97***, 104787. https://doi.org/10.1016/j.landusepol.2020.104787

Rozos, E., Dimitriadis, P., & Bellos, V. (2022). Machine learning in assessing the performance of hydrological models. *Hydrology* ***9***(1), 1-17. https://doi.org/10.3390/hydrology9010005

Sakaa, B., Elbeltagi, A., Boudibi, S., Chaffaï, H., Islam, A.R.M.T., Kulimushi, L.C., ... & Wong, Y.J. (2022). Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. *Environmental Science and Pollution Research*, ***29***(32), 48491-48508. https://doi.org/10.1007/s11356-022-18644-x

Sangwan, V., & Bhardwaj, R. (2024). Application of machine learning model for assessing water quality index. In Integrated Management of Water Resources in India: A Computational Approach: Optimizing for Sustainability and Planning (pp. 291-308). Cham: *Springer Nature Switzerland.* https://doi.org/10.1007/978-3-031-62079-9_16

Scheel, A.M., Schijen, M., & Lakens, D. (2021). An excess of positive results: Comparing the standard Psychology literature with registered reports. *Nature Human Behaviour,* ***5***, 1544-1553. https://doi.org/10.1038/s41562-021-01121-7

Serter, M., & Gumusburun Ayalp, G. (2024). A holistic analysis on risks of post-disaster reconstruction using RStudio Bibliometrix. *Sustainability*, ***16***(21), 9463. https://doi.org/10.3390/su16219463

Shalby, A., Elshemy, M., & Zeidan, B.A. (2020). Assessment of climate change impacts on water quality parameters of Lake Burullus, Egypt. *Environmental Science and Pollution Research*, ***27***(26), 32157-32178. https://doi.org/10.1007/s11356-019-06105-x

Sharma, P., Singh, S., & Sharma, S. D. (2022). Artificial neural network approach for hidrologca river flow time series forecasting. *Agricultural Research,* ***11***(3), 465-476. https://doi.org/10.1007/s40003-021-00585-5

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental modelling & software,***22**(4), 464-475. https://doi.org/10.1016/j.envsoft.2006.02.001

Siddiq, B., Javed, M.F., & Aldrees, A. (2025). Machine learning-driven surface water quality prediction: an intuitive GUI solution for forecasting TDS and DO levels. *Water Quality Research Journal*, ***60***(4), 514-546. https://doi.org/10.2166/wqrj.2025.005

Simian, D., Şerban, M. E., & Bărbulescu, A. (2025). Machine learning-based multifaceted analysis framework for comparing and selecting water quality indices. *Water Resources Management*, ***39***(2), 847-863. https://doi.org/10.1007/s11269-024-03993-8

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for deep learning in NLP. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645-3650). https://doi.org/10.18653/v1/P19-1355

Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., & John, A. (2022). A review of earth artificial intelligence. *Computers & Geosciences*, ***159***, 105034. https://doi.org/10.1016/j.cageo.2022.105034

Tekin, A. (2025). Global Research Trends and Thematic Development on the Biological Protective Effects of Boron: A Comprehensive Bibliometric Analysis (1990-2025). *Journal of Anatolian Environmental and Animal Sciences,* ***10***(6), 844-860. https://doi.org/10.35229/jaes.1773399

Turan, T., Turan, G., & Kırbaş, İ. (2025). Yapay zekanın çevresel ayak izi: Enerji, su ve elektronik atık üzerine çok yönlü bir analiz. *Uluslararası Mühendislik Tasarım Ve Teknoloji Dergisi,* *7*(2), 78-89. https://doi.org/10.70669/ijedt.1612846

Uddin, M.G., Bamal, A., Diganta, M.T.M., Sajib, A.M., Rahman, A., Abioui, M., & Olbert, A.I. (2025). The role of optimizers in developing data-driven model for predicting lake water quality incorporating advanced water quality model. *Alexandria Engineering Journal*, ***122***, 411-435. https://doi.org/10.1016/j.aej.2025.03.022

Uddin, M. G., Nash, S., Diganta, M. T. M., Rahman, A., & Olbert, A. I. (2022). Robust machine learning algorithms for predicting coastal water quality index. *Journal of Environmental Management*, ***321***, 115923. https://doi.org/10.1016/j.jenvman.2022.115923

Uddin, M.G., Nash, S., Rahman, A., & Olbert, A.I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection*, ***169***, 808-828. https://doi.org/10.1016/j.psep.2022.11.073

Vélez-Estévez, A., Pérez, I.J., García-Sánchez, P., & Moreno-García, M.N. (2022). New trends in bibliometric APIs: A comparative analysis. *Scientometrics,* ***127***(12), 7517-7555. https://doi.org/10.1007/s11192-022-04384-3

Wang, D., Zhang, C., Li, A., Guo, Y., Zhang, H., & Tan, C. (2025). Spatio-temporal analysis and prediction for raw water quality of drinking water source by improved RNN algorithm. *Journal of Water Process Engineering*, ***71***, 107164. https://doi.org/10.1016/j.jwpe.2025.107164

Xia, L., Hao, X., & Zhou, Y. (2024). Machine learning-based global trends and the development prospects of wastewater treatment: a bibliometric analysis. *Journal of Environmental Chemical Engineering*, ***12***(3), 112732. https://doi.org/10.1016/j.jece.2024.112732

Xu, J., Mo, Y., Zhu, S., Wu, J., Jin, G., Wang, Y. G., ..., & Li, L. (2024). Assessing and predicting water quality index with key water parameters by machine learning models in coastal cities, China. *Heliyon*, ***10***(13). https://doi.org/10.3390/w16111485

Yucesoy, E., Egrioglu, E., & Baş, E. (2025). An intuitionistic fuzzy Gaussian process regression function approach for forecasting problem. *Computational Economics,* 1-18. https://doi.org/10.1007/s10614-025-10373-7

Zamani, M.G., Nikoo, M.R., Al-Rawas, G., Nazari, R., Rastad, D., & Gandomi, A.H. (2024). Hybrid WT–CNN–GRU-based model for the estimation of reservoir water quality variables considering spatio-temporal features. *Journal of Environmental Management*, ***358***, 120756. https://doi.org/10.1016/j.jenvman.2024.120756

Zare, M.S., Nikoo, M.R., Al-Rawas, G., Al-Wardy, M., Nazari, R., Karimi, M., & Gandomi, A.H. (2025). Integrating machine learning and data envelopment analysis for reliable reservoir water quality index assessment considering uncertainty. *Hydrological Sciences Journal*, ***70***(7), 1109-1124. https://doi.org/10.1080/02626667.2025.2473475

Zare, M.S., Nikoo, M.R., Al-Rawas, G., Nazari, R., Al-Wardy, M., Etri, T., & Gandomi, A.H. (2024). Integrated ensemble learning approach for multi-depth water quality estimation in reservoir environments. *Journal of Water Process Engineering*, ***66***, 105840. https://doi.org/10.1016/j.jwpe.2024.105840

Zhan, C., Dai, Z., Yin, S., Carroll, K.C., & Soltanian, M.R. (2024). Conceptualizing future groundwater models through a ternary framework of multisource data, human expertise, and machine intelligence. *Water Research*, ***257***, 121679. https://doi.org/10.1016/j.watres.2024.121679

Zheng, Y., Zhang, X., Zhou, Y., Zhang, Y., Zhang, T., & Farmani, R. (2025). Deep representation learning enables cross-basin water quality prediction under data-scarce conditions. *npj Clean Water*, ***8***(1), 33. https://doi.org/10.1038/s41545-025-00466-2

Zhi, W., Appling, A.P., Golden, H.E., Podgorski, J., & Li, L. (2024). Deep learning for water quality. *Nature Water*, ***2***(3), 228-241. https://doi.org/10.5281/zenodo.7056647

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health,* ***1***(2), 107-116. https://doi.org/10.1016/j.eehl.2022.06.001