
A Bibliometric-Systematic Literature Review (B-SLR) of Machine Learning-Based Water Quality Prediction: Current State and Projections

[J. Adriana Muñoz-Alegría](#)*, [Jorge Núñez](#), [Ricardo Oyarzún](#), [Cristian Alfredo Chávez](#), José Luis Arumí, [Lien Rodríguez-López](#)

Posted Date: 9 September 2025

doi: 10.20944/preprints202509.0779.v1

Keywords: water quality; machine learning; explainable artificial intelligence; XAI; bibliometric analysis; B-SLR; systematic literature review



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

A Bibliometric-Systematic Literature Review (B-SLR) of Machine Learning-Based Water Quality Prediction: Current State and Projections

J. Adriana Muñoz-Alegría ^{1,*}, Jorge Núñez ², Ricardo Oyarzún ^{2,3,4}, Cristian Alfredo Chávez ⁵, José Luis Arumí ^{3,6} and Lien Rodríguez-López ⁷

¹ Doctorate Program in Energy, Water and Environment, University of La Serena, La Serena 1700000

² Department of Mining Engineering, University of La Serena, La Serena 1700000, Chile

³ Water Research Center for Agriculture and Mining (CRHIAM), ANID FONDAP Center, Universidad de Concepción, Concepción 4070411, Chile

⁴ Center for Advanced Studies in Arid Zones (CEAZA), La Serena 1700000, Chile

⁵ Department of Mechanical Engineering, University of La Serena 1700000, Chile

⁶ Department of Water Resources, Universidad de Concepción, Chillán 3812120, Chile

⁷ Faculty of Engineering, Universidad San Sebastián, Concepción 4030000, Chile

* Correspondence: jeimmy.munoz@userena.cl

Abstract

Prediction of freshwater quality, both surface and groundwater, is a key component for the sustainable management of water resources. This study analyzes the state of the art on the application of artificial intelligence (AI), machine learning (ML), and deep learning (DL) models in water quality prediction, using a structured framework for conducting systematic literature reviews known as Bibliometric-Systematic Literature Review (B-SLR). A total of 1822 articles were collected from the Scopus database (2000–2024), of which 274 were finally selected through an automated filtering process and manual validation. The analysis allowed us to identify the most used algorithms (e.g., Random Forest, XGB, LSTM, CNN), the most studied water bodies (e.g., rivers, aquifers), and the most used key quality indicators (e.g., WQI, EWQI, DO, nitrates). In addition, a transition towards the use of hybrid and explainable models was identified, with increasing application of interpretability techniques. The application of the B-SLR approach allowed for obtaining a more robust, replicable, and in-depth view of the field, facilitating the identification of thematic patterns, gaps, and research opportunities.

Keywords: water quality; machine learning; explainable artificial intelligence; XAI; bibliometric analysis; B-SLR; systematic literature review

1. Introduction

Surface and groundwater pollution is a problem of global concern, as it can negatively affect freshwater availability, a key aspect for the conservation of ecosystems and the health of the population. This is aggravated by the strong pressure exerted by a growing demand for fresh water to sustain population growth and associated economic development [1–3]. Thus, the availability of good-quality freshwater sources has a direct impact on the ability to ensure water security, especially when nearly 2.2 billion people lack access to safe drinking water and another 4.2 billion do not have safe sanitation systems [4].

Water quality is determined by different factors, such as weather conditions [5], seasonal changes in hydrological regime [6], local geological conditions [7], human activities or land use, among others. Water quality has traditionally been assessed with *in situ* probes or by sampling and subsequent analysis of biological and physical-chemical parameters in the laboratory. Finally, the

water quality has been increasingly forecasted with physically-based simulation models, such as QUAL2K [8–10] and WASP [11,12], as well as statistical models such as multivariate analysis [13,14], and linear regression [15,16].

In this context, the application in recent decades of artificial intelligence (AI) [17–20], especially machine learning (ML) models [21–24] and deep learning (DL) [25–31], have contributed significantly to the development of water quality prediction techniques for environmental management purposes, strengthening the path towards the fulfillment of the water-related Sustainable Development Goals (SDGs) [32]. These prediction models can handle large amounts of data and adapt to complex relationships. In fact, the use of ML for water quality studies is highlighted as an example of the United Nations Global Acceleration Framework, for the scope of the SDG-6 [33,34].

An important aspect in the application of AI/ML/DL for water quality prediction is that its use allows anticipating variations in physical-chemical and biological parameters of water, using either historical data, remote sensing, or hydrological models, therefore optimizing environmental management and evidence-based decision-making. In this regard, two main approaches are identified in the applications used for water quality prediction: a) those that seek the prediction of individual parameters known as water quality parameters (WQP) [10,35–38] and (b) those whose purpose is to integrate these parameters into a water quality indicator, traditionally known as Water Quality Index (WQI) [39–43]. Regarding the former, we can mention, only as a given example, the prediction of total phosphorus concentrations in Taihu Lake, China, using the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm and the Long Short-Term Memory (LSTM) algorithm together with the use of Shapley Additive Explanations (SHAP) as an interpretability method [44]. Another example is the study [45], which applies the Random Forest (RF) method for the prediction of 14 physicochemical parameters using historical data from the Loa River, in the xeric region of Chile. With respect to the second approach, an example is the study [46], where a set of ML models is used to estimate the WQI in the Bug River, Ukraine. A high predictive performance was also found in the study [47] where different ML models (i.e., Extreme Gradient Boosting (XGB), Support Vector Regression (SVR), Artificial Neural Network (ANN), Random Forest Regression (RFR)) were evaluated to predict the Entropy Weighted Water Quality Index (EWQI) and estimate groundwater quality in Kerala, India.

The aforementioned studies are examples of how AI/ML/DL methods have been consolidated as water quality prediction techniques, given their robustness and efficiency to extract patterns of complex systems, as water systems, making these techniques an interesting alternative to conventional water quality modeling techniques [48–50]. Indeed, systematic reviews of literature at a global level show the high interest of researchers in the use of AI/ML/DL for water quality prediction. For example, the authors in [51] reviewed 876 articles within the period 2015–2022, showing that the United States, England, Iran, India, and China have emerged as major contributors to the field of water quality prediction with ML and DL. Likewise, the authors in [52] reviewed 249 articles on water quality using Internet of Things (IoT) models and machine learning. In another study [53], 253 articles were reviewed, finding that DL optimizes the processing of large volumes of data through parallel computing, facilitating the effective prediction of water quality, although its success depends on the quality of the data used. Finally, in [54], the literature was reviewed in terms of groundwater quality prediction studies using AI/ML/DL, finding that the ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) techniques have proven to be efficient and accurate tools for such purposes.

Despite the intrinsic value of the aforementioned literature review studies, it is important to highlight a key conceptual distinction between the two main review approaches. One is known as Bibliometric Analysis, that is, a quantitative method that allows the identification of emerging trends, thematic networks, and research gaps [55–57]. Another corresponds to Systematic Review of the literature, a rigorous and structured method used to collect and evaluate all available evidence on a research question [58–60]. Although the two approaches complement each other, and together they are a promising methodology to elaborate robust reviews of the scientific literature, to date no work,

to the best of our knowledge, has addressed the study of the state of the art on the application of AI/ML/DL for water quality prediction through the integration of both approaches. Thus, such a dual-approach review corresponds to the main objective of this work, as this contribution proposes to analyze the state of the art on the application of AI/ML/DL for water quality prediction in the 2000-2024 time frame, using the Bibliometric-Systematic approach Literature Review (B-SLR) [61]. It is hoped that the results of this work can guide researchers interested in the study, assessment and prediction of water quality using AI/ML/DL techniques.

2. Materials and Methods

We employed a structured framework for conducting systematic literature reviews, known as the Bibliometric-Systematic Literature Review (B-SLR), following the guidelines outlined in [61]. This approach was used to assess the current status, emerging trends, and research gaps in the application of AI/ML/DL techniques to freshwater (surface and groundwater) quality studies. The methodology B-SLR combines the bibliometric analysis (BA) [57,62,63] with the systematic literature review (SLR) [64–66]. Thus, the B-SLR approach facilitates the broadening of topic scope, expanding the domain of knowledge available to researchers working in the field [60]. According to the adopted methodology, this work was developed in three sequential stages: a) Data collection, b) Bibliometric analysis, and c) Systematic Review of the Literature, as illustrated in Figure 1.

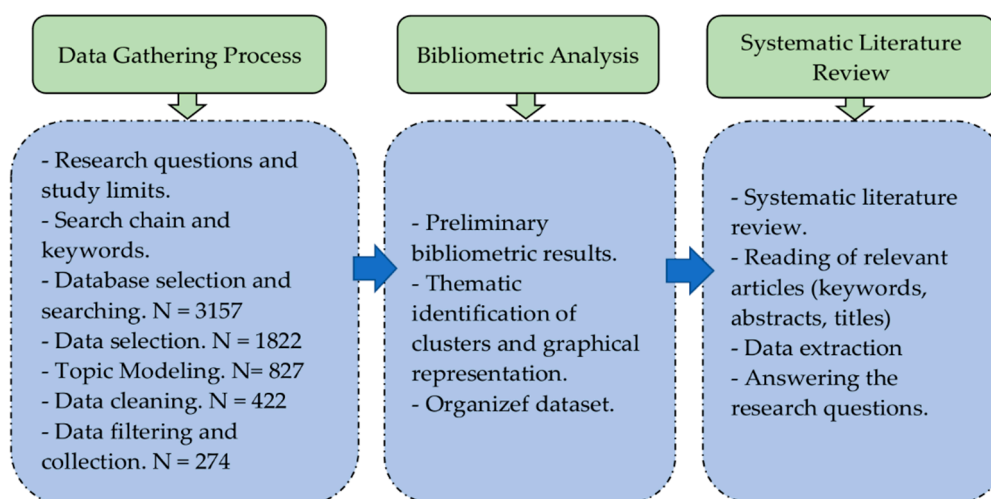


Figure 1. Stages of the B-SLR methodology.

2.1. Data Gathering Process

This stage aims to identify the research topic and the structural framework to establish the research questions, keywords, search strings, database, time scope of the research, techniques to carry out the search and analysis tools, where each stage generates a result that serves as an input for the next stage.

The research questions raised in this study seek to identify a set of patterns behind the application of AI/ML/DL techniques in water quality prediction. Table 1 presents the guiding research questions of this study with their corresponding justification, following the guidelines of the systematic review of the literature [65].

Research Questions	Justification
RQ1. What are the most commonly used AI/ML/DL algorithms for predicting water quality	To establish a general overview of the research topic.

RQ2. Which AI/ML/DL algorithm provides the most accurate estimation of water quality?	To identify knowledge gaps in AI/ML/DL prediction models.
RQ3. What limitations have been identified in water quality prediction using AI/ML/DL techniques?	To uncover potential research opportunities and future work
RQ4. What emerging variants currently exist in AI/ML/DL models for estimating water quality?	To identify current trends in AI/ML/DL techniques for water quality prediction.
RQ5. What are the key water quality indicators used to assess natural water sources?	To review and understand the factors that determine water quality.

To obtain the bibliographic information necessary to answer the research questions posed, the following line of code was generated for the search and construction of the collection of bibliographic references:

```
water AND quality AND prediction AND machine AND learning) OR TITLE-ABS-KEY (water AND quality AND prediction AND artificial AND intelligence) OR TITLE-ABS-KEY (water AND quality AND prediction AND deep AND learning)) AND PUBYEAR > 1999 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SUBJAREA, "ENVI") OR LIMIT-TO (SUBJAREA, "ENGI") OR LIMIT-TO (SUBJAREA, "EART") OR LIMIT-TO (SUBJAREA, "MULT"))).
```

The temporal scope of this study was defined from January 2000 to December 2024. Bibliographic references were obtained from the Scopus database (<https://www.scopus.com/home.uri>), which offers a comprehensive and global overview of scientific production. Scopus is one of the most widely used sources in bibliometric analysis and is recognized for its reliability and high-quality data, particularly in academic research related to water resources and water quality [17,56,57,67–69].

The selection criteria for documents included in the bibliographic reference collection were as follows: a) Publications classified as “Research Article” or “Review,” excluding conference proceedings, books, book chapters, theses, and reports; b) Publications written in English; c) Publications within the subject areas of Environmental Science, Engineering, Earth and Planetary Sciences, and Multidisciplinary; d) Articles containing the keywords specified in the search string. In this first retrieval, a total of 1822 articles were obtained.

Bibliographic references were exported from Scopus in CSV format and subsequently imported into the R programming environment [63]. A comprehensive list of 29,144 Scimago journals (<https://www.scimagojr.com/>) for the year 2023 was incorporated to filter the 1,822 original articles selected for B-SLR analysis, retaining only those published in journals classified within Q1 and Q2 quartiles, in accordance with the recommendations outlined in reference [61]. Publications to 2014 were excluded, establishing the analysis period between 2015 and 2024. This procedure involved excluding publications prior to 2014, thereby establishing the analysis time frame from 2015 to 2024. In addition, a systematic cleaning process was conducted, which included the removal of duplicate rows, incomplete records in the fields Index Keywords, Author Keywords, DOI, and Abstract, as well as the elimination of entries with duplicated DOIs and false positives [70], that is. These publications did not align with the objective of the present study.

The automation of this process was achieved through the implementation of an iterative text mining approach based on Topic Modeling, using the TextMiner package in R as a specialized tool for constructing and analyzing thematic models [71]. The procedure consisted of: a) Merging the Title, Abstract, and Author Keywords fields from Scopus into a single column; b) Applying the Topic Modeling procedure to detect reference topics for each record; c) Grouping the records into 30 clusters, each characterized by the five most frequent topics; and d) Identifying terms that, being representative of the Title, Keywords, and Abstract, indicated that the publication fell outside the scope of the study.

Records containing the terms selected in step d) were removed from the database, and a new iteration was performed. In each iteration, the appropriateness of each removal was confirmed through manual inspection.

The record elimination procedure, assisted by Topic Modeling, resulted in a preliminary database of 827 records. To ensure that the publications were aligned with the scope of the study, only those published in journals whose titles included the expressions Hydro/ Enviro/Water were retained. Furthermore, only journals with at least three published articles were considered. This filtering process yielded a refined database of 424 articles. Subsequently, all records underwent manual inspection, which involved reviewing the title, abstract, introduction, methodology, and conclusions to exclude publications that did not meet the thematic criteria of the study. After this final validation step, the resulting database comprised 276 articles published between 2015 and 2024. These records were used in the subsequent stages of the bibliometric analysis (BA) and systematic literature review (SLR).

2.2. Bibliometric Analysis (BA)

The bibliometric analysis was conducted using the Bibliometric package in the R programming language [63]. This package has been widely applied in bibliometric studies related to hydrology [56,72] and water quality [73,74]. The analysis included the generation and visualization of maps and graphs corresponding to performance analysis [17,56,62] and science mapping [57,75,76].

2.3. Systematic Literature Review (SLR)

A comprehensive reading stage was conducted for the 276 selected articles, conducted using the Systematic Literature Review (SLR) approach described in [60], in order to extract key data and information to address the research questions. This methodology enabled the identification, collection, analysis, and synthesis of relevant information to answer the research questions posed [77]. Its robustness stems primarily from the transparency of its implementation process, which ensures the reproducibility of the review [78,79].

Each study was classified as either a Review or a Research article. Within the Research category, articles were further classified according to the type of water body under study: surface water or groundwater. A particular case involved two Review articles [80,81] that addressed both systems (groundwater and surface water); this condition did not affect the classification by water body type, as these studies were not part of the Research category.

3. Results

3.1. Bibliometric Performance Analysis

The annual scientific output on water quality prediction using AI/ML/DL has increased in the study period, as shown in Figure 2, exhibiting an exponential increase, especially marked since 2020.

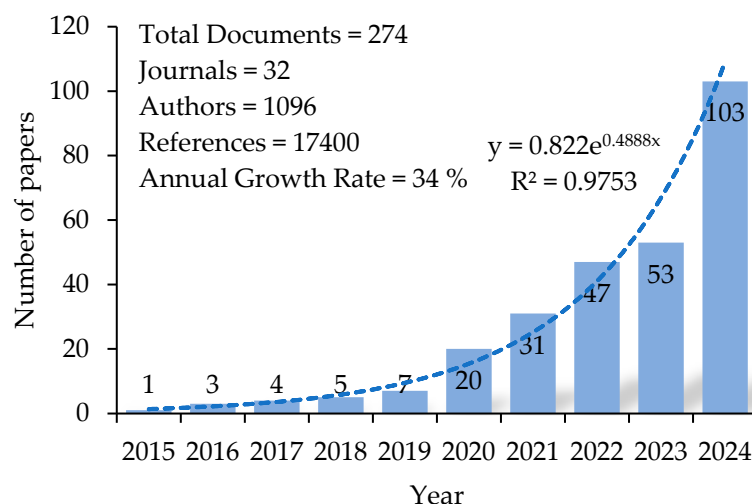


Figure 2. Annual publication output on freshwater quality prediction using AI/ML/DL.

The growing interest in research on the prediction of water quality using AI/ML/DL is within a global context driven by climate change, and increased pollution and demand for fresh water [5,82]. This evolution indicates that the analyzed domain has passed an initial stage of exploration, positioning itself in a phase of maturity. Likewise, it is possible to expect an even more significant development in the coming years, anticipating a continuous strengthening of both interest and research in this field [83,84].

3.1.1. Journals

When ranking the top 10 of the 32 sources based on the H-index and Total Citations (TC) [85], it is observed that the journals Water (Switzerland), Journal of Hydrology, Environmental Science and Pollution Research, Science of the Total Environment and Water Research are the publications with the highest number of papers (first 5) within the reference collection (Table 2).

Table 2. Journals Local impact.

ID	Journals	H Index	TC
1	Water (Switzerland)	19	1568
2	Journal of Hydrology	18	2231
3	Environmental Science and Pollution Research	14	803
4	Water Research	10	823
5	Science of the Total Environment	9	737
6	Journal of Environmental Management	7	193
7	International Journal of Environmental Research and Public Health	6	103
8	Environmental Monitoring and Assessment	5	115
9	Hydrological Processes	5	82
10	Process Safety and Environmental Protection	5	147

3.1.2. Most Cited Documents

The most cited documents within the analyzed collection of bibliographic references were identified. Both global and local citation counts were considered to assess the reach and influence of the research from complementary perspectives. Local citations reflect the impact of a document within the specific dataset under analysis, while global citations represent the total number of times each article has been cited in the Scopus database.

Table 3 provides the top 10 most cited publications and corresponding (leading) authors in terms of local and global citations. The referred papers reveal important contributions to research on water quality prediction using AI/ML/DL techniques.

Table 3. Top ten Local citations and Global citations.

Ranking	First author	Year	LC ¹	GC ²	Reference
1	Rahim Barzegar	2020	32	330	[86]
2	Rahim Barzegar	2016	15	149	[87]
3	Jun Yung Ho	2019	13	101	[88]
4	Amir Hamzeh Haghiabi	2018	13	290	[38]
5	Xiaoliang Ji	2017	11	10	[89]
6	Elham Fijani	2019	10	146	[90]
7	Sani Isah Abba	2020	9	91	[91]
8	Muhammed Sit	2020	9	273	[80]
9	Roohollah Noori	2015	9	66	[92]
10	Bachir Sakaa	2022	7	66	[93]

¹ LC= Local citation. ² GC= Global citation.

According to Table 3, the most cited reference corresponds to the work of Barzegar [86] entitled "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model", which investigates the prediction of water quality in Lake Prespa, Greece, and proposes a hybrid CNN-LSTM model to predict dissolved oxygen (DO; mg/L) and chlorophyll-a (Chl-a; µg/L).

Second, the study [87] compares different machine learning methods (such as artificial neural networks and support vector machines) for the prediction of water quality parameters of northwest Iran's Aji-Chay River, demonstrating the effectiveness of these techniques in improving environmental monitoring. Its high number of global citations reflects the interest in using water quality prediction to support the transition toward more sustainable water resources management.

Third, the work of [88] examined the application of a decision tree model for predicting the Water Quality Index (WQI) in the Klang River. The research successfully showed that the number of water quality parameters needed for monitoring can be reduced while maintaining prediction accuracy above a 75% benchmark. Finally, the most recent study of the list [93], proposes a hybrid approach that combines the Random Forest algorithm with an improved version of the SMO algorithm for support vector machines. Applied to the Saf-Saf River Basin, this model improved WQI prediction, highlighting the growing role of optimized hybrid approaches in watershed management.

3.2. Bibliometric Science Mapping

3.2.1. Network Analysis on Co-Occurrence of Authors' Keywords

The network analysis of co-occurrence of author keywords presented in Figure 3 reveals a conceptual structure derived from studies on machine learning and water quality management. Each node represents a keyword, with its size indicating the degree of connectivity (number of links to other terms). Connecting lines reflect co-occurrence relationships, and their thickness represents the strength of those associations. Nodes are grouped into three thematic communities, distinguished by color: group 1 (blue), group 2 (orange), and group 3 (green), enabling the identification of conceptual subdomains within the field.

Central terms such as support vector machine, deep learning, water quality predictions, and water quality management, along with technical acronyms like ML (machine learning), AI (artificial intelligence), SHAP (SHapley Additive exPlanations), ANN, CNN (convolutional neural networks),

LSTM (long short-term memory), and random forest, exhibit high connectivity. This suggests their integrative role in the analyzed literature and highlights their methodological relevance in the development of predictive and explanatory models applied to water systems. The visualization supports the identification of thematic clusters, methodological relationships, and emerging areas at the intersection of artificial intelligence and environmental management.

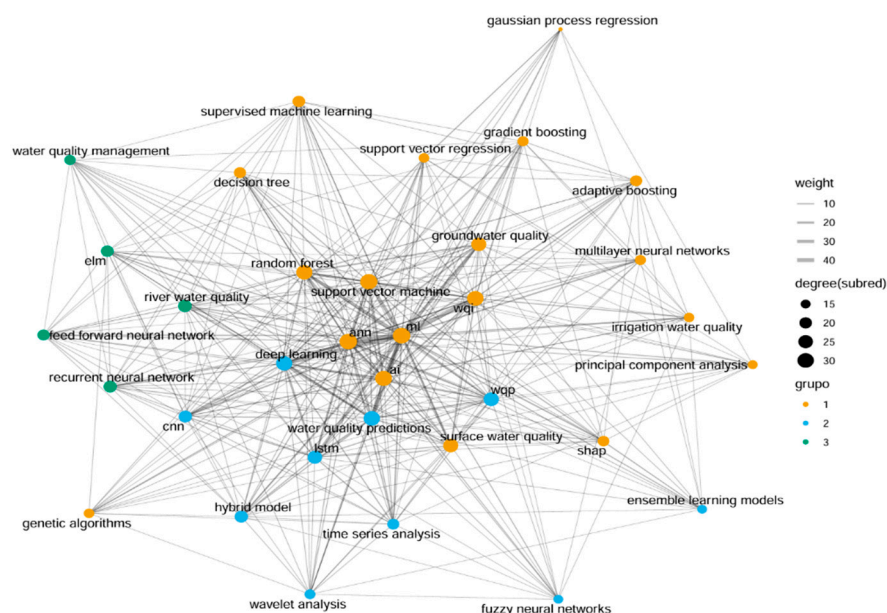


Figure 3. Network analysis co-occurrence of the author's keywords.

3.2.2. World Cloud

Another method for identifying the frequency of the most common terms used by the authors in the collection corresponds to the word cloud generated from the keywords. Figure 4 shows a predominant interest in “water quality,” which serves as the central axis of the study corpus. Its association with indicators such as “WQI,” “WQP,” and water bodies like “groundwater quality” and “surface water” reaffirms the environmental orientation of the studies. From a methodological standpoint, there is a predominant interest in techniques such as “random forest,” “deep learning,” “ML,” and “ANN,” along with specific architectures like “LSTM,” “CNN,” “RNN,” and “SVM,” used in modeling and prediction processes of hydrological parameters.

Complementing this trend is the appearance of terms like “hybrid model,” “ensemble learning,” and “transfer learning,” reflecting the adoption of integrated strategies and techniques for transferring pretrained representations aimed at improving model generalization and efficiency. Finally, “XAI” and “SHAP” reveal an interest in model explainability, pointing toward advanced analytical systems for the management and evaluation of water sustainability.

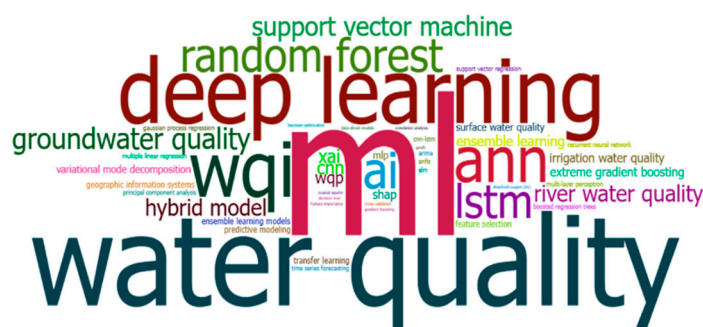


Figure 4. World cloud of water quality prediction by AI/ML/DL.

3.2.3. Thematic Map with Authors' Keywords

Figure 5 presents a thematic map derived from bibliometric analysis, grouping clusters associated with distinct research subtopics in the field of water quality prediction using AI, ML, and DL. This map enables the evaluation of each topic in terms of centrality and maturity, identifying established, emerging, and declining areas. In the quadrant of Basic themes, terms like “water quality prediction,” “deep learning,” and “LSTM” reflect the consolidation of deep neural networks as a methodological foundation. Notably, LSTM networks have been highlighted in multiple studies for their strength in modeling complex temporal dynamics in water quality parameter prediction [31,86,94,95]. Furthermore, hybrid models and the ANFIS technique have provided enhanced precision and flexibility in complex scenarios, particularly in surface water contexts [87,96–98].

In the quadrant of Motor themes, terms such as “WQI,” “ANN,” “AI,” and “ML” stand out, indicating that artificial neural networks and machine learning models have been extensively studied and applied, yielding significant results. Conversely, “transfer learning” emerges as a nascent exploratory line within water quality prediction, although several studies [6,94,99–102] have demonstrated its potential to improve predictions in data-scarce environments.

In the quadrant of Niche themes, the terms “data-driven models,” “lake water quality,” and “prediction intervals” suggest that lake water analysis and data-driven modeling represent growing niches within water-related research, though they have yet to become dominant in AI/ML/DL-based water quality prediction.

In the quadrant of Emerging or declining themes, terms such as “support vector machine,” “feature selection,” and “gaussian process regression” indicate that these methods (SVR, GPR), although historically robust and widely used in this field, may be losing relevance compared to more modern approaches such as deep learning. Additionally, procedures like feature selection may have already transitioned into standard practice within AI/ML/DL applications. Meanwhile, “extreme gradient boosting” and “SHAP” are associated with advanced ML techniques but have not yet reached the prominence of core topics. The emerging discipline known as Explainable Artificial Intelligence (XAI), or Interpretable Machine Learning (iML), seeks to address the challenges posed by the opacity of black-box techniques in AI/ML/DL. Its application has expanded in recent years across various engineering domains [83,103–105], and has recently been implemented in the field of water quality, particularly in groundwater studies [106,107].

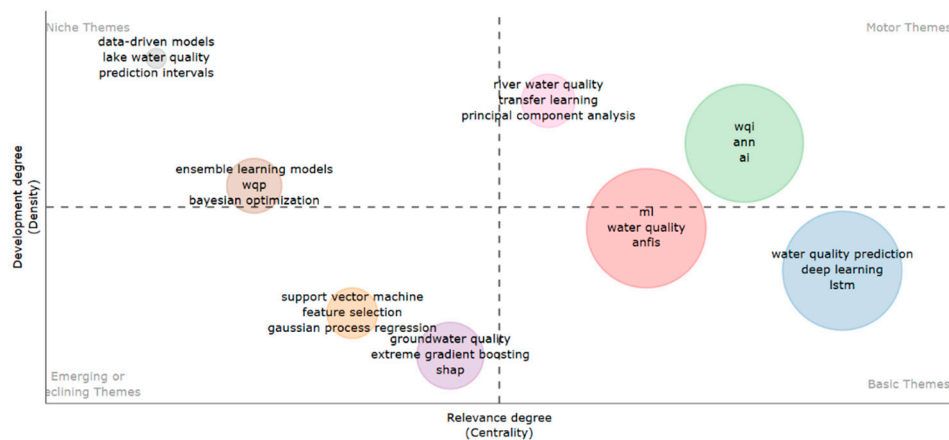


Figure 5. Thematic map based on the author's keyword.

3.2.4. Thematic Evolution and Trend Topics with Keywords Plus

The thematic evolution in the BA is useful to visualize more clearly the change and evolution of research topics over time, allowing the identification of windows of opportunity for future research

work. To have a more global view of the thematic evolution and current trends in water quality prediction using AI/ML/DL, Figure 6 presents the thematic evolution based on the Keywords Plus. These words are automatically generated by the BA algorithms from sources such as the titles of the references cited in the documents, allowing for to assessment, from a complementary perspective, of the areas that are related to the central axis of the research search. Thus, the application of models based on ANN and decision support techniques is maintained over time, while models based on SVM have been less constant in recent years.

In recent years, cutting-edge ML and DL techniques have increasingly been applied to environmental challenges such as pollution, climate change, water quality, and water availability. These approaches are often integrated with Geographic Information Systems (GIS) technologies [1,2,108–110], enhancing spatial analysis capabilities.

While GIS is now widely used in the field of water resources, it has yet to become a dominant tool for predicting water quality through AI/ML/DL methods. Notably, interest in developing predictive models for surface water quality has remained consistent throughout the study period. Since 2015, ML has emerged as the leading approach, demonstrating strong applicability in environmental monitoring, wastewater treatment, and water quality assessment, indicating a sustained research focus on water resources management.

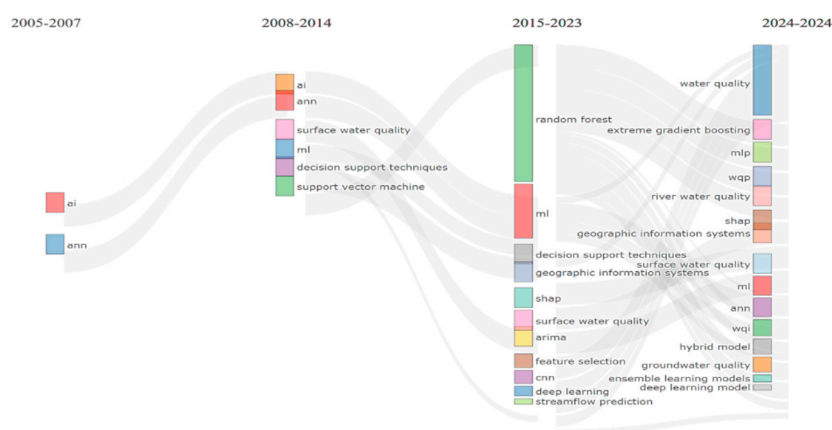


Figure 6. Thematic evolution of AI/ML/DL applications in water quality prediction.

3.2.5. Social Structure

Figure 7 presents the map of main collaborations between countries, which reveals a highly interconnected structure of scientific cooperation. Countries such as China and the United States stand out, acting as central nodes, evidencing their leadership in the production of knowledge and in the articulation of international efforts. This dynamic reflects not only the shared interest in addressing the challenges of water quality prediction but also the growing need to integrate data, methodologies, and interdisciplinary approaches on a global scale. Countries such as India, Australia, the United Kingdom, Italy, and Germany stand out for their collaborative connections as well.

However, the map also reveals a striking absence of Latin American countries, suggesting a potential asymmetry in scientific visibility. This underrepresentation may be partially attributed to the methodological filter applied in this study, which included only publications indexed in high-impact Q1 and Q2 journals. While this criterion ensures scientific rigor, it may inadvertently exclude valuable regional research that, due to structural or editorial limitations, does not reach these publication venues. As a result, the map not only visualizes collaboration intensity but also reflects broader epistemic inequalities in global scientific discourse.

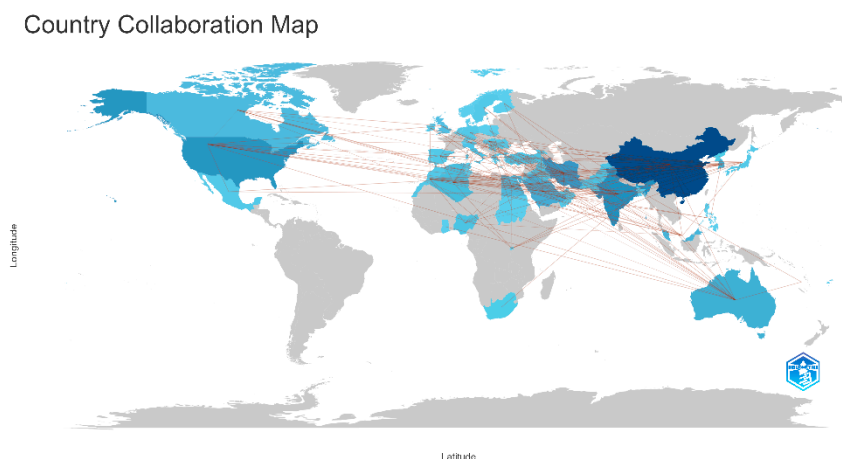


Figure 7. Collaboration with country authors.

3.3. Systematic Literature Reviews Results

The systematic review allowed us to classify each work into different categories according to: Approach (Review or Research articles), body of water of the study, that is, Surface water (rivers, reservoirs and lakes) or Groundwater, as presented in Figure 8. Figure 8(a) shows the classification of the 274 articles collected according to their type: Research Articles and Review Articles. The disparity between the two approaches suggests that, although research in the area is growing rapidly, a significant gap persists in terms of consolidating knowledge through systematic reviews of the state of the art. This represents a valuable opportunity to strengthen the field through studies that integrate and synthesize existing findings (such as ours).

Within Research articles, 75% ($n = 190$) correspond to studies focused on surface water quality, while the remaining 25% ($n = 65$) address groundwater quality issues (Figure 8(b)). This bias towards the study of surface waters can be attributed to the relevance of watershed monitoring for the sustainable management of water resources. This monitoring allows us to understand the complex interactions between biological, chemical, physical, and environmental factors that determine water quality, as well as to anticipate the appearance of alterations to water quality. This information is crucial for informed decision-making and long-term planning, which are essential for preserving the health of water bodies and the ecosystems that depend on them [111].

In Figure 8(c), we present the distribution of works considering only surface water bodies, i.e., rivers, lakes, and reservoirs. It is shown a major focus in rivers, which could be because they are particularly exposed to pollution caused by anthropogenic activities [40,112]. In addition, during periods of drought, rivers are more easily and rapidly affected by the reduction of the flow, affecting the quality of the water in terms of its physical, chemical and biological properties [113]. The study [58], made a significant contribution with the exhaustive review of the literature in the period 2000-2020 on the modeling of river water quality in the world, reviewing 209 research articles from Scopus journals. The results showed that most of the study areas are Asian countries, such as China, Iran, India, Malaysia, Taiwan, Korea, Iraq, Bangladesh and Thailand, accounting for more than 50% of research papers. These results are consistent with the findings of the current work for the period 2015-2024. Based on the classification shown in Figure 8, this study delves into the two water bodies that are currently of greatest interest in the application of AI/ML/DL techniques for water quality prediction: rivers and groundwater. For this reason, a detailed description of the literature for them is presented below.

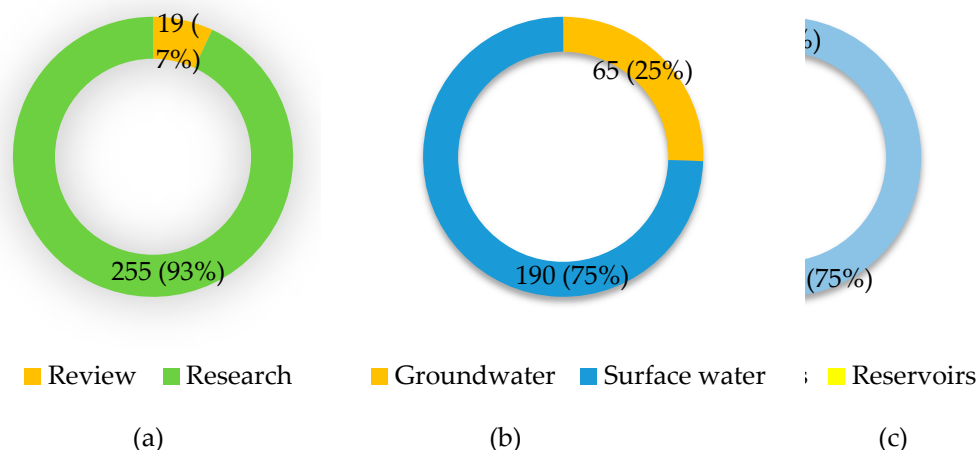


Figure 8. Classification of publications in water quality prediction according to (a) Approach: Article research versus Review. (b) Body of water: Underground versus Surface, (c) Surface water: River, lake, reservoir.

3.3.1. Prediction of River Water Quality Using AI/ML/DL

Focusing on the publications on water quality prediction in rivers, Table 4 presents the characterization of a random sample of the database, equivalent to 40% (n=57).

Table 4. Characterization of predictive models in a representative sample of river water quality prediction studies (n = 57).

ID	River	Algorithm	Approach	Reference
1	Yangtze River, China	CNN-LSTM	WQP	[114]
2	Delaware River Basin, USA	XGB, RF, KNN	WQP	[115]
3	Sheshui River in Wuhan, China	RF, SSA-CNN-LSTM	WQP	[112]
4	Vaigai, Madurai, and Tamil Nadu Rivers, India	Optimization algorithm and LSTM	WQP	[116]
5	Upper Red River Basin (URRB), USA	TL, FFNNs	WQP	[100]
6	The South Platte River, Colorado, USA	EBM, SWEBM	WQP	[117]
7	Cauvery River, India	AO-SVM	WQI	[118]
8	Indian Rivers	DT, RF, GBT, ANN, SVM	WQP	[43]
9	Cauvery River, India	CNN	WQP	[119]
10	Han River, South Korea	RF, SVR, XGB, LGB, and a hybrid model. SHAP, LIME	WQI	[120]
11	Tanjiang River, China	SVR	WQP	[121]
12	Des Moines, Iowa, and Cedar Rivers, Iowa, USA	LSTM, GRU	WQP	[122]
13	Mahanadi River, India	LSTM, GRU, XGB	WQI	[123]
14	Oyster River, New Hampshire, USA	CNN-LSTM	WQP	[124]
15	Li River and Liu River, China	SSA, GRU, SHAP	WQP	[125]
16	Fujian River Network, China	WA-LSTM-TL	WQP	[101]
17	Euphrates River, Iraq	RC, DR, REPT, AR	WQP	[126]
18	Ohio River, USA	LSTM	WQP	[109]
19	USA Rivers	RF	WQP	[127]
20	Xiaofu River, China	LSTM	WQP	[128]
21	Lijiang River, China	BPNN, SVR, GRU	WQP	[129]
22	Drinking water quality, South Korea	LSTM, GRU	WQP	[130]

23	Indian, Rivers*	DT, LR, Ridge, Lasso, SVR, RF, ET, ANN	WQI	[131]
24	Júcar River, Spain	RF, XGB, SHAP	WQP	[132]
25	Bullfrog River, Tampa, Florida USA	SVM, RF, XGB, ANN, SHAP	WQP	[133]
26	Talar River, Iran	EN, AMT, REPT	WQP	[134]
27	Wadi Saf-Saf River, Algeria	SMO-SVM, RF	WQI	[93]
28	Pearl River, China	CEEMDAN -LSTM	WQP	[135]
29	Fuyang River, China	RF, PLS	WQP	[136]
30	Synthetic data set Wabash River, USA	SVMR	WQP	[137]
31	Yamuna River, India	LSTM, SVR, CNN-LSTM	WQP	[31]
32	Kelantan River, Malaysia	KNN, ANN, DT, RF, GB	WQP	[138]
33	Langat River, Malaysia	ANN	WQP	[139]
34	Mid-Atlantic and Pacific Northwest USA, River Basin	SVR, XGB	WQP	[140]
35	Santiago-Guadalajara River, Mexico	SLR, MLR	WQI	[141]
36	Danube, Tisa, and Sava Rivers, Vojvodina Province, Serbia	Naïve Bayes algorithm	WQI	[142]
37	Yamuna River, India	ANFIS-GP, ANFIS-SC	WQP	[143]
38	Fanno Creek in Oregon, USA	DRNN, SVM, ANN	WQP	[144]
39	Dongjiang River, China	WT-MLR, WT-SVM, WT-ANN, WT-RF	WQP	[145]
40	Klang and Penang Rivers, Malaysia	MLP, SVM, RF, BDT	WQI	[146]
41	Nakdong River, South Korea	CEEMDAN, CSA, MARS	WQP	[147]
42	Luan River, Tangshan China	1-DRCNN*, BiGRU	WQP	[148]
43	Tyhume, Bloukrans, Buffalo Rivers Province of South Africa	ANN, MLP, RBF	WQP	[149]
44	Kinta River, Malaysia	EANN-GA, EANN, FFNN, NNE	WQI	[150]
45	Xin'anjiang River, China	CNN-LSTM, CEEMDAN	WQP	[151]
46	The Juhe River, Sanhe China	PSO-DBN-LSSVR	WQP	[25]
47	Burnett River, Australia	kPCA, RNN, FFNN, SVR, GRNN	WQP	[152]
48	Nakdong River, South Korea	CNN-LSTM	WQP	[153]
49	Sefid Rud River, Iran	W-MGGP, GEP, DWT	WQP	[154]
50	Talar River, Iran	RF, RFC	WQI	[155]
51	Yangtze River, Jiangsu, China	IABC-BPNN	WQP	[156]
52	Klang River, Malaysia	DT	WQI	[88]
53	Langat River, Malaysia	MLP-FFA	WQP	[157]
54	Tireh River, Iran	ANN, GMDH, SVM	WQP	[38]
55	Danube Delta River, Romania	ANN, KNN, BPNN	WQI	[158]
56	Sefidrood River, Iran	SVM	WQP	[92]
57	Aji-Chay River, Iran	ANN, ANFIS, WT	WQP	[87]

*Indian water quality data from Kaggle, 1-DRCNN: One-dimensional residual convolutional neural networks.

Recent advances in AI/ML/DL techniques have significantly enhanced the prediction of river water quality, as demonstrated by the 57 studies summarized in Table 4. These investigations employ a range of algorithmic approaches, including CNN-LSTM, XGB, RF, LSTM, SVM, and hybrid models to estimate key indicators such as the Water Quality Parameter (WQP) and the Water Quality Index (WQI) across rivers in Asia, America, Europe, and Africa. Artificial Neural Networks (ANNs) have played a prominent role in this domain, owing to their capacity to model complex nonlinear relationships and anticipate fluctuations in water quality, as confirmed by studies [133] and [138]. In [43], ANN-based models were evaluated alongside Gradient Boosted Trees (GBT), Decision Trees (DT), Support Vector Machines (SVM), and Random Forests (RF) for predicting the WQI of the Indian river system, with GBT achieving the highest performance. Similarly, [155] implemented four standalone decision tree models and twelve hybrid configurations to estimate the WQI of the Talar River in Iran. The Bagging-Random Tree (BA-RT) approach yielded the most accurate results, reinforcing the superiority of hybrid models over simpler alternatives, as also noted in [159].

The integration of optimization algorithms has proven to significantly enhance the accuracy of machine learning (ML) models in predicting river water quality. For instance, studies [25] and [93] employed Particle Swarm Optimization (PSO) and Sequential Minimal Optimization (SMO), respectively, to strengthen models such as PSO-DBN-LSSVR for the Juhe River in China and SMO-SVM for the Wadi Saf-Saf River in Algeria. Additionally, the application of Wavelet Transform (WT) has been key in identifying relevant variables and reducing noise, as demonstrated in studies on the Aji-Chay River (Iran) [87], Fujian (China) [101], Dongjiang (China) [145], and Sefid Rud (Iran) [154]. These enhancements have led to improved predictive performance in models such as WT-SVM, WT-RF, and WT-ANN. Another widely applicable technique is XGB, which has been successfully implemented in basins such as the Delaware River (USA) [115], the Han River (South Korea), and regions of the northwestern United States [140], due to its ability to capture complex interactions among predictors and deliver high predictive accuracy.

Hybrid architectures that integrate Convolutional Neural Networks (CNNs) with LSTM units have also demonstrated strong performance in river water quality prediction, as shown in studies on the Yangtze River [114,156]. In [112], the SSA-CNN-LSTM model was applied to the Sheshui River, achieving effective integration of temporal and spatial patterns for estimating the Water Quality Parameter (WQP). Other studies have explored variants such as Gated Recurrent Units (GRU) [122,125], Bidirectional GRU (BiGRU) networks [148], and hybrid approaches like ANFIS-GP and ANFIS-SC [143] to address nonlinear modeling challenges. In this context, emerging techniques such as Transfer Learning (TL) have shown considerable promise by enabling the reuse of previously acquired knowledge in new environments, thereby enhancing predictive performance as demonstrated in studies on the Fujian river system in China [100,101].

Finally, the geographical diversity of the studies listed in Table 4 spans a wide range of river systems, including the Yangtze [114,156], Sheshui [112], Tanjiang [121], Li and Liu [125], Pearl [135], Fuyang [136], Xiaofu [128], Lijiang [129], Juhe [25], Euphrates [126,143], Júcar [132], Yamuna [31,143], Langat [139], Klang [146], Sefid Rud [154], Talar [134], Sefidrood [92], Aji-Chay [87], Nakdong [147,153], Oyster River [124], Upper Red River Basin [100], Danube [142], Burnett [152], Kelantan [138], and Bullfrog [133]. These rivers span diverse climatic zones from tropical and temperate to arid and mountainous and reflect a broad spectrum of hydrological and geological contexts. This territorial breadth highlights the versatility of AI/ML/DL models in adapting to varied environmental conditions, reinforcing their value as effective tools for the sustainable management of water resources.

3.3.2. Prediction of River Water Quality Using AI/ML/DL

Regarding groundwater, Table 5 presents a random sample of 40% (n=26) of the publications on groundwater quality prediction found in our database, spanning various regions of the world. Diverse environmental, hydrogeological, socioeconomic, and technological factors have shaped the development of advanced computational approaches for predicting groundwater quality.

These approaches range from traditional statistical models and standalone machine learning algorithms (e.g., decision trees, support vector machines, etc.) to hybrid, ensemble methods and SHAP-enhanced models. They also include the integration of geospatial tools, remote sensing data, and cloud-based platforms for data processing and visualization. The most frequently studied contaminants are nitrates, arsenic, fluoride, heavy metals, and total dissolved solids. In addition to studies focused on individual water quality parameters (WQPs), there is research incorporating integrative water quality indices such as the Water Quality Index (WQI), Irrigation Water Quality Index (IWQI), Entropy-weighted Water Quality Index (EWQI), and Groundwater Quality Index (GWQI).

Table 5. Characterization of predictive models in a representative sample of groundwater quality prediction studies (n = 26).

ID	Region	Parameters	Algorithm	Reference
1	Madrid, Spain	Nitrate concentrations	DT, RF, AdaBoost, ExT	[160]
2	Songyuan City, China	Strontium (Sr ²⁺)	GAN, KNN, GPR	[161]
3	Mekong Delta región, Vietnam	Salinity levels	Bagging, CatBoost, ExT, HGB*, XGB, DT, RF, LightGBM, KNN, SHAP	[162]
4	Eden Valley, Cumbria, North West England	Nitrate concentrations	DT, XGB, RF, KNN, SHAP	[163]
5	Kerala, India	EWQI	XGB, SVR, ANN, RF	[47]
6	The Mitidja plain, northern Algeria	IWQI	LSTM	[164]
7	Groundwater dataset	Salinity levels	GMDH algorithm	[165]
8	Tamil Nadu, India	IWQI	SVM, ANN, LRM, RT, GPR, BRT	[166]
9	North China Plain, Beijing	Arsenic (As) and fluoride (F ⁻) concentrations	XGB, RF, SVM,	[167]
10	Eastern India	WQI	MLP-ANN	[168]
11	Raipur district, Chhattisgarh, India	WQI	ANN-LR	
12	Midwestern United States Hawasinah catchment	Redox Conditions	GBM, XGB, RF CatBoost regression, ExT	[169]
13	Wilayat Al-Khaburah, Oman	TDS	regression, Bagging regression	[170]
14	Vehari, Punjab Province of Pakistan	WQI	ANN, RF, LR	[171]
15	Northeast of Tamil Nadu, India	WQI	GB, RF, DT, KNN, MLP, XGB, SVR	[172]
16	Qom City, Iran	Nitrate concentration	KNN, SVR, RF	[173]
17	Savar, Dhaka district, Bangladesh	GWQI*	LR, SVM, ANN	[174]
18	Al Qunfudhah, Saudi Arabia	WQI	CNN, XGB, SHAP	[175]
19	Fars Province, Iran	WQI	RF, BRT, MnLR	[176]
20	Wendeng District, China	WQI	LSTM	[177]

21	Taiwan Groundwater Pollution Monitoring Standard	Heavy Metal Concentrations	SVR, KNN, MLP, GBR, LIME, SHAP	[178]
22	Middle Black Sea Region of Turkey	WQP	CNN, RF, XGB, DNN	[179]
23	Noida, Uttar Pradesh, India	WQP	MLR, SVR, DT	[180]
24	The Akot basin, Akola district of Maharashtra, India	IWQI	ANN, LSTM, MLR	[181]
25	North Carolina, USA	Nitrate concentrations	RF	[182]
26	Dezful Aquifer, Iran	TDS	RF	[183]

The wide variety of AI/ML/DL models highlights a transition from traditional approaches such as ANN, LSTM, Multilayer Perceptron (MLP) and CNN, to more sophisticated models such as XGB and LightGB. These models are generally applied in studies that require modeling nonlinear relationships or time series, and have been widely used to estimate water quality indices as well. From the reading of the studies listed in Table 5, it is possible to identify a current trend in the use of ensemble-type algorithms, such as Random Forest, Gradient Boosting and Bagging, which are considered robust methods capable of handling the high dimensionality and multicollinearity of predictive models [47,162].

Other ensemble algorithms such as CatBoost, Bagging, and Extra Trees have also been frequently used to predict specific WQP such as nitrates [160,163,174,182], salinity levels [162,171], metals [161,178], as well as water quality indices such as WQI, IWQI, EWQI, and GWQ. These algorithms allow multiple hydrochemical variables to be handled, and reliable predictions to be constructed in complex environments. Regression models have also been widely used in this area, such as the use of Multinomial Logistic Regression (MnLR), which allows water quality to be classified into multiple categories, being useful in environmental risk and zoning studies [176].

Among the emerging models, Generative Adversarial Network (GAN) [161] and Group Method of Data Handling (GMDH) [165] stand out. They are generally used in contexts with complex dynamics, particularly with scarce data, and be particularly useful in predictions of Sr^{2+} and salinity levels [161]. In addition, explanatory models such as SHAP and LIME represent a significant advance in the interpretability and explainability of predictive models applied to groundwater quality [178]. These techniques, typical of the field of XAI, are important because they allow the outputs of complex algorithms to be broken down into quantifiable contributions of each input variable, allowing one to understand how much each factor influences the final prediction. For example, in [162] SHAP was applied to assess the impact of physicochemical factors on salinity levels in multiple aquifers, revealing key spatial patterns using interpretable maps. Similarly, [163] used SHAP in a nitrate prediction model to identify the main geoenvironmental pollution-related drivers in a UK aquifer, integrating interpretation and prediction into a unified framework.

In this sense, the integration of GIS technologies, together with XAI models, strengthens the capacity to spatially represent the results, detect risk areas and generate useful tools for water management. Together, the AI/ML/DL models allow the prediction of groundwater quality to be approached from an interdisciplinary, explanatory and applicable perspective, configuring themselves as key tools for water sustainability in complex environmental scenarios [106,168].

On the other hand, in the hydrogeological context, studies are conducted in regions characterized by a wide range of factors, including aquifers affected by saline intrusion (Vietnam, Iran), arid zones (Algeria, Saudi Arabia), and areas with carbonate lithology that induce water hardness (India, USA). Natural processes such as weathering, evapotranspiration, and the influence of volcanic emissions also contribute to the spatial and temporal variability of hydrochemical parameters [161,177,178]. These hydrogeological factors, combined with local climatic conditions, shape physically complex and dynamic environments that demand predictive models with high

adaptive capacity. In this context, the suitability of groundwater quality for human consumption is conditioned by these factors, among them, variations in well depth, which can significantly alter water mineralization and contaminant mobility due to interactions with lithological formations, hydraulic gradients, and specific redox conditions [172,184]. Therefore, the applied algorithms must be capable of capturing aquifer heterogeneity by integrating geological and hydrodynamic variables that directly influence groundwater quality.

Finally, at the regional level, Asia leads scientific production, with strong representation from India, China, Iran, and Bangladesh. These countries face severe water stress, which justifies the scientific community's growing interest in groundwater quality and the development of advanced predictive models.

4. Answering the Research Questions

Consistent with the B-SLR framework adopted in this study, the driven research questions are answered below.

RQ1. What are the most widely used AI/ML/DL algorithms in water quality prediction?

The literature review showed that assembly models, such as Bagging and Boosting, have established themselves as predominant techniques in the prediction of both surface water and groundwater quality [185]. Their effectiveness is supported by studies that demonstrate their ability to reduce mean absolute errors and quadratic errors [186]. In addition, recent research has explored variants of these models, such as Grid Search Random Forest (GS-RF) and XGB, to optimize prediction accuracy on parameters such as turbidity and different nutrients [187], showing that algorithms such as CatBoost Regression (CBR) offer advantages in terms of stability and adaptability, particularly in handling heterogeneous datasets, minimizing overfitting, and maintaining consistent performance across varying environmental conditions and input configurations [171].

These assembly models have been successfully applied in diverse hydrological contexts, including urban rivers and Rural [58], and their use extends to the prediction of water quality indices in multivariable and complex scenarios [188]. Moreover, their ability to handle large volumes of data and correlated variables characterizes them as robust tools in environmental studies.

On the other hand, models based on ANNs have evolved considerably in recent years, incorporating more flexible architectures and optimization techniques such as genetic algorithms, wavelet transforms and hybrid strategies inspired by nature, improving performance against nonlinear and highly noisy datasets [107,159]. Its applicability has extended from the prediction of individual parameters to multivariable estimates of water quality, including dissolved oxygen and organic pollutants. Also, models based on fuzzy logic have gained relevance due to their ability to handle uncertainty (inherent in environmental data). Indeed, research in Saudi Arabia, India, and China has shown that fuzzy logic, when integrated with neural networks or evolutionary algorithms, exhibits suitable performances in regions with fragmented or scattered data [98,189,190]. Techniques such as Extreme Learning Machine (ELM), in combination with RF, have also been used to extend their applicability in environments with high hydrological variability [125].

Additionally, algorithms such as SVM, SVR, and RF remain highly popular due to their effectiveness in estimating water quality parameters. These methods offer non-parametric solutions that learn directly from observed data, facilitating uncertainty management and contextualized interpretation of pollution patterns [27].

The bibliometric analysis also reflects a sustained increase in integrating XAI techniques in research oriented to water quality prediction, to interpret interpreting the internal behavior of complex models and facilitating transparent decision-making. Techniques such as SHAP and LIME have been implemented to overcome the "black box" of ML/DL algorithms, allowing researchers to assess and visualize the influence of each predictor variable on the results [83,105,191]. XAI has been shown to improve model reliability and provide useful explanations for water quality management [106,192,193]. A prominent example is [194], in which an interpretable learning framework based on

SHAP and RF is developed, applied to hydrodynamic scenarios. This approach allowed us to understand the impact of environmental variables on water quality, reinforcing the usefulness of XAI in complex studies of aquatic systems.

Figure 9 (a) summarizes the predictive models identified in this study, classifying and organizing the AI/ML/DL algorithms most commonly used in freshwater quality prediction studies, both for surface water and groundwater systems. The percentage of the models with the highest applicability in predicting water quality in both surface and groundwater is presented in Figure 9(b).

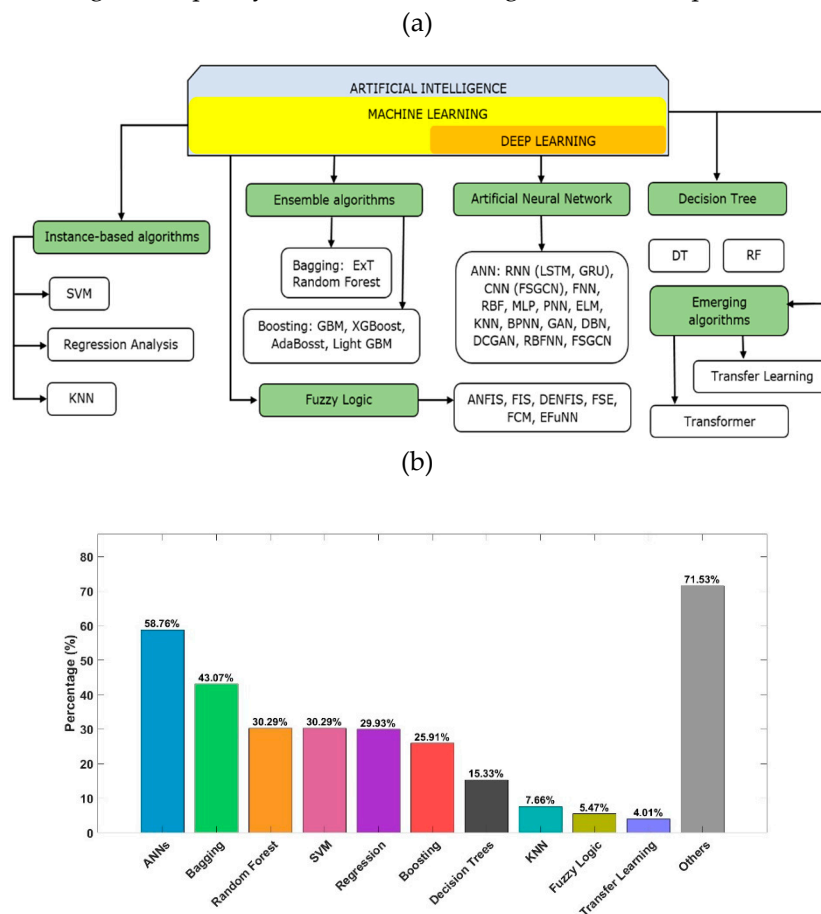


Figure 9. ML/DL models most used in water quality prediction: (a) Classification; (b) Percentage of applicability.

Finally, building reliable predictive models involves following a rigorous workflow that includes database consolidation, raw data preprocessing, proper predictive algorithm selection, model training, and validation. Recent literature underscores that each stage is critical to ensure the accuracy and robustness of the predictive model [48,195–197]. In particular, the selection of the algorithm must be aligned with the nature of the data and the specific objectives of a given study, ensuring interpretable, adaptable and relevant results for environmental decision-making.

RQ2. Which AI/ML/DL algorithm allows a better estimate of water quality?

The literature review identified a wide range of AI/ML/DL models developed both as standalone approaches and hybrid frameworks to enhance water quality prediction. Defining a single robust predictive model remains a challenge for researchers. However, certain algorithms stand out for their efficiency. For instance, the Light Gradient Boosting Machine (LightGBM) has emerged as a highly effective option due to its ability to process large datasets, fast training speed, and optimized architecture designed to minimize computational cost while maximizing predictive accuracy. Recent studies have reported accuracies exceeding 90% in the estimation of physicochemical parameters [198–200]. This technique is particularly notable for its consistent performance in water bodies with high variability.

The XGB algorithm has demonstrated remarkable performance in classifying both surface and groundwater quality, achieving accuracy levels close to 89% [201,202]. Its integration with XAI techniques such as SHAP enables the interpretation of key indicators such as zinc, nitrates, and chlorides, thereby improving the transparency of predictive models [201].

Time-series-based models, particularly LSTM architectures and their hybrid variants such as LSTM-CNN and CEEMDAN-LSTM, have achieved accuracies ranging from 90% to 93% by capturing complex dynamic patterns [6,44,86,114,153,203]. The incorporation of Transfer Learning further enhances performance by enabling efficient adaptation under data limitations [99,101,102]. Although MLP models have a simpler structure, they yield accuracies between 85% and 89% in sequential scenarios. Optimizing their parameters through genetic algorithms has improved precision in hydrologically realistic contexts [195,204–206].

Finally, hybrid approaches that combine multiple algorithms have reached accuracies above 92%, standing out for their adaptability to outlier values and their generalization capacity [155,207]. In summary, although no single model consistently outperforms across all contexts, algorithms that integrate interpretability, deep architecture, and hybrid strategies have delivered more accurate and reliable results. The selection of the most suitable model depends on the type of data, the analytical objectives, and the specific conditions of the water system under study.

RQ3. What limitations have been identified in the use of AI/ML/DL for water quality prediction?

A relevant aspect identified in this work is the study of spatial and temporal variations in the evaluation of water quality in the context of missing data, primarily due to missing data caused by measurement system failures, operational errors, environmental phenomena, and the non-continuous sampling frequency of water quality data. These limitations constrain the availability of reliable datasets for classification and evaluation purposes. In this context, Time Series Analysis (TSA) models combined with machine learning approaches, particularly Long Short-Term Memory (LSTM) networks have proven highly effective in addressing these challenges and estimating future values of water quality parameters (WQP) or water quality indices (WQI) based on historical data [208–210].

It was observed that DL models face various challenges compared to traditional physical models for water quality prediction. These include the complexity of internal structure and parameter adjustment, reliance on large data sets for effective training, and a lack of physical constraints, which can make it difficult to explain prediction results. Similarly, obtaining high-reliability data for certain water quality parameters can be difficult, limiting the applicability of deep learning approaches [211]. However, these challenges have begun to be overcome with the development of hybrid models and the use of interpretable approaches [26,86,112,212,213]. In general, the accuracy of the predictions of AI/ML/DL models is influenced by the availability and quality of historical data, and the models developed can be sensitive to variations in environmental conditions over time.

Recent literature has shown that some of these limitations can be mitigated through the use of LSTM models combined with Transfer Learning (TL) techniques [99,101], particularly instance-based approaches such as TrAdaBoost [94]. This model inherits the strengths of both LSTM and TL, offering powerful capabilities to capture long-term dependencies in time series and the flexibility to leverage related knowledge from complete datasets to fill large-scale consecutive data gaps [94]. Notably, prediction performance can be further enhanced by applying wavelet transforms to suppress noise in time-series signals, serving as an optimization mechanism for predictive modeling [87]. In this regard, model performance depends heavily on how variables are cleaned, transformed, and selected. For this reason, optimization algorithms such as Particle Swarm Optimization (PSO) [214,215], among others [112,116,216], are widely used.

Finally, variations in model performance caused by seasonal changes, extreme events, or point-source contamination which can significantly affect prediction accuracy have been addressed using Generative Adversarial Networks (GANs). These networks enable the simulation of abnormal or extreme conditions that are poorly represented in real datasets [161].

RQ4. What emerging variants currently exist in AI/ML/DL models for estimating water quality?

A current trend in water resource management using AI/ML/DL is the development of hybrid algorithms, which have gained relevance as a strategy to improve the accuracy of water quality parameter estimation. For instance, models based on variational mode decomposition optimized by the sparrow search algorithm (SSA-VMD), combined with Bidirectional Gated Recurrent Units (BiGRU), have achieved over 96% efficiency in the case of Qiandao Lake, China [217]. These hybrid approaches enable the capture of both nonlinear patterns and temporal dynamics in hydro-environmental data.

Besides, the Transformer architecture, has recently gained significant attention in deep learning research due to its superior performance over state-of-the-art LSTM models in time series forecasting and prediction tasks [27]. Finally, there is growing interest in enhancing the interpretability of machine learning models, given their “black box” nature. In this regard, Explainable AI (XAI) techniques have been implemented to identify the relative impact of water quality parameters on model predictions. Recent literature highlights successful applications of SHAP in models estimating salinity, dissolved oxygen, and the prediction of heavy metal concentrations, contributing to more informed decision-making in water management [162,163,218].

RQ5. What are the key water quality indicators used to assess natural water sources?

Water quality assessment in natural sources relies on a set of physicochemical and biological parameters that characterize both the environmental status and the suitability of water for various uses. According to the scientific literature reviewed, the most frequently used parameters in predictive studies employing AI/ML/DL techniques include dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), temperature, pH, electrical conductivity (EC), chlorophyll-a (Chl-a), nitrates, phosphates, and coliform bacteria. These indicators are considered conventional and are regulated by international environmental standards due to their relevance in characterizing both surface and groundwater bodies.

Beyond these essential parameters, there is a growing trend in the literature toward incorporating complementary variables such as heavy metals (e.g., arsenic, copper, lead), nutrients (NO_3^- , NO_2^- , NH_4^+ , PO_4^{3-}), and trace metals (Fe, Mn, Zn, Cu, Cr), which enhance the discrimination between different water quality categories. Additionally, the integration of hydrogeological, meteorological, land use, and socioeconomic variables has proven valuable in enriching predictive models by capturing the influence of external factors on water body dynamics [127,219,220]. In this regard, meteorological, land use, socioeconomic, and hydrogeological variables [221], for example, help illustrate how human activities can alter the export of chemical elements through changes in vegetation cover ultimately affecting water quality [222].

5. Contribution and Future Work

The review of the scientific literature performed in this work for the 2015-2024 period, through the integrated methodology of bibliometric and systematic review (B-SLR), delved into the AI/ML/DL techniques for the prediction of freshwater quality (surface water, groundwater). It also allowed us to identify key authors in the field, research gaps, emerging trends and models with high predictive performance. The main contribution of this study lies in its methodological approach that allowed identifying the most relevant works applying AI/ML/DL models in both surface waters and groundwater quality-related studies. It highlights an evolution of the models from traditional approaches such as ANN to hybrid and explainable architectures such as XGB, LSTM-CNN and SHAP.

The results of this research confirmed the findings of the study [58], which has received over 452 citations and reviewed predictive models for river water quality within the 2000–2020 timeframe. However, by applying the B-SLR methodology, the present study significantly expanded the scope of analysis—updating the literature to 2024, identifying emerging trends, and incorporating studies on freshwater groundwater as a potential source of drinking water. This broader and more integrative approach enabled a deeper thematic mapping of the field, revealing underexplored areas

and offering a more comprehensive understanding of current scientific developments. Regarding groundwater quality, the study [107] of the time frame (1994-2022) highlighted that ANNs are the most used in water quality modeling and that nitrate is the most studied parameter. Complementing these insights, the present work identifies a transition from more traditional approaches such as ANNs to more sophisticated models such as XGB, LightGB, and hybrid models with LSTM and CNN. Finally, and as future research, we propose to expand the spectrum of water quality parameters considered in predictive models, incorporating emerging pollutants such as microplastics, pharmaceuticals and persistent organic compounds which are of current concern [193] in water management.

6. Conclusions

This research demonstrates that the Bibliometric-Systematic Literature Review (B-SLR) approach constitutes a robust methodology for analyzing the state of the art in highly dynamic scientific fields, such as water quality prediction using AI/ML/DL techniques. By integrating the structured rigor of systematic review with the analytical depth of bibliometric analysis, the B-SLR enabled the identification of domain-specific trends, thematic mapping of knowledge, assessment of scientific impact, and detection of gaps in the literature—offering a more comprehensive, precise, and context-aware understanding of the field.

The findings reveal that ensemble models (e.g., Bagging, Boosting), deep neural networks (LSTM, CNN, MLP), and hybrid approaches have overcome the limitations of conventional methods, delivering greater accuracy, adaptability, and the ability to handle incomplete or nonlinear data. The integration of explainable artificial intelligence (XAI) techniques, such as SHAP and LIME, has facilitated the development of more transparent and reliable models, enhanced result interpretation and supporting informed decision-making. Accordingly, the selection of the optimal predictive model depends on multiple factors, including the type of water body, geographic context, data availability, and the specific objectives of the study. In this regard, hybrid and interpretable models emerge as the most promising alternatives for addressing current challenges in water quality prediction.

Regarding the methodological approach employed in this study, the application of B-SLR enabled the refinement of an initial database of 1,822 articles into a final corpus of 274 highly relevant publications, through automated procedures and manual validation. This process ensured the quality and relevance of the analyzed studies, reinforcing the reliability of the results obtained. Furthermore, a detailed classification was achieved, covering the most frequently used algorithms, the types of water bodies studied, key quality indicators, and the methodological limitations faced by predictive models. Ultimately, this work establishes a replicable and scalable methodological foundation for future research.

Author Contributions: Conceptualization, J.A.M.A., J.N., R.O., C.A.H., J.L.A., and L.R.L.; methodology, J.A.M.A.; software, J.A.M.A. and J.N.; validation, J.A.M.A. and J.N.; formal analysis, J.A.M.A., J.N., R.O., C.A.H., J.L.A., and L.R.L.; investigation, J.A.M.A., J.N. and R.O.; resources, J.N. and R.O.; data curation, J.A.M.A. and J.N.; writing—original draft preparation, J.A.M.A., J.N. and R.O.; writing—review and editing, J.A.M.A., J.N., R.O., C.A.H., J.L.A., and L.R.L.; visualization, J.A.M.A. and J.N.; supervision, J.N. and R.O.; project administration, R.O.; funding acquisition R.O., J.N. and J.L.A.

Funding: This research was supported by DIDULS Regular PR2553851 Project of the University of La Serena and ANID/FONDAP/1523A0001. The APC was sponsored by the CRHIAM Water Center, Universidad de Concepción, Chillán, Chile.

Data Availability Statement: This work has been developed upon public databases as described in the manuscript.

Acknowledgments: Ricardo Oyarzún and Jorge Núñez acknowledges the financial support of DIDULS/ULS, through the project PR2553851 (University of La Serena, Chile). José Luis Arumí and Ricardo Oyarzún acknowledges the financial support of the Water Research Center CRHIAM: ANID/FONDAP/1523A0001.

Conflicts of Interest: “The authors declare no conflicts of interest.”

Abbreviations

The following abbreviations are used in this manuscript:

AMT	Alternating Model Tree
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANFIS-GP	Adaptive Neuro Fuzzy Inference System – Grid Partitioning
ANFIS-SC	ANFIS – Subtractive Clustering
ANN	Artificial Neural Network
AO-SVM	Aquila Optimization Support Vector Machine
AR	Additive Regression
AdaBoost	Adaptive Boosting
BDT	Boosted Decision Tree
BiGRU	Bi-directional Gated Recurrent Units
BMEF	Bayesian Maximum Entropy-based Fusion
BNN	Bayesian Neural Network
BPNN	Backpropagation Neural Network
CART	Classification and Regression Tree
CatBoost	Categorical Boosting
CEEMD	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Network
CSA	Crow Search Algorithm
DBN	Deep Belief Network
DCGAN	Deep Convolutional Generative Adversarial Network
DENFIS	Dynamic Evolving Neural-Fuzzy Inference System
DNN	Deep Neural Network
DR	Discretization Regression
DRNN	Deep Recurrent Neural Network
DT	Decision Tree
DWT	Discrete Wavelet Transform
EANN	Emotional Artificial Neural Network
EANN-GA	Emotional Artificial Neural Network – Genetic Algorithm

EBM	Ensemble Bagged Machine
EFuNN	Evolving Fuzzy Neural Network
ELN	Extreme Learning Machine
EN	Elastic Network
ET	Extra Tree Regression
EWQI	Entropy-weighted Water Quality Index
ExT	Extra Trees
FFNNs	Feedforward Neural Networks
FNN	Feed-forward Neural Network
FSGCN	Functional-Structural Sub-Region Graph Convolutional Network
FFA	Firefly Algorithm
GAN	Generative Adversarial Network
GB	Gradient Boosting
GBM	Gradient Boosting Machine
GBR	Gradient Boosting Regression
GBT	Gradient Boosted Trees
GEP	Gene Expression Programming
GMDH	Group Method of Data Handling
GNB	Gaussian Naïve Bayes
GPR	Gaussian Process Regression
GRNN	Generalized Regression Neural Network
GRU	Gated Recurrent Unit
GS-RF	Grid Search Random Forest
GS-SVR	Grid Search Support Vector Regression
GWQI	Groundwater Quality Index
HGB	Histogram Gradient Boosting
IABC-BP	Improved Artificial Bee Colony – Backpropagation
IWQI	Irrigation Water Quality Index
KNN	K-Nearest Neighbours
LIME	Local Interpretable Model-agnostic Explanations
LR	Logistic Regression
LSSVR	Least Squares Support Vector Regression
LSTM	Long Short-Term Memory
LightGBM	Light Gradient Boosting Machine
MARS	Multivariate Adaptive Regression Spline
MLR	Multiple Linear Regression
MLRF	Multi-label Classification Through Random Forest

MLP	Multi-Layer Perceptron
MnLR	Multinomial Logistic Regression
NNE	Neural Network Ensemble
PLS	Partial Least Squares
PNN	Probabilistic Neural Network
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
RBFNN	Radial Basis Function Neural Network
RC	Random Committee
REPT	Reduced Error Pruning Tree
RF	Random Forest
RFC	Randomizable Filtered Classification
RNN	Recurrent Neural Network
RR	Ridge Regression
SDGs	Sustainable Development Goals
SHAP	SHapley Additive exPlanations
SLR	Simple Linear Regression
SMO-SVM	Sequential Minimal Optimization - Support Vector Machine
SSA-CNN-LSTM	Sparrow Search Algorithm - Convolutional Neural Network - Long Short-Term Memory
SVM	Support Vector Machines
SVR	Support Vector Regression
SVMR	Support Vector Machine Regression
SWEBM	Stochastic Weighted Ensemble Bagged Machine
TDS	Total Dissolved Solids
TL	Transfer Learning
WA	Wavelet Analysis
W-MGGP	Wavelet-Multigene Genetic Programming
WQI	Water Quality Index
WQP	Water Quality Parameters
WT	Wavelet Transform
XAI	eXplainable Artificial Intelligence
XGB	eXtreme Gradient Boosting

References

1. W. Ahmed, S. Mohammed, A. El-Shazly, and S. Morsy, "Tigris River water surface quality monitoring using remote sensing data and GIS techniques," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 26, no. 3, pp. 816–825, 2023, doi: <https://doi.org/10.1016/j.ejrs.2023.09.001>.
2. A. Gaagai *et al.*, "Application of Water Quality Indices, Machine Learning Approaches, and GIS to Identify Groundwater Quality for Irrigation Purposes: A Case Study of Sahara Aquifer, Doucen Plain, Algeria," *Water*, vol. 15, no. 2, 2023, doi: 10.3390/w15020289.
3. M. H. Rahaman, H. Sajjad, S. Hussain, Roshani, M. Masroor, and A. Sharma, "Surface water quality prediction in the lower Thoubal river watershed, India: A hyper-tuned machine learning approach and DNN-based sensitivity analysis," *J. Environ. Chem. Eng.*, vol. 12, no. 3, p. 112915, 2024, doi: <https://doi.org/10.1016/j.jece.2024.112915>.
4. ONU, *Informe mundial de las Naciones Unidas sobre el desarrollo de los recursos hídricos 2020: agua y cambio climático*. Ciudad de México, 2020. [Online]. Available: <https://es.unesco.org/themes/water-security/wwap/wwdr/2020>
5. J. Gao *et al.*, "Bibliometric analysis of climate change and water quality," *Hydrobiologia*, vol. 850, no. 16, pp. 3441–3459, 2023, doi: 10.1007/s10750-023-05270-y.
6. J. Pyo *et al.*, "Long short-term memory models of water quality in inland water environments," *Water Res. X*, vol. 21, p. 100207, 2023, doi: <https://doi.org/10.1016/j.wroa.2023.100207>.
7. E. E. Hussein, M. Y. Jat Baloch, A. Nigar, H. F. Abualkhair, F. K. Aldawood, and E. Tageldin, "Machine Learning Algorithms for Predicting the Water Quality Index," *Water*, vol. 15, no. 20, 2023, doi: 10.3390/w15203540.
8. N. Ahmad Kamal, N. S. Muhammad, and J. Abdullah, "Scenario-based pollution discharge simulations and mapping using integrated QUAL2K-GIS," *Environ. Pollut.*, vol. 259, p. 113909, 2020, doi: <https://doi.org/10.1016/j.envpol.2020.113909>.
9. S. K. Mummdivarapu, S. Rehana, and Y. R. S. Rao, "Mapping and assessment of river water quality under varying hydro-climatic and pollution scenarios by integrating QUAL2K, GEFC, and GIS," *Environ. Res.*, vol. 239, p. 117250, 2023, doi: <https://doi.org/10.1016/j.envres.2023.117250>.
10. J. Sarafaraz, F. Ahmadzadeh Kaleybar, J. Mahmoudi Karamjavan, and N. Habibzadeh, "Predicting river water quality: An imposing engagement between machine learning and the QUAL2Kw models (case study: Aji-Chai, river, Iran)," *Results Eng.*, vol. 21, p. 101921, 2024, doi: <https://doi.org/10.1016/j.rineng.2024.101921>.
11. Y.-Y. Chueh, C. Fan, and Y.-Z. Huang, "Copper concentration simulation in a river by SWAT-WASP integration and its application to assessing the impacts of climate change and various remediation strategies," *J. Environ. Manage.*, vol. 279, p. 111613, 2021, doi: <https://doi.org/10.1016/j.jenvman.2020.111613>.
12. S. Prajapati, P. Sabokruhie, M. Brinkmann, and K.-E. Lindenschmidt, "Modelling Transport and Fate of Copper and Nickel across the South Saskatchewan River Using WASP–TOXI," *Water (Switzerland)*, vol. 15, no. 2, 2023, doi: 10.3390/w15020265.
13. R. Alam, Z. Ahmed, S. M. Seefat, and K. T. K. Nahin, "Assessment of surface water quality around a landfill using multivariate statistical method, Sylhet, Bangladesh," *Environ. Nanotechnology, Monit. Manag.*, vol. 15, p. 100422, 2021, doi: <https://doi.org/10.1016/j.enmm.2020.100422>.
14. R. Isaac *et al.*, "Assessment of seasonal impacts on Water Quality in Yamuna river using Water Quality Index and Multivariate Statistical approaches," *Waste Manag. Bull.*, vol. 2, no. 3, pp. 145–153, 2024, doi: <https://doi.org/10.1016/j.wmb.2024.07.006>.
15. A. C. P. Fernandes, A. R. Fonseca, F. A. L. Pacheco, and L. F. Sanches Fernandes, "Water quality predictions through linear regression - A brute force algorithm approach," *MethodsX*, vol. 10, p. 102153, 2023, doi: <https://doi.org/10.1016/j.mex.2023.102153>.
16. M. Galoie, A. Motamedi, J. Fan, and M. Moudi, "Prediction of water quality under the impacts of fine dust and sand storm events using an experimental model and multivariate regression analysis," *Environ. Pollut.*, vol. 336, p. 122462, 2023, doi: <https://doi.org/10.1016/j.envpol.2023.122462>.
17. D. K. Pandey, A. I. Hunjra, R. Bhaskar, and M. A. S. Al-Faryan, "Artificial intelligence, machine learning and big data in natural resources management: A comprehensive bibliometric review of literature spanning 1975–2022," *Resour. Policy*, vol. 86, p. 104250, 2023, doi: <https://doi.org/10.1016/j.resourpol.2023.104250>.

18. X. Li *et al.*, "Bibliometric analysis of artificial intelligence in wastewater treatment: Current status, research progress, and future prospects," *J. Environ. Chem. Eng.*, vol. 12, no. 4, p. 113152, 2024, doi: <https://doi.org/10.1016/j.jece.2024.113152>.
19. C. Gonzales-Inca *et al.*, "Geospatial Artificial Intelligence (GeoAI) in the Integrated Hydrological and Fluvial Systems Modeling: Review of Current Applications and Trends," *Water*, vol. 14, no. 14, 2022, doi: 10.3390/w14142211.
20. J. Park, J. Ahn, J. Kim, Y. Yoon, and J. Park, "Prediction and Interpretation of Water Quality Recovery after a Disturbance in a Water Treatment System Using Artificial Intelligence," *Water*, vol. 14, no. 15, 2022, doi: 10.3390/w14152423.
21. M. Aliaga-Alvarado, V. Gómez-Escalonilla, and P. Martínez-Santos, "Identification of non-conventional groundwater resources by means of machine learning in the Aconcagua basin, Chile," *J. Hydrol. Reg. Stud.*, vol. 49, p. 101502, 2023, doi: <https://doi.org/10.1016/j.ejrh.2023.101502>.
22. D. N. Khoi, N. T. Quan, D. Q. Linh, P. T. T. Nhi, and N. T. D. Thuy, "Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam," *Water*, vol. 14, no. 10, 2022, doi: 10.3390/w14101552.
23. G. Mallya, M. M. Hantush, and R. S. Govindaraju, "A Machine Learning Approach to Predict Watershed Health Indices for Sediments and Nutrients at Ungauged Basins," *Water*, vol. 15, no. 3, 2023, doi: 10.3390/w15030586.
24. F. Ghobadi and D. Kang, "Application of Machine Learning in Water Resources Management: A Systematic Literature Review," *Water*, vol. 15, no. 4, 2023, doi: 10.3390/w15040620.
25. J. Yan, Y. Gao, Y. Yu, H. Xu, and Z. Xu, "A Prediction Model Based on Deep Belief Network and Least Squares SVR Applied to Cross-Section Water Quality," *Water*, vol. 12, no. 7, 2020, doi: 10.3390/w12071929.
26. Y. Zhou, "Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques," *J. Hydrol.*, vol. 589, p. 125164, 2020, doi: <https://doi.org/10.1016/j.jhydrol.2020.125164>.
27. K. P. Tripathy and A. K. Mishra, "Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions," *J. Hydrol.*, vol. 628, p. 130458, 2024, doi: <https://doi.org/10.1016/j.jhydrol.2023.130458>.
28. Y. Zheng, J. Wei, W. Zhang, Y. Zhang, T. Zhang, and Y. Zhou, "An ensemble model for accurate prediction of key water quality parameters in river based on deep learning methods," *J. Environ. Manage.*, vol. 366, p. 121932, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.121932>.
29. C. Chellaiah *et al.*, "Integrating deep learning techniques for effective river water quality monitoring and management," *J. Environ. Manage.*, vol. 370, p. 122477, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.122477>.
30. D. V. V. Prasad *et al.*, "Analysis and prediction of water quality using deep learning and auto deep learning techniques," *Sci. Total Environ.*, vol. 821, p. 153311, 2022, doi: <https://doi.org/10.1016/j.scitotenv.2022.153311>.
31. S. Khullar and N. Singh, "Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation," *Environ. Sci. Pollut. Res.*, vol. 29, no. 9, pp. 12875–12889, 2022, doi: 10.1007/s11356-021-13875-w.
32. United Nations Department of Economic and Social Affairs, "The Sustainable Development Goals Report 2023: Special Edition.," New York, 2023. [Online]. Available: <https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf>
33. UN-Water, "The Sustainable Development Goal 6 Global Acceleration Framework," Geneva, Switzerland, 2020. [Online]. Available: <https://unsceb.org/sdg-6-global-acceleration-framework>
34. M. de Carvalho Marques, A. A. Mohamed, and P. Feitosa, "Sustainable development goal 6 monitoring through statistical machine learning – Random Forest method," *Clean. Prod. Lett.*, vol. 8, p. 100088, 2025, doi: <https://doi.org/10.1016/j.clpl.2024.100088>.
35. L. Rodríguez-López *et al.*, "Machine Learning Algorithms for the Estimation of Water Quality Parameters in Lake Llanquihue in Southern Chile," *Water*, vol. 15, no. 11, 2023, doi: 10.3390/w15111994.

36. A. T. Ansari, N. Nigar, H. M. Faisal, and M. K. Shahzad, "AI for clean water: efficient water quality prediction leveraging machine learning," *Water Pract. Technol.*, vol. 19, no. 5, pp. 1986–1996, 2024, doi: 10.2166/wpt.2024.120.
37. M. G. Zamani, M. R. Nikoo, F. Niknazar, G. Al-Rawas, M. Al-Wardy, and A. H. Gandomi, "A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy," *J. Clean. Prod.*, vol. 416, p. 137885, 2023, doi: <https://doi.org/10.1016/j.jclepro.2023.137885>.
38. A. H. Haghiabi, A. H. Nasrolahi, and A. Parsaie, "Water quality prediction using machine learning methods," *Water Qual. Res. J.*, vol. 53, no. 1, pp. 3–13, 2018, doi: 10.2166/wqrj.2018.025.
39. N. Satish, J. Anmala, M. R. R. Varma, and K. Rajitha, "Performance of Machine Learning, Artificial Neural Network (ANN), and stacked ensemble models in predicting Water Quality Index (WQI) from surface water quality parameters, climatic and land use data," *Process Saf. Environ. Prot.*, vol. 192, pp. 177–195, 2024, doi: <https://doi.org/10.1016/j.psep.2024.10.054>.
40. S. Z. Farzana, D. R. Paudyal, S. Chadalavada, and M. J. Alam, "Temporal Dynamics and Predictive Modelling of Streamflow and Water Quality Using Advanced Statistical and Ensemble Machine Learning Techniques," *Water*, vol. 16, no. 15, 2024, doi: 10.3390/w16152107.
41. M. G. Uddin, S. Nash, M. T. Mahammad Diganta, A. Rahman, and A. I. Olbert, "Robust machine learning algorithms for predicting coastal water quality index," *J. Environ. Manage.*, vol. 321, p. 115923, 2022, doi: <https://doi.org/10.1016/j.jenvman.2022.115923>.
42. I. I. S. Shamsuddin, Z. Othman, and N. S. Sani, "Water Quality Index Classification Based on Machine Learning: A Case from the Langat River Basin Model," *Water*, vol. 14, no. 19, 2022, doi: 10.3390/w14192939.
43. S. Singh, A. Das, and P. Sharma, "Predictive modeling of water quality index (WQI) classes in Indian rivers: Insights from the application of multiple Machine Learning (ML) models on a decennial dataset," *Stoch. Environ. Res. Risk Assess.*, vol. 38, no. 8, pp. 3221 – 3238, 2024, doi: 10.1007/s00477-024-02741-z.
44. J. Yao, S. Chen, and X. Ruan, "Interpretable CEEMDAN-FE-LSTM-transformer hybrid model for predicting total phosphorus concentrations in surface water," *J. Hydrol.*, vol. 629, p. 130609, 2024, doi: <https://doi.org/10.1016/j.jhydrol.2024.130609>.
45. V. Flores, I. Bravo, and M. Saavedra, "Water Quality Classification and Machine Learning Model for Predicting Water Quality Status—A Study on Loa River Located in an Extremely Arid Environment: Atacama Desert," *Water*, vol. 15, no. 16, 2023, doi: 10.3390/w15162868.
46. A. Masood, M. Niazkar, M. Zakwan, and R. Piraei, "A Machine Learning-Based Framework for Water Quality Index Estimation in the Southern Bug River," *Water*, vol. 15, no. 20, 2023, doi: 10.3390/w15203543.
47. C. D. Aju, A. L. Achu, M. P. Mohammed, M. C. Raicy, G. Gopinath, and R. Reghunath, "Groundwater quality prediction and risk assessment in Kerala, India: A machine-learning approach," *J. Environ. Manage.*, vol. 370, p. 122616, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.122616>.
48. M. Zhu *et al.*, "A review of the application of machine learning in water quality evaluation," *Eco-Environment Heal.*, vol. 1, no. 2, pp. 107–116, 2022, doi: <https://doi.org/10.1016/j.eehl.2022.06.001>.
49. A. F. del Castillo *et al.*, "Improving river water quality prediction with hybrid machine learning and temporal analysis," *Ecol. Inform.*, vol. 82, p. 102655, 2024, doi: <https://doi.org/10.1016/j.ecoinf.2024.102655>.
50. T. Yan, A. Zhou, and S.-L. Shen, "Prediction of long-term water quality using machine learning enhanced by Bayesian optimisation," *Environ. Pollut.*, vol. 318, p. 120870, 2023, doi: <https://doi.org/10.1016/j.envpol.2022.120870>.
51. C. Liu, J. Xu, X. Li, Z. Yu, and J. Wu, "Water resource forecasting with machine learning and deep learning: A scientometric analysis," *Artif. Intell. Geosci.*, vol. 5, p. 100084, 2024, doi: <https://doi.org/10.1016/j.aiig.2024.100084>.
52. P. Jayaraman, K. K. Nagarajan, P. Partheeban, and V. Krishnamurthy, "Critical review on water quality analysis using IoT and machine learning models," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 1, p. 100210, 2024, doi: <https://doi.org/10.1016/j.jjime.2023.100210>.
53. W. Li, Y. Zhao, Y. Zhu, Z. Dong, F. Wang, and F. Huang, "Research progress in water quality prediction based on deep learning technology: a review," *Environ. Sci. Pollut. Res.*, vol. 31, no. 18, pp. 26415–26431, 2024, doi: 10.1007/s11356-024-33058-7.

54. N. F. Che Nordin, N. S. Mohd, S. Koting, Z. Ismail, M. Sherif, and A. El-Shafie, "Groundwater quality forecasting modelling using artificial intelligence: A review," *Groundw. Sustain. Dev.*, vol. 14, p. 100643, 2021, doi: <https://doi.org/10.1016/j.gsd.2021.100643>.
55. X. Li, Y. Li, and G. Li, "A scientometric review of the research on the impacts of climate change on water quality during 1998–2018," *Environ. Sci. Pollut. Res.*, vol. 27, no. 13, pp. 14322–14341, 2020, doi: [10.1007/s11356-020-08176-7](https://doi.org/10.1007/s11356-020-08176-7).
56. S. Bose, A. Mazumdar, and S. Basu, "Evolution of groundwater quality assessment on urban area- a bibliometric analysis," *Groundw. Sustain. Dev.*, vol. 20, p. 100894, 2023, doi: <https://doi.org/10.1016/j.gsd.2022.100894>.
57. N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021, doi: <https://doi.org/10.1016/j.jbusres.2021.04.070>.
58. Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," *J. Hydrol.*, vol. 585, p. 124670, 2020, doi: <https://doi.org/10.1016/j.jhydrol.2020.124670>.
59. S. Cojbasic *et al.*, "Application of machine learning in river water quality management: a review," *Water Sci. Technol.*, vol. 88, no. 9, pp. 2297–2308, 2023, doi: [10.2166/wst.2023.331](https://doi.org/10.2166/wst.2023.331).
60. B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009, doi: <https://doi.org/10.1016/j.infsof.2008.09.009>.
61. G. Marzi, M. Balzano, A. Caputo, and M. M. Pellegrini, "Guidelines for Bibliometric-Systematic Literature Reviews: 10 steps to combine analysis, synthesis and theory development," *Int. J. Manag. Rev.*, vol. n/a, no. n/a, doi: <https://doi.org/10.1111/ijmr.12381>.
62. W. M. Lim, S. Kumar, and N. Donthu, "How to combine and clean bibliometric data and use bibliometric tools synergistically: Guidelines using metaverse research," *J. Bus. Res.*, vol. 182, p. 114760, 2024, doi: <https://doi.org/10.1016/j.jbusres.2024.114760>.
63. M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, no. 4, pp. 959–975, 2017, doi: <https://doi.org/10.1016/j.joi.2017.08.007>.
64. M. A. Mohammed, C. De-Pablos-Heredero, and J. L. Montes Botella, "A Systematic Literature Review on the Revolutionary Impact of Blockchain in Modern Business," *Appl. Sci.*, vol. 14, no. 23, 2024, doi: [10.3390/app142311077](https://doi.org/10.3390/app142311077).
65. M. Lahami, A. Jmal Maalej, and M. Krichen, "A systematic literature review on dynamic testing of blockchain oriented software," *Sci. Comput. Program.*, vol. 240, p. 103211, 2025, doi: <https://doi.org/10.1016/j.scico.2024.103211>.
66. B. Z. Rousso, E. Bertone, R. Stewart, and D. P. Hamilton, "A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes," *Water Res.*, vol. 182, p. 115959, 2020, doi: <https://doi.org/10.1016/j.watres.2020.115959>.
67. J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quant. Sci. Stud.*, vol. 1, no. 1, pp. 377 – 386, 2020, doi: [10.1162/qss_a_00019](https://doi.org/10.1162/qss_a_00019).
68. A. O. Baarimah *et al.*, "Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis," *Case Stud. Chem. Environ. Eng.*, vol. 10, p. 100926, 2024, doi: <https://doi.org/10.1016/j.cscee.2024.100926>.
69. M. Gusenbauer, "Searchsmart.org: Guiding researchers to the best databases and search systems for systematic reviews and beyond," *Res. Synth. Methods*, vol. 15, no. 6, pp. 1200–1213, 2024, doi: <https://doi.org/10.1002/jrsm.1746>.
70. R. van Dinter, B. Tekinerdogan, and C. Catal, "Automation of systematic literature reviews: A systematic literature review," *Inf. Softw. Technol.*, vol. 136, p. 106589, 2021, doi: <https://doi.org/10.1016/j.infsof.2021.106589>.
71. CRAN, "TextmineR: Functions for Text Mining and Topic Modeling." <https://cran.r-project.org/package=textmineR>

72. A. Kassem, A. Sefelnasr, A. A. Ebraheem, and M. Sherif, "Seawater intrusion physical models: A bibliometric analysis and review of mitigation strategies," *J. Hydrol.*, vol. 634, p. 131135, 2024, doi: <https://doi.org/10.1016/j.jhydrol.2024.131135>.
73. Z. Phiri, N. T. Moja, T. T. I. Nkambule, and L.-A. de Kock, "Utilization of biochar for remediation of heavy metals in aqueous environments: A review and bibliometric analysis," *Heliyon*, vol. 10, no. 4, p. e25785, 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e25785>.
74. H. Liu *et al.*, "Impacts of green roofs on water, temperature, and air quality: A bibliometric review," *Build. Environ.*, vol. 196, p. 107794, 2021, doi: <https://doi.org/10.1016/j.buildenv.2021.107794>.
75. M. J. Biazatti, A. C. A. Justi, R. F. Souza, and J. C. de Carvalho Miranda, "Soybean biorefinery and technological forecasts based on a bibliometric analysis and network mapping," *Environ. Dev.*, vol. 52, p. 101074, 2024, doi: <https://doi.org/10.1016/j.envdev.2024.101074>.
76. V. Batagelj and M. Cerinšek, "On bibliographic networks," *Scientometrics*, vol. 96, no. 3, pp. 845–864, 2013, doi: [10.1007/s11192-012-0940-1](https://doi.org/10.1007/s11192-012-0940-1).
77. H. P. Pandey, T. N. Maraseni, and A. A. Apan, "Enhancing systematic literature review adapting 'double diamond approach,'" *Heliyon*, vol. 10, no. 22, p. e40581, 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e40581>.
78. M. Gusenbauer and S. P. Gauster, "How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide," *Technol. Forecast. Soc. Change*, vol. 212, p. 123833, 2025, doi: <https://doi.org/10.1016/j.techfore.2024.123833>.
79. K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, 2015, doi: <https://doi.org/10.1016/j.infsof.2015.03.007>.
80. M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," *Water Sci. Technol.*, vol. 82, no. 12, pp. 2635–2670, 2020, doi: [10.2166/wst.2020.369](https://doi.org/10.2166/wst.2020.369).
81. R. Huang, C. Ma, J. Ma, X. Huangfu, and Q. He, "Machine learning in natural and engineered water systems," *Water Res.*, vol. 205, p. 117666, 2021, doi: <https://doi.org/10.1016/j.watres.2021.117666>.
82. G. W. Tefera, R. L. Ray, and V. P. Singh, "Surface water quality under climate change scenarios in the Bosque watershed, Central Texas of United States," *Ecohydrol. Hydrobiol.*, 2024, doi: <https://doi.org/10.1016/j.ecohyd.2024.07.008>.
83. S. Ramya, S. Srinath, and P. Tuppad, "Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI," *Case Stud. Chem. Environ. Eng.*, vol. 10, 2024, doi: [10.1016/j.cscee.2024.100822](https://doi.org/10.1016/j.cscee.2024.100822).
84. L. M. Sidek *et al.*, "Developing an ensemble machine learning model for predicting water quality index in Johor River Basin," *Environ. Sci. Eur.*, vol. 36, no. 1, p. 67, 2024, doi: [10.1186/s12302-024-00897-7](https://doi.org/10.1186/s12302-024-00897-7).
85. J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Natl. Acad. Sci.*, vol. 102, no. 46, pp. 16569–16572, 2005, doi: [10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102).
86. R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model," *Stoch. Environ. Res. Risk Assess.*, vol. 34, no. 2, pp. 415–433, 2020, doi: [10.1007/s00477-020-01776-2](https://doi.org/10.1007/s00477-020-01776-2).
87. R. Barzegar, J. Adamowski, and A. A. Moghaddam, "Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran," *Stoch. Environ. Res. Risk Assess.*, vol. 30, no. 7, pp. 1797–1819, 2016, doi: [10.1007/s00477-016-1213-y](https://doi.org/10.1007/s00477-016-1213-y).
88. J. Y. Ho *et al.*, "Towards a time and cost effective approach to water quality index class prediction," *J. Hydrol.*, vol. 575, pp. 148–165, 2019, doi: <https://doi.org/10.1016/j.jhydrol.2019.05.016>.
89. X. Ji, X. Shang, R. A. Dahlgren, and M. Zhang, "Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China," *Environ. Sci. Pollut. Res.*, vol. 24, no. 19, pp. 16062–16076, 2017, doi: [10.1007/s11356-017-9243-7](https://doi.org/10.1007/s11356-017-9243-7).
90. E. Fijani, R. Barzegar, R. Deo, E. Tziritis, and K. Skordas, "Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time

- environmental monitoring of water quality parameters," *Sci. Total Environ.*, vol. 648, pp. 839–853, 2019, doi: <https://doi.org/10.1016/j.scitotenv.2018.08.221>.
91. S. I. Abba *et al.*, "Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index," *Environ. Sci. Pollut. Res.*, vol. 27, no. 33, pp. 41524–41539, 2020, doi: [10.1007/s11356-020-09689-x](https://doi.org/10.1007/s11356-020-09689-x).
 92. R. Noori, H.-D. Yeh, M. Abbasi, F. T. Kachosangi, and S. Moazami, "Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand," *J. Hydrol.*, vol. 527, pp. 833–843, 2015, doi: <https://doi.org/10.1016/j.jhydrol.2015.05.046>.
 93. B. Sakaa *et al.*, "Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin," *Environ. Sci. Pollut. Res.*, vol. 29, no. 32, pp. 48491–48508, 2022, doi: [10.1007/s11356-022-18644-x](https://doi.org/10.1007/s11356-022-18644-x).
 94. Z. Chen *et al.*, "A transfer Learning-Based LSTM strategy for imputing Large-Scale consecutive missing data and its application in a water quality prediction system," *J. Hydrol.*, vol. 602, p. 126573, 2021, doi: <https://doi.org/10.1016/j.jhydrol.2021.126573>.
 95. Z. Jamshidzadeh, M. Ehteram, and H. Shabaniyan, "Bidirectional Long Short-Term Memory (BiLSTM) - Support Vector Machine: A new machine learning model for predicting water quality parameters," *Ain Shams Eng. J.*, vol. 15, no. 3, p. 102510, 2024, doi: <https://doi.org/10.1016/j.asej.2023.102510>.
 96. T.-C. Chen, "Application of wavelet theory to enhance the performance of machine learning techniques in estimating water quality parameters (case study: Gao-Ping River)," *Water Sci. Technol.*, vol. 87, no. 5, pp. 1294–1315, 2023, doi: [10.2166/wst.2023.047](https://doi.org/10.2166/wst.2023.047).
 97. K. Wang, L. Liu, X. Ben, D. Jin, Y. Zhu, and F. Wang, "Hybrid deep learning based prediction for water quality of plain watershed," *Environ. Res.*, vol. 262, p. 119911, 2024, doi: <https://doi.org/10.1016/j.envres.2024.119911>.
 98. M. S. Manzar *et al.*, "New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable: A case study from Saudi Arabia," *Ecol. Inform.*, vol. 70, p. 101696, 2022, doi: <https://doi.org/10.1016/j.ecoinf.2022.101696>.
 99. X. Chen, W. Sun, T. Jiang, and H. Ju, "Enhanced prediction of river dissolved oxygen through feature- and model-based transfer learning," *J. Environ. Manage.*, vol. 372, p. 123310, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.123310>.
 100. K. Khodkar *et al.*, "Stream salinity prediction in data-scarce regions: Application of transfer learning and uncertainty quantification," *J. Contam. Hydrol.*, vol. 266, p. 104418, 2024, doi: <https://doi.org/10.1016/j.jconhyd.2024.104418>.
 101. S. Chen, J. Huang, P. Wang, X. Tang, and Z. Zhang, "A coupled model to improve river water quality prediction towards addressing non-stationarity and data limitation," *Water Res.*, vol. 248, p. 120895, 2024, doi: <https://doi.org/10.1016/j.watres.2023.120895>.
 102. L. Peng *et al.*, "TLT: Recurrent fine-tuning transfer learning for water quality long-term prediction," *Water Res.*, vol. 225, p. 119171, 2022, doi: <https://doi.org/10.1016/j.watres.2022.119171>.
 103. L. Longo *et al.*, "Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions," *Inf. Fusion*, vol. 106, p. 102301, 2024, doi: <https://doi.org/10.1016/j.inffus.2024.102301>.
 104. J. Núñez, C. B. Cortés, and M. A. Yáñez, "Explainable Artificial Intelligence in Hydrology: Interpreting Black-Box Snowmelt-Driven Streamflow Predictions in an Arid Andean Basin of North-Central Chile," *Water*, vol. 15, no. 19, 2023, doi: [10.3390/w15193369](https://doi.org/10.3390/w15193369).
 105. H. A. Madni *et al.*, "Water-Quality Prediction Based on H2O AutoML and Explainable AI Techniques," *Water*, vol. 15, no. 3, 2023, doi: [10.3390/w15030475](https://doi.org/10.3390/w15030475).
 106. F. Alshehri and A. Rahman, "Coupling Machine and Deep Learning with Explainable Artificial Intelligence for Improving Prediction of Groundwater Quality and Decision-Making in Arid Region, Saudi Arabia," *Water*, vol. 15, no. 12, 2023, doi: [10.3390/w15122298](https://doi.org/10.3390/w15122298).
 107. R. Haggerty, J. Sun, H. Yu, and Y. Li, "Application of machine learning in groundwater quality modeling - A comprehensive review," *Water Res.*, vol. 233, p. 119745, 2023, doi: <https://doi.org/10.1016/j.watres.2023.119745>.

108. H. Ibrahim *et al.*, "Evaluation and Prediction of Groundwater Quality for Irrigation Using an Integrated Water Quality Indices, Machine Learning Models and GIS Approaches: A Representative Case Study," *Water*, vol. 15, no. 4, 2023, doi: 10.3390/w15040694.
109. S. H. Rahat *et al.*, "Remote sensing-enabled machine learning for river water quality modeling under multidimensional uncertainty," *Sci. Total Environ.*, vol. 898, p. 165504, 2023, doi: <https://doi.org/10.1016/j.scitotenv.2023.165504>.
110. K. Huangfu, J. Li, X. Zhang, J. Zhang, H. Cui, and Q. Sun, "Remote Estimation of Water Quality Parameters of Medium- and Small-Sized Inland Rivers Using Sentinel-2 Imagery," *Water*, vol. 12, no. 11, 2020, doi: 10.3390/w12113124.
111. J. O'Grady, D. Zhang, N. O'Connor, and F. Regan, "A comprehensive review of catchment water quality monitoring using a tiered framework of integrated sensing technologies," *Sci. Total Environ.*, vol. 765, p. 142766, 2021, doi: <https://doi.org/10.1016/j.scitotenv.2020.142766>.
112. Y. Bai, M. Peng, and M. Wang, "A River Water Quality Prediction Method Based on Dual Signal Decomposition and Deep Learning," *Water*, vol. 16, no. 21, 2024, doi: 10.3390/w16213099.
113. A. Najah Ahmed *et al.*, "Machine learning methods for better water quality prediction," *J. Hydrol.*, vol. 578, p. 124084, 2019, doi: <https://doi.org/10.1016/j.jhydrol.2019.124084>.
114. M. Zhang, Z. Zhang, X. Wang, Z. Liao, and L. Wang, "The Use of Attention-Enhanced CNN-LSTM Models for Multi-Indicator and Time-Series Predictions of Surface Water Quality," *Water Resour. Manag.*, vol. 38, no. 15, pp. 6103–6119, 2024, doi: 10.1007/s11269-024-03946-1.
115. J. M. Sadler, L. E. Koenig, G. Gorski, A. M. Carter, and R. O. Hall Jr., "Evaluating a process-guided deep learning approach for predicting dissolved oxygen in streams," *Hydrol. Process.*, vol. 38, no. 9, p. e15270, 2024, doi: <https://doi.org/10.1002/hyp.15270>.
116. S. Rajagopal, S. S. Ganesh, A. Karthick, and T. Sampradeepraj, "Environmental water quality prediction based on COOT-CSO-LSTM deep learning," *Environ. Sci. Pollut. Res.*, vol. 31, no. 42, pp. 54525–54533, 2024, doi: 10.1007/s11356-024-34750-4.
117. M. Poursaeid, A. H. Poursaeed, and S. Shabanlou, "Water quality fluctuations prediction and Debi estimation based on stochastic optimized weighted ensemble learning machine," *Process Saf. Environ. Prot.*, vol. 188, pp. 1160–1174, 2024, doi: <https://doi.org/10.1016/j.psep.2024.05.146>.
118. V. J. K. K. Shanmugaiah, and F. J. J. Shobana Bai, "AO-SVM: a machine learning model for predicting water quality in the cauvery river," *Environ. Res. Commun.*, vol. 6, no. 7, p. 75025, Jul. 2024, doi: 10.1088/2515-7620/ad6061.
119. R. K. Poluru, S. Sundararajan, V. S. S. Balakrishnan, S. V, and M. Rajagopal, "Predicting nitrous oxide contaminants in Cauvery basin using region-based convolutional neural network," *Groundw. Sustain. Dev.*, vol. 26, p. 101194, 2024, doi: <https://doi.org/10.1016/j.gsd.2024.101194>.
120. Z. Lin, J. Y. Lim, and J.-M. Oh, "Innovative interpretable AI-guided water quality evaluation with risk adversarial analysis in river streams considering spatial-temporal effects," *Environ. Pollut.*, vol. 350, p. 124015, 2024, doi: <https://doi.org/10.1016/j.envpol.2024.124015>.
121. W. Liu *et al.*, "Analysis of dissolved oxygen influencing factors and concentration prediction using input variable selection technique: A hybrid machine learning approach," *J. Environ. Manage.*, vol. 357, p. 120777, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.120777>.
122. G. Saha, C. Shen, J. Duncan, and R. Cibin, "Performance evaluation of deep learning based stream nitrate concentration prediction model to fill stream nitrate data gaps at low-frequency nitrate monitoring basins," *J. Environ. Manage.*, vol. 357, p. 120721, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.120721>.
123. R. B. Singh, K. C. Patra, B. Pradhan, and A. Samantra, "HDTO-DeepAR: A novel hybrid approach to forecast surface water quality indicators," *J. Environ. Manage.*, vol. 352, p. 120091, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.120091>.
124. Y. Hu, C. Liu, and W. M. Wollheim, "Prediction of riverine daily minimum dissolved oxygen concentrations using hybrid deep learning and routine hydrometeorological data," *Sci. Total Environ.*, vol. 918, p. 170383, 2024, doi: <https://doi.org/10.1016/j.scitotenv.2024.170383>.

125. Z. Wang, Q. Wang, Z. Liu, and T. Wu, "A deep learning interpretable model for river dissolved oxygen multi-step and interval prediction based on multi-source data fusion," *J. Hydrol.*, vol. 629, p. 130637, 2024, doi: <https://doi.org/10.1016/j.jhydrol.2024.130637>.
126. M. Al-Mukhtar, A. Srivastava, L. Khadke, T. Al-Musawi, and A. Elbeltagi, "Prediction of Irrigation Water Quality Indices Using Random Committee, Discretization Regression, REPTree, and Additive Regression," *Water Resour. Manag.*, vol. 38, no. 1, pp. 343–368, 2024, doi: [10.1007/s11269-023-03674-y](https://doi.org/10.1007/s11269-023-03674-y).
127. B. E. S. Zhang, C. T. Driscoll, and T. Wen, "Human and natural impacts on the U.S. freshwater salinization and alkalinization: A machine learning approach," *Sci. Total Environ.*, vol. 889, p. 164138, 2023, doi: <https://doi.org/10.1016/j.scitotenv.2023.164138>.
128. L. Luo, Y. Zhang, W. Dong, J. Zhang, and L. Zhang, "Ensemble Empirical Mode Decomposition and a Long Short-Term Memory Neural Network for Surface Water Quality Prediction of the Xiaofu River, China," *Water*, vol. 15, no. 8, 2023, doi: [10.3390/w15081625](https://doi.org/10.3390/w15081625).
129. R. Xu *et al.*, "Feature Extraction and Prediction of Water Quality Based on Candlestick Theory and Deep Learning Methods," *Water*, vol. 15, no. 5, 2023, doi: [10.3390/w15050845](https://doi.org/10.3390/w15050845).
130. Y. Im, G. Song, J. Lee, and M. Cho, "Deep Learning Methods for Predicting Tap-Water Quality Time Series in South Korea," *Water*, vol. 14, no. 22, 2022, doi: [10.3390/w14223766](https://doi.org/10.3390/w14223766).
131. J. Mohd Zebaral Hoque, N. A. Ab. Aziz, S. Alelyani, M. Mohana, and M. Hosain, "Improving Water Quality Index Prediction Using Regression Learning Models," *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, 2022, doi: [10.3390/ijerph192013702](https://doi.org/10.3390/ijerph192013702).
132. D. Y. Dorado-Guerra, G. Corzo-Pérez, J. Paredes-Arquiola, and M. Á. Pérez-Martín, "Machine learning models to predict nitrate concentration in a river basin," *Environ. Res. Commun.*, vol. 4, no. 12, p. 125012, 2023, doi: [10.1088/2515-7620/acabb7](https://doi.org/10.1088/2515-7620/acabb7).
133. I. C. Adedeji, E. Ahmadisharaf, and Y. Sun, "Predicting in-stream water quality constituents at the watershed scale using machine learning," *J. Contam. Hydrol.*, vol. 251, p. 104078, 2022, doi: <https://doi.org/10.1016/j.jconhyd.2022.104078>.
134. K. Khosravi, A. Golkarian, A. M. Melesse, and R. C. Deo, "Suspended sediment load modeling using advanced hybrid rotation forest based elastic network approach," *J. Hydrol.*, vol. 610, p. 127963, 2022, doi: <https://doi.org/10.1016/j.jhydrol.2022.127963>.
135. C. Song and L. Yao, "A hybrid model for water quality parameter prediction based on CEEMDAN-IALO-LSTM ensemble learning," *Environ. Earth Sci.*, vol. 81, no. 9, p. 262, 2022, doi: [10.1007/s12665-022-10380-2](https://doi.org/10.1007/s12665-022-10380-2).
136. Y. Hou *et al.*, "A study on water quality parameters estimation for urban rivers based on ground hyperspectral remote sensing technology," *Environ. Sci. Pollut. Res.*, vol. 29, no. 42, pp. 63640–63654, 2022, doi: [10.1007/s11356-022-20293-z](https://doi.org/10.1007/s11356-022-20293-z).
137. T. Balson and A. S. Ward, "A machine learning approach to water quality forecasts and sensor network expansion: Case study in the Wabash River Basin, United States," *Hydrol. Process.*, vol. 36, no. 6, p. e14619, 2022, doi: <https://doi.org/10.1002/hyp.14619>.
138. N. H. A. Malek, W. F. Wan Yaacob, S. A. Md Nasir, and N. Shaadan, "Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques," *Water*, vol. 14, no. 7, 2022, doi: [10.3390/w14071067](https://doi.org/10.3390/w14071067).
139. N. N. M. Rizal, G. Hayder, and K. A. Yusof, "Water Quality Predictive Analytics Using an Artificial Neural Network with a Graphical User Interface," *Water*, vol. 14, no. 8, 2022, doi: [10.3390/w14081221](https://doi.org/10.3390/w14081221).
140. H. Weierbach *et al.*, "Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning," *Water*, vol. 14, no. 7, 2022, doi: [10.3390/w14071032](https://doi.org/10.3390/w14071032).
141. A. del Castillo, C. Yebra-Montes, M. Verduzco Garibay, J. de Anda, A. Garcia-Gonzalez, and M. S. Gradilla-Hernández, "Simple Prediction of an Ecosystem-Specific Water Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning," *Water*, vol. 14, no. 8, 2022, doi: [10.3390/w14081235](https://doi.org/10.3390/w14081235).
142. M. Ilić, Z. Srdjević, and B. Srdjević, "Water quality prediction based on Naïve Bayes algorithm," *Water Sci. Technol.*, vol. 85, no. 4, pp. 1027–1039, 2022, doi: [10.2166/wst.2022.006](https://doi.org/10.2166/wst.2022.006).
143. S. Arora and A. K. Keshari, "Dissolved oxygen modelling of the Yamuna River using different ANFIS models," *Water Sci. Technol.*, vol. 84, no. 10–11, pp. 3359–3371, 2021, doi: [10.2166/wst.2021.466](https://doi.org/10.2166/wst.2021.466).

144. S. V. Moghadam, A. Sharafati, H. Feizi, S. M. S. Marjaie, S. B. H. S. Asadollah, and D. Motta, "An efficient strategy for predicting river dissolved oxygen concentration: application of deep recurrent neural network model," *Environ. Monit. Assess.*, vol. 193, no. 12, p. 798, Nov. 2021, doi: 10.1007/s10661-021-09586-x.
145. C. Xu, X. Chen, and L. Zhang, "Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models," *J. Environ. Manage.*, vol. 295, p. 113085, Oct. 2021, doi: 10.1016/j.jenvman.2021.113085.
146. A. Najah *et al.*, "Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia," *Int. J. Environ. Sci. Technol.*, vol. 18, no. 4, pp. 1009–1018, 2021, doi: 10.1007/s13762-021-03139-y.
147. S. Kim *et al.*, "Assessment of the total organic carbon employing the different nature-inspired approaches in the Nakdong River, South Korea," *Environ. Monit. Assess.*, vol. 193, no. 7, p. 445, 2021, doi: 10.1007/s10661-021-08907-4.
148. J. Yan, J. Liu, Y. Yu, and H. Xu, "Water Quality Prediction in the Luan River Based on 1-DRCNN and BiGRU Hybrid Neural Network Model," *Water*, vol. 13, no. 9, 2021, doi: 10.3390/w13091273.
149. K. J. Setshedi, N. Mutingwende, and N. P. Ngqwala, "The Use of Artificial Neural Networks to Predict the Physicochemical Characteristics of Water Quality in Three District Municipalities, Eastern Cape Province, South Africa," *Int. J. Environ. Res. Public Health*, vol. 18, no. 10, 2021, doi: 10.3390/ijerph18105248.
150. S. I. Abba *et al.*, "Comparative implementation between neuro-emotional genetic algorithm and novel ensemble computing techniques for modelling dissolved oxygen concentration," *Hydrol. Sci. J.*, vol. 66, no. 10, pp. 1584–1596, 2021, doi: 10.1080/02626667.2021.1937179.
151. J. Sha, X. Li, M. Zhang, and Z.-L. Wang, "Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks," *Water*, vol. 13, no. 11, 2021, doi: 10.3390/w13111547.
152. Y.-F. Zhang, P. Fitch, and P. J. Thorburn, "Predicting the Trend of Dissolved Oxygen Based on the kPCA-RNN Model," *Water*, vol. 12, no. 2, 2020, doi: 10.3390/w12020585.
153. S.-S. Baek, J. Pyo, and J. A. Chun, "Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach," *Water*, vol. 12, no. 12, 2020, doi: 10.3390/w12123399.
154. M. Jamei, I. Ahmadianfar, X. Chu, and Z. M. Yaseen, "Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach," *J. Hydrol.*, vol. 589, p. 125335, 2020, doi: <https://doi.org/10.1016/j.jhydrol.2020.125335>.
155. D. T. Bui, K. Khosravi, J. Tiefenbacher, H. Nguyen, and N. Kazakis, "Improving prediction of water quality indices using novel hybrid machine-learning algorithms," *Sci. Total Environ.*, vol. 721, p. 137612, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.137612>.
156. S. Chen, G. Fang, X. Huang, and Y. Zhang, "Water Quality Prediction Model of a Water Diversion Project Based on the Improved Artificial Bee Colony–Backpropagation Neural Network," *Water*, vol. 10, no. 6, 2018, doi: 10.3390/w10060806.
157. B. Raheli, M. T. Aalami, A. El-Shafie, M. A. Ghorbani, and R. C. Deo, "Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River," *Environ. Earth Sci.*, vol. 76, no. 14, p. 503, 2017, doi: 10.1007/s12665-017-6842-z.
158. C. Stoica *et al.*, "Water quality of Danube Delta systems: ecological status and prediction using machine-learning algorithms," *Water Sci. Technol.*, vol. 73, no. 10, pp. 2413–2421, 2016, doi: 10.2166/wst.2016.097.
159. Q. Zhang and X. You, "Recent Advances in Surface Water Quality Prediction Using Artificial Intelligence Models," *Water Resour. Manag.*, vol. 38, no. 1, pp. 235–250, 2024, doi: 10.1007/s11269-023-03666-y.
160. V. Gómez-Escalonilla, E. Montero-González, S. Díaz-Alcaide, M. Martín-Loeches, M. R. del Rosario, and P. Martínez-Santos, "A machine learning approach to site groundwater contamination monitoring wells," *Appl. Water Sci.*, vol. 14, no. 12, p. 250, 2024, doi: 10.1007/s13201-024-02320-1.
161. J. Zhang *et al.*, "Improving prediction of groundwater quality in situations of limited monitoring data based on virtual sample generation and Gaussian process regression," *Water Res.*, vol. 267, p. 122498, 2024, doi: <https://doi.org/10.1016/j.watres.2024.122498>.

162. H. Jeong *et al.*, "Spatial prediction of groundwater salinity in multiple aquifers of the Mekong Delta region using explainable machine learning models," *Water Res.*, vol. 266, p. 122404, 2024, doi: <https://doi.org/10.1016/j.watres.2024.122404>.
163. X. Li *et al.*, "Identifying the spatial pattern and driving factors of nitrate in groundwater using a novel framework of interpretable stacking ensemble learning," *Environ. Geochem. Health*, vol. 46, no. 11, p. 482, 2024, doi: [10.1007/s10653-024-02201-1](https://doi.org/10.1007/s10653-024-02201-1).
164. A. Boufekane, M. Meddi, D. Maizi, and G. Busico, "Performance of artificial intelligence model (LSTM model) for estimating and predicting water quality index for irrigation purposes in order to improve agricultural production," *Environ. Monit. Assess.*, vol. 196, no. 11, p. 1049, 2024, doi: [10.1007/s10661-024-13211-y](https://doi.org/10.1007/s10661-024-13211-y).
165. A. Lal, A. Sharan, K. Sharma, A. Ram, D. K. Roy, and B. Datta, "Scrutinizing different predictive modeling validation methodologies and data-partitioning strategies: new insights using groundwater modeling case study," *Environ. Monit. Assess.*, vol. 196, no. 7, p. 623, 2024, doi: [10.1007/s10661-024-12794-w](https://doi.org/10.1007/s10661-024-12794-w).
166. I. Khan and M. Ayaz, "Sensitivity analysis-driven machine learning approach for groundwater quality prediction: Insights from integrating ENTROPY and CRITIC methods," *Groundw. Sustain. Dev.*, vol. 26, p. 101309, 2024, doi: <https://doi.org/10.1016/j.gsd.2024.101309>.
167. W. Cao *et al.*, "Prediction of arsenic and fluoride in groundwater of the North China Plain using enhanced stacking ensemble learning," *Water Res.*, vol. 259, p. 121848, 2024, doi: <https://doi.org/10.1016/j.watres.2024.121848>.
168. C. R. Das and S. Das, "Coastal groundwater quality prediction using objective-weighted WQI and machine learning approach," *Environ. Sci. Pollut. Res.*, vol. 31, no. 13, pp. 19439–19457, 2024, doi: [10.1007/s11356-024-32415-w](https://doi.org/10.1007/s11356-024-32415-w).
169. T. Chatterjee, U. R. Gogoi, A. Samanta, A. Chatterjee, M. K. Singh, and S. Pasupuleti, "Identifying the Most Discriminative Parameter for Water Quality Prediction Using Machine Learning Algorithms," *Water*, vol. 16, no. 3, 2024, doi: [10.3390/w16030481](https://doi.org/10.3390/w16030481).
170. A. J. Tesoriero, S. A. Wherry, D. I. Dupuy, and T. D. Johnson, "Predicting Redox Conditions in Groundwater at a National Scale Using Random Forest Classification," *Environ. Sci. Technol.*, vol. 58, no. 11, pp. 5079–5092, Mar. 2024, doi: [10.1021/acs.est.3c07576](https://doi.org/10.1021/acs.est.3c07576).
171. H. E. Elzain *et al.*, "An innovative approach for predicting groundwater TDS using optimized ensemble machine learning algorithms at two levels of modeling strategy," *J. Environ. Manage.*, vol. 351, p. 119896, 2024, doi: <https://doi.org/10.1016/j.jenvman.2023.119896>.
172. J. Iqbal *et al.*, "Hydrogeochemistry and prediction of arsenic contamination in groundwater of Vehari, Pakistan: comparison of artificial neural network, random forest and logistic regression models," *Environ. Geochem. Health*, vol. 46, no. 1, p. 14, 2023, doi: [10.1007/s10653-023-01782-7](https://doi.org/10.1007/s10653-023-01782-7).
173. L. Krishnamoorthy and V. R. Lakshmanan, "Groundwater quality assessment using machine learning models: a comprehensive study on the industrial corridor of a semi-arid region," *Environ. Sci. Pollut. Res.*, 2024, doi: [10.1007/s11356-024-34119-7](https://doi.org/10.1007/s11356-024-34119-7).
174. H. Mahboobi, A. Shakiba, and B. Mirbagheri, "Improving groundwater nitrate concentration prediction using local ensemble of machine learning models," *J. Environ. Manage.*, vol. 345, p. 118782, 2023, doi: <https://doi.org/10.1016/j.jenvman.2023.118782>.
175. A. M. Sajib, M. T. M. Diganta, A. Rahman, T. Dabrowski, A. I. Olbert, and M. G. Uddin, "Developing a novel tool for assessing the groundwater incorporating water quality index and machine learning approach," *Groundw. Sustain. Dev.*, vol. 23, p. 101049, 2023, doi: <https://doi.org/10.1016/j.gsd.2023.101049>.
176. R. Masoudi, S. R. Mousavi, P. D. Rahimabadi, M. Panahi, and A. Rahmani, "Assessing data mining algorithms to predict the quality of groundwater resources for determining irrigation hazard," *Environ. Monit. Assess.*, vol. 195, no. 2, p. 319, 2023, doi: [10.1007/s10661-022-10909-9](https://doi.org/10.1007/s10661-022-10909-9).
177. C. Liu, M. Xu, Y. Liu, X. Li, Z. Pang, and S. Miao, "Predicting Groundwater Indicator Concentration Based on Long Short-Term Memory Neural Network: A Case Study," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, 2022, doi: [10.3390/ijerph192315612](https://doi.org/10.3390/ijerph192315612).

178. T.-M.-T. Huynh *et al.*, "Predicting Heavy Metal Concentrations in Shallow Aquifer Systems Based on Low-Cost Physiochemical Parameters Using Machine Learning Techniques," *Int. J. Environ. Res. Public Health*, vol. 19, no. 19, 2022, doi: 10.3390/ijerph191912180.
179. M. Taşan, S. Taşan, and Y. Demir, "Estimation and uncertainty analysis of groundwater quality parameters in a coastal aquifer under seawater intrusion: a comparative study of deep learning and classic machine learning methods," *Environ. Sci. Pollut. Res.*, vol. 30, no. 2, pp. 2866–2890, 2023, doi: 10.1007/s11356-022-22375-4.
180. K. Banerjee *et al.*, "A Machine-Learning Approach for Prediction of Water Contamination Using Latitude, Longitude, and Elevation," *Water*, vol. 14, no. 5, 2022, doi: 10.3390/w14050728.
181. S. Kouadri, C. B. Pande, B. Panneerselvam, K. N. Moharir, and A. Elbeltagi, "Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models," *Environ. Sci. Pollut. Res.*, vol. 29, no. 14, pp. 21067–21091, 2022, doi: 10.1007/s11356-021-17084-3.
182. K. P. Messier *et al.*, "Modeling groundwater nitrate exposure in private wells of North Carolina for the Agricultural Health Study," *Sci. Total Environ.*, vol. 655, pp. 512–519, 2019, doi: <https://doi.org/10.1016/j.scitotenv.2018.11.022>.
183. M. Sakizadeh, "Spatial analysis of total dissolved solids in Dezful Aquifer: Comparison between universal and fixed rank kriging," *J. Contam. Hydrol.*, vol. 221, pp. 26–34, 2019, doi: <https://doi.org/10.1016/j.jconhyd.2019.01.001>.
184. F. Abbas *et al.*, "Machine Learning Models for Water Quality Prediction: A Comprehensive Analysis and Uncertainty Assessment in Mirpurkhas, Sindh, Pakistan," *Water*, vol. 16, no. 7, 2024, doi: 10.3390/w16070941.
185. M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," *J. Hydrol.*, vol. 598, p. 126266, 2021, doi: <https://doi.org/10.1016/j.jhydrol.2021.126266>.
186. A. Aldrees, H. H. Awan, M. F. Javed, and A. M. Mohamed, "Prediction of water quality indexes with ensemble learners: Bagging and boosting," *Process Saf. Environ. Prot.*, vol. 168, pp. 344–361, 2022, doi: <https://doi.org/10.1016/j.psep.2022.10.005>.
187. Y. Chen *et al.*, "Water Quality Inversion of a Typical Rural Small River in Southeastern China Based on UAV Multispectral Imagery: A Comparison of Multiple Machine Learning Algorithms," *Water*, vol. 16, no. 4, 2024, doi: 10.3390/w16040553.
188. A. Najah, A. El-Shafie, O. A. Karim, and A. H. El-Shafie, "Application of artificial neural networks for water quality prediction," *Neural Comput. Appl.*, vol. 22, no. 1, pp. 187–201, 2013, doi: 10.1007/s00521-012-0940-3.
189. S. Gulati and A. Pal, "Tuning fuzzy logic controller with SGWO for river water quality modelling," *Mater. Today Proc.*, vol. 54, pp. 733–737, 2022, doi: <https://doi.org/10.1016/j.matpr.2021.10.467>.
190. J. Dilipkumar and P. Shanmugam, "Fuzzy-based global water quality assessment and water quality cells identification using satellite data," *Mar. Pollut. Bull.*, vol. 193, p. 115148, 2023, doi: <https://doi.org/10.1016/j.marpolbul.2023.115148>.
191. A. Kalyakulina, I. Yusipov, A. Moskalev, C. Franceschi, and M. Ivanchenko, "eXplainable Artificial Intelligence (XAI) in aging clock models," *Ageing Res. Rev.*, vol. 93, p. 102144, 2024, doi: <https://doi.org/10.1016/j.arr.2023.102144>.
192. M. K. Nallakaruppan, E. Gangadevi, M. L. Shri, B. Balusamy, S. Bhattacharya, and S. Selvarajan, "Reliable water quality prediction and parametric analysis using explainable AI models," *Sci. Rep.*, vol. 14, no. 1, p. 7520, 2024, doi: 10.1038/s41598-024-56775-y.
193. S. Kundu, P. Datta, P. Pal, K. Ghosh, A. Das, and B. K. Das, "Unveiling the hidden connections: Using explainable artificial intelligence to assess water quality criteria in nine giant rivers," *J. Clean. Prod.*, vol. 492, p. 144861, 2025, doi: <https://doi.org/10.1016/j.jclepro.2025.144861>.
194. X. Nong, C. Lai, L. Chen, and J. Wei, "A novel coupling interpretable machine learning framework for water quality prediction and environmental effect understanding in different flow discharge regulations of hydro-projects," *Sci. Total Environ.*, vol. 950, p. 175281, 2024, doi: <https://doi.org/10.1016/j.scitotenv.2024.175281>.

195. A. Juna *et al.*, "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron," *Water*, vol. 14, no. 17, 2022, doi: 10.3390/w14172592.
196. U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, "Efficient Water Quality Prediction Using Supervised Machine Learning," *Water*, vol. 11, no. 11, 2019, doi: 10.3390/w11112210.
197. R. Pany, A. Rath, and P. C. Swain, "Water quality assessment for River Mahanadi of Odisha, India using statistical techniques and Artificial Neural Networks," *J. Clean. Prod.*, vol. 417, p. 137713, 2023, doi: <https://doi.org/10.1016/j.jclepro.2023.137713>.
198. S. Yang *et al.*, "Spatial Mapping and Prediction of Groundwater Quality Using Ensemble Learning Models and SHapley Additive exPlanations with Spatial Uncertainty Analysis," *Water*, vol. 16, no. 17, 2024, doi: 10.3390/w16172375.
199. S. Heydari, M. Reza Nikoo, A. Mohammadi, and R. Barzegar, "Two-stage meta-ensembling machine learning model for enhanced water quality forecasting," *J. Hydrol.*, vol. 641, p. 131767, 2024, doi: <https://doi.org/10.1016/j.jhydrol.2024.131767>.
200. L. Rodríguez-López *et al.*, "Estimation of Water Quality Parameters through a Combination of Deep Learning and Remote Sensing Techniques in a Lake in Southern Chile," *Remote Sens.*, vol. 15, no. 17, 2023, doi: 10.3390/rs15174157.
201. T. Zhang, J. Wu, H. Chu, J. Liu, and G. Wang, "Interpretable Machine Learning Based Quantification of the Impact of Water Quality Indicators on Groundwater Under Multiple Pollution Sources," *Water*, vol. 17, no. 6, 2025, doi: 10.3390/w17060905.
202. Z. Yao, Z. Wang, J. Huang, N. Xu, X. Cui, and T. Wu, "Interpretable prediction, classification and regulation of water quality: A case study of Poyang Lake, China," *Sci. Total Environ.*, vol. 951, p. 175407, 2024, doi: <https://doi.org/10.1016/j.scitotenv.2024.175407>.
203. S. Shadkani, Y. Hemmatzadeh, A. Saber, and M. Mohammadi Sergini, "Enhanced predictive modeling of dissolved oxygen concentrations in riverine systems using novel hybrid temporal pattern attention deep neural networks," *Environ. Res.*, vol. 263, p. 120015, 2024, doi: <https://doi.org/10.1016/j.envres.2024.120015>.
204. S. Y. Abuzir and Y. S. Abuzir, "Machine learning for water quality classification," *Water Qual. Res. J.*, vol. 57, no. 3, pp. 152–164, 2022, doi: 10.2166/wqrj.2022.004.
205. Q. Zhang *et al.*, "Real-time prediction of river chloride concentration using ensemble learning," *Environ. Pollut.*, vol. 291, p. 118116, 2021, doi: <https://doi.org/10.1016/j.envpol.2021.118116>.
206. S. Fertikh, H. Boutaghane, M. Boumaaza, A. Belaadi, and S. Bouslah, "Assessment and prediction of water quality indices by machine learning-genetic algorithm and response surface methodology," *Model. Earth Syst. Environ.*, vol. 10, no. 4, pp. 5573–5604, 2024, doi: 10.1007/s40808-024-02079-z.
207. M. Fooladi, M. R. Nikoo, R. Mirghafari, C. A. Madramootoo, G. Al-Rawas, and R. Nazari, "Robust clustering-based hybrid technique enabling reliable reservoir water quality prediction with uncertainty quantification and spatial analysis," *J. Environ. Manage.*, vol. 362, p. 121259, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.121259>.
208. M. Ghashghaie, H. Eslami, and K. Ostad-Ali-Askari, "Applications of time series analysis to investigate components of Madiyan-rood river water quality," *Appl. Water Sci.*, vol. 12, no. 8, p. 202, 2022, doi: 10.1007/s13201-022-01693-5.
209. A. K. Bojer *et al.*, "Machine learning and remote sensing based time series analysis for drought risk prediction in Borena Zone, Southwest Ethiopia," *J. Arid Environ.*, vol. 222, p. 105160, 2024, doi: <https://doi.org/10.1016/j.jaridenv.2024.105160>.
210. S. Huan, "A novel interval decomposition correlation particle swarm optimization-extreme learning machine model for short-term and long-term water quality prediction," *J. Hydrol.*, vol. 625, p. 130034, 2023, doi: <https://doi.org/10.1016/j.jhydrol.2023.130034>.
211. A. Yadav, A. Raj, and B. Yadav, "Enhancing local-scale groundwater quality predictions using advanced machine learning approaches," *J. Environ. Manage.*, vol. 370, p. 122903, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.122903>.
212. S. Soleymani Hasani, M. E. Arias, H. Q. Nguyen, O. M. Tarabih, Z. Welch, and Q. Zhang, "Leveraging explainable machine learning for enhanced management of lake water quality," *J. Environ. Manage.*, vol. 370, p. 122890, 2024, doi: <https://doi.org/10.1016/j.jenvman.2024.122890>.

213. D. Ezzat, M. Soliman, E. Ahmed, and A. E. Hassanien, "An optimized explainable artificial intelligence approach for sustainable clean water," *Environ. Dev. Sustain.*, vol. 26, no. 10, pp. 25899–25919, 2024, doi: 10.1007/s10668-023-03712-0.
214. S. Maroufpoor, M. Jalali, S. Nikmehr, N. Shiri, J. Shiri, and E. Maroufpoor, "Modeling groundwater quality by using hybrid intelligent and geostatistical methods," *Environ. Sci. Pollut. Res.*, vol. 27, no. 22, pp. 28183–28197, 2020, doi: 10.1007/s11356-020-09188-z.
215. M. I. Shah, M. F. Javed, A. Alqahtani, and A. Aldrees, "Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data," *Process Saf. Environ. Prot.*, vol. 151, pp. 324–340, 2021, doi: <https://doi.org/10.1016/j.psep.2021.05.026>.
216. H. Moayedi, M. Salari, S. A.-J. Ali, A. A. Dehrashid, and H. Azadi, "Modeling the total hardness (TH) of groundwater in aquifers using novel hybrid soft computing optimizer models," *Environ. Earth Sci.*, vol. 83, no. 13, p. 392, 2024, doi: 10.1007/s12665-024-11618-x.
217. J. Jiao, Q. Ma, S. Huang, F. Liu, and Z. Wan, "A hybrid water quality prediction model based on variational mode decomposition and bidirectional gated recursive unit," *Water Sci. Technol.*, vol. 89, no. 9, pp. 2273–2289, 2024, doi: 10.2166/wst.2024.133.
218. R. K. Makumbura *et al.*, "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 Eng.*, vol. 23, p. 102831, 2024, doi: <https://doi.org/10.1016/j.rineng.2024.102831>.
219. M. Bordbar, G. Busico, M. Sirna, D. Tedesco, and M. Mastrocicco, "A multi-step approach to evaluate the sustainable use of groundwater resources for human consumption and agriculture," *J. Environ. Manage.*, vol. 347, p. 119041, 2023, doi: <https://doi.org/10.1016/j.jenvman.2023.119041>.
220. J. M. Lee, K.-S. Ko, and K. Yoo, "A machine learning-based approach to predict groundwater nitrate susceptibility using field measurements and hydrogeological variables in the Nonsan Stream Watershed, South Korea," *Appl. Water Sci.*, vol. 13, no. 12, p. 242, 2023, doi: 10.1007/s13201-023-02043-9.
221. H. Zheng, Y. Liu, W. Wan, J. Zhao, and G. Xie, "Large-scale prediction of stream water quality using an interpretable deep learning approach," *J. Environ. Manage.*, vol. 331, p. 117309, 2023, doi: <https://doi.org/10.1016/j.jenvman.2023.117309>.
222. Z. Zhang, J. Huang, S. Duan, Y. Huang, J. Cai, and J. Bian, "Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed," *Ecol. Indic.*, vol. 140, p. 108977, 2022, doi: <https://doi.org/10.1016/j.ecolind.2022.108977>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.