

Evaluating Effective AI Models for Surface and Groundwater Quality Prediction: A Review

Mohammad Khoshnam ^a , Hojat Karami ^a , Amin Shams ^{a*} , Ozgur Kisi ^{b, c, d} 

^a Faculty of Civil Engineering, Semnan University, Semnan, Iran

^b Department of Civil Engineering, Lübeck University of Applied Sciences, 23562, Lübeck, Germany

^c Department of Civil Engineering, Ilia State University, 0162, Tbilisi, Georgia

^d School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, South Korea

ARTICLE INFO

Article history:

Received 03 August 2026

Revised 18 September 2026

Accepted 30 September 2026

Available online 01 July 2027

Keywords:

Artificial intelligence

Water quality

Surface water

Groundwater

Water management

Hydro-informatics

Sustainability

ABSTRACT

Artificial intelligence (AI) techniques have become important tools for predicting and assessing water quality because they can capture nonlinear relationships among complex environmental variables. This review systematically analyzes AI-based approaches for surface water and groundwater quality prediction and identifies factors that influence model performance. Following PRISMA guidelines, we conducted a systematic review using five scientific databases, screening 224 records and including 35 peer-reviewed studies. We analyzed the reviewed studies by AI model type, water resource category, input parameters, dataset characteristics, hydrological conditions, and prediction performance. The results indicate that hybrid and ensemble models generally outperform standalone algorithms by improving feature extraction, reducing uncertainty, and better capturing nonlinear relationships. GPR showed high accuracy for WQI prediction with uncertainty below 2%, RF-SVM achieved strong performance for BOD prediction ($R^2 = 0.908$), and LSTM-based and metaheuristic models improved DO prediction. Despite their high accuracy, AI models still face challenges related to interpretability, data availability, and regional transferability. This review compares AI model suitability under different hydrological conditions and proposes future directions in explainable AI, uncertainty quantification, and adaptive real-time water quality monitoring.

How to cite this article: Khoshnam, M., Karami, H., Shams, A., Kisi, O. Evaluating Effective AI Models for Surface and Groundwater Quality Prediction: A Review. Civil Engineering and Applied Solutions. 2027; 3(3): 190–211. doi:10.22080/ceas.2026.32582.1127.

1. Introduction

Water is a limited, essential natural resource that plays a fundamental role in sustaining life, ecosystems, and socio-economic development. As the world population continues to rise rapidly (estimated to exceed 9 billion by 2050), demand for clean, accessible water is rising at a record rate. This increased demand puts immense pressure on existing water supplies, some of which are already stressed by pollution, global warming, and overuse. Maintaining water quality is therefore crucial not only for public health but also for environmental protection and sustainable development. However, traditional methods of quantifying water quality, while effective, are slow, resource-intensive, and limited in providing real-time or predictive results.

In these situations, incorporating artificial intelligence (AI) into water quality measurement has emerged as a promising, advanced approach. AI technologies can handle large, complex sensor and remote-sensing datasets to provide more accurate, efficient, and timely evaluations of water conditions. AI can also identify hidden patterns, predict contamination events, and support decision-making in water resource management. These are potential applications, but a major scientific gap remains in

* Corresponding author.

E-mail address: amin.shams@semnan.ac.ir (A. Shams).



<https://doi.org/10.22080/ceas.2026.32582.1127>

ISSN: 3092-7749/© 2027 The Author(s). Published by University of Mazandaran.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/deed.en>)

systematically using and testing AI models across a broad spectrum of hydrological conditions. Research that addresses these differences is critical to unlocking AI's full potential for managing water quality and building scalable, adaptive, and robust systems. Filling this research gap not only improves our technological solutions for water problems but also supports global sustainability goals by protecting one of the world's greatest assets.

Elkiran et al. [1] presented single-step and multi-step prediction models with AI and ensemble techniques for dissolved oxygen (DO) at three Yamuna River stations. They used AI models such as back propagation neural network (BPNN), adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM), along with auto-regressive integrated moving average (ARIMA), and each yielded better results than the others at different stations. Additionally, ensemble techniques such as simple average ensemble (SAE), weighted average ensemble (WAE), and neural network ensemble (NNE) further improved prediction accuracy and showed that NNE was the most robust. The combination of various algorithms with ensemble methods in the suggested study offers a way to improve predictions for DO and other water quality parameters.

Hameed et al. [2] applied an artificial neural network (ANN) to predict water quality index (WQI) from tropical environments using monthly water quality measurements collected over ten years from 2001 to 2010. The researchers compared two models: BPNN and radial basis function neural network (RBFNN). ANN predicted the WQI satisfactorily, but RBFNN performed best. This study identifies RBFNN as a powerful tool for river water quality assessments, with strong potential to model real systems involving nonlinear dynamics and stochastic behavior for each water quality parameter.

Najah Ahmed et al. [3] evaluated the enhanced wavelet de-noising technique (WDT)-ANFIS for predicting water quality parameters influenced by urbanization. Considering the Johor River dataset, this work evaluates the accuracy and reliability of WDT-ANFIS, ANFIS, multi-layer perceptron (MLP)-ANN, and radial basis function neural networks (RBF-ANN). The former was found to be the best among these for simulating complex and dynamic processes. Of the varied scenarios studied, scenario 2, which included the data from the upstream station, showed the best prediction. The study further opens up possibilities for WDT-ANFIS to make river basin management and policy-making more effective through improved water-quality forecasting.

Ahmed et al. [4] used machine learning (ML) to forecast water quality with the fewest parameters and used 663 samples obtained from Rawal Lake in Pakistan. Their results showed that polynomial regression and gradient boosting (GB) were highly successful in predicting WQI, while the MLP algorithm performed best for water quality class (WQC). Their work, therefore, has great utility for assessing water quality in the subject models using the mentioned ML techniques.

Azad et al. [5] investigated genetic algorithm (GA), differential evolution (DE), and ant colony optimization (ACO) for continuous domains (ACOR) evolutionary algorithms for optimization of the ANFIS model to predict the water quality parameters of irrigation and drinking water, namely electrical conductivity (EC), sodium absorption ratio (SAR), and total hardness (TH). The results showed that the integrated models were highly accurate, and ANFIS-DE performed more efficiently than the other methods. In the current research, these algorithms significantly reduced uncertainty and overfitting in the ANFIS model during training and validation. Among them, ANFIS-DE performed better in terms of performance measures and prediction accuracy. In a nutshell, the study showed that metaheuristic algorithms can effectively enhance ANFIS for predicting river water quality parameters.

Wang et al. [6] estimated the WQI by remote sensing data and algorithms in the Ebinur Lake watershed of China. Accordingly, WQI values ranged from 56.6133 to 2,886.5198, representing generally low water quality in the arid region. The coefficient of correlation (R) was high and increased with increasing orders of derivatives up to order 1.6, improving the predictive capability of the models. Hence, among all 22 generated models, the ratio index (RI), difference index (DI), and normalized difference index (NDI)-based models performed best at the 1.6-order derivative. Thus, this new algorithm shows strong potential for application in other semi-arid regions.

Granata et al. [7] applied an indirect method to estimate key effluent quality parameters. In this method, they considered several characteristics of an urban catchment: drainage area; percentages of residential, institutional, commercial, and industrial areas; open space areas; freeway and impervious area percentages; precipitation depth; and runoff. It compared two ML algorithms: support vector regression (SVR) and regression tree. Both models were robust and reliable and showed strong generalization capability. However, SVR outperformed the regression tree in the estimation of total suspended solids (TSS), total dissolved solids (TDS), and chemical oxygen demand (COD) based on the root mean square error (RMSE) and coefficient of determination (R^2). For biochemical oxygen demand (BOD), both models performed similarly. ML algorithms can also help estimate values for sizing treatment units using basin characteristics when direct measurements are not available. Their study showed this could be an effective methodology for real-time management of wastewater treatment plants and environmental planning.

Hridoy et al. [8] tested different machine learning models, such as extreme gradient boosting (XGB), random forest (RF), and SVM, for water quality classification and WQI prediction. The authors used grid search to tune model parameters and Shapley additive explanations (SHAP) to interpret predictions and identify key water-quality predictors. The tests found that ensemble machine learning models, especially the XGB model, provided outstanding classification results, with an accuracy of 99.65%. For WQI prediction, XGB demonstrated outstanding performance, with an R^2 of 0.9685. According to SHAP analysis, the main water quality predictors were DO and BOD, followed by turbidity, nitrate, and EC.

Li et al. [9] used machine learning models to predict ultrapure water quality one hour in advance using total organic carbon (TOC) and resistivity. They compared six approaches and chose Optuna-XGB for TOC prediction. The model performed well, with R^2 values of 0.9664 and 0.9670, respectively. We used SHAP and partial dependence plots to interpret the results and identify the major factors influencing water quality. Return water resistivity was the major factor in both cases, and we observed significant

nonlinear interactions between reverse osmosis rejection rate and operational conditions.

Mena et al. [10] presented an integrated framework for water quality assessment and regulatory compliance in river basins under hydrological uncertainty. The methodology combines a long short-term memory (LSTM) deep learning (DL) algorithm for streamflow forecasting, Monte Carlo simulation for uncertainty propagation, hydrological disaggregation through the HBV model, and process-based water quality modeling. The methodology was tested in the Júcar River basin in eastern Spain, characterized by high regulation, hydrological uncertainty, and water quality problems. The process-based water quality model simulates BOD, phosphorus, and ammonium concentrations. The framework reproduced the spatial and temporal patterns of water pollution and quantified uncertainty in hydrological forecasting. Accuracies in the chemical status classification reached 90% for BOD, 84.4% for phosphorus, and 85.3% for ammonium.

Due to the extensive utilization of AI techniques in predicting water parameters [11–14], and also because studies have shown that even one AI technique can be utilized in various aspects of water engineering [15], analyzing and classifying studies conducted in the field of water quality prediction via AI can help to achieve a better and more detailed horizontal understanding in this field.

2. Research method

In this article, to develop a suitable package of "AI application in water quality determination," we first defined a method for searching the latest studies in this field. Then, using keywords including "AI", "water quality", "water resources," etc., our survey began. We used the ScienceDirect, MDPI, Wiley Online Library, Nature, and Springer reference databases to access relevant research articles. Table 1 illustrates the data extraction and methodological characteristics. After receiving several articles on "artificial intelligence application in water quality determination", all non-English literature and conference abstracts were omitted. By reviewing each article with focus and precision on the points and keywords, we extracted a suitable abstract from each article. Finally, we matched and ordered the abstracts to prepare this article. All research articles are properly referenced. Fig. 1 demonstrates the PRISMA flow diagram of this review.

Table 1. Data extraction and methodological characteristics of the studies included in the systematic review.

Databases	ScienceDirect, MDPI, Wiley, Nature, Springer
Search period	e.g., exact years
Search terms	exact Boolean string
Initial records	224
Duplicates	57
Screened	148
Full-text assessed	56
Final studies	35
Inclusion criteria	explicit
Exclusion criteria	explicit

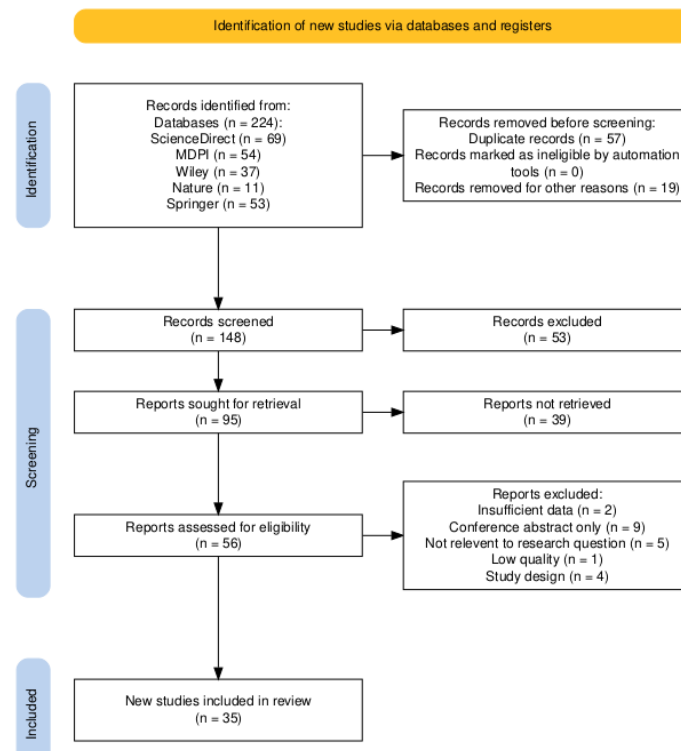


Fig. 1. PRISMA flow diagram.

This systematic review followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. We conducted a comprehensive literature search across five scientific databases: ScienceDirect, MDPI, Wiley Online Library, Nature, and Springer. We defined the search period as 2017-2024 to focus on recent developments in artificial intelligence applications for water quality prediction. The search strategy was based on combinations of keywords and Boolean operators, including ("artificial intelligence" or "machine learning" or "deep learning" or "neural network") and ("water quality" or "water quality index" or "surface water" or "groundwater") and ("prediction" or "assessment" or "modeling"). The initial search identified 224 records. After removing 57 duplicates, we screened 148 articles for title and abstract relevance. We then assessed 56 articles through full-text review, and finally included 35 studies in this review. Studies were included if they (i) were published as peer-reviewed journal articles, (ii) were written in English, (iii) applied AI-based methods for water quality assessment or prediction, and (iv) provided sufficient information regarding model development and evaluation. We excluded conference papers, review articles, non-English publications, and studies unrelated to AI-based water quality prediction. The screening and selection process followed the PRISMA framework presented in Fig. 1.

Journal reputation and name did not influence article selection. We used article quality and topic alone as evaluation criteria to keep the review on course. Therefore, the selected articles share a theme and topic, enabling a more reliable, valid, and meaningful conclusion. We limited the articles to the most recent publication years. This ensured that recent studies support the results and minimized the possibility that the selected articles had already been reviewed in previous review articles. In addition, we selected articles with minimal geographic overlap to examine a wider range of regions.

3. Artificial intelligence application in water quality determination

3.1. Surface Water

Wang et al. [16] used water quality monitoring data from Shandong province, China. They devised an AI system to identify spatial and temporal variations in water quality and understand the relationships between water quality measures and sources of industrial pollution. In this respect, the method proposes an intelligent approach to assess complex relationships between pollution sources and water quality data and to investigate interactions among water quality indicators. It then employs cross-correlation analysis to map water quality indicators strongly correlated with industrial pollution sources. The findings align with real-world conditions observed in water quality and pollution sources. These maps can help identify any sudden changes in discharge due to pollution and locate industries that affect water quality. Second, the Apriori method identified high-frequency indicator pairs across monitoring stations, showing that water quality parameters vary spatially and change with industrial activity over time. Using LSTM, the neural network predicted how industrial pollution would affect future water quality. This will increase our ability to manage pollutant discharge and help enhance river basin water quality.

Nasir et al. [17] focused on water quality classification using ML techniques and proposed an intelligent, real-time approach to water quality monitoring. They extended this with different standard variants and ensemble models to conduct a comprehensive analysis. This dataset consists of over 1600 samples collected from various places in India, ranging from 2005 to 2014. The metadata parameters include DO, pH, conductivity, BOD, nitrate, fecal coliform, and total coliform. The ML classifiers included SVM, random forest (RF), logistic regression, categorical boosting (CATBoost), XGB, DT, and MLP. Performance metrics calculated from the confusion matrix included precision, sensitivity, F1 score, and accuracy. Precision-recall curves, receiver operating characteristic (ROC) curves, and average precision curves were also obtained. Among these, CATBoost achieved the highest maximum precision (94.51%), followed by the RF classifier (94%), then the MLP (88.6%), XGB (88.1%), DT (81.6%), SVM (80.7%), and logistic regression (72.9%). However, in the stacking ensemble model, various meta-classifiers, including DT and CATBoost, achieved 100% accuracy. Thus, CATBoost became the best classifier and the best stacking model, improving accuracy. Because of its strong performance metrics, the CATBoost classifier is a good tool to apply to different datasets. Such analyses can be further enriched in cooperation with experts in ML implementation and other fields, such as hardware implementation, internet of things (IoT), and biomedical research.

Bui et al. [18] examined the performance of 16 data-mining algorithms, including four standalone techniques such as RF, M5 Prime (M5P), random tree (RT), and reduced error pruning tree (REPT), and 12 hybrid versions, which are standalone combined with bagging, CV parameter selection (CVPS), and randomizable filtered classification (RFC) for prediction of WQI every month in a humid area in Mazandaran and Semnan provinces in northern Iran. In addition to developing new WQI prediction algorithms, the aim was to capture other aspects of water science in areas without regular water quality monitoring stations. Among the parameters, fecal coliform concentration was identified as the most influential factor in determining the WQI, followed by BOD, DO, EC, COD, turbidity, total solids (TS), and pH. Different combinations of the selected variables may affect model performance differently, and irregularities in model-input variation may occur across catchments. A high R value indicates good performance; otherwise, low R values reduce predictive power. Generally, hybrid algorithms somewhat outperformed the standalone algorithms, although the results differed in practice. For instance, the BA-RT model performed best, followed by RF, bagging-RF, etc. However, it performed poorly for extreme WQI values. Most algorithms tended to overestimate WQI, except RT, BA-RT, and CVPS-REPT. These will be especially valuable in catchments with limited monitoring networks, which are common in most developing countries because of the high cost of measurement. However, suitability varies by algorithm and cannot be generalized across all contexts. BA-RT is a highly promising tool for managing groundwater quality in the humid regions of northern Iran and may be less effective elsewhere.

Sajib et al. [19] evaluated the water quality for the Tongi canal in Bangladesh. They used an unweighted Root Mean Square

(RMS)-WQI model along with the eight ML algorithms, namely ensembles of trees (EN), Gaussian process regression (GPR), linear regression (LR), neural networks (NN), stepwise linear regression (SLR), SVM, regression of trees (TR), and XGB. We developed these models to predict the WQI score. We used various model evaluation techniques, including RMSE, Mean absolute error (MAE), R^2 , Nash-Sutcliffe efficiency (NSE), model efficiency factor (MEF), PERI, Z statistics, and Taylor statistics, to identify the most accurate model. We also performed sensitivity and uncertainty analyses using R^2 and Monte Carlo simulation (MCS) methods. The main findings are: (1) all parameters of water quality exceeded the guideline values prescribed by Bangladesh. (2) The WQI score showed the "poor" quality of water in the Tongi canal, and thus, it is not suitable for any purpose. GPR performed best across various metrics, including RMSE, MAE, R^2 , NSE, MEF, Z, and Taylor statistics, for predicting the WQI score. It performed well in evaluating the ecological state of urban waters. The uncertainty analysis showed that the RMS-WQI model reduces uncertainty to below 2% without key issues such as eclipsing or ambiguity. Therefore, the RMS-WQI model seems appropriate for assessing water quality in similar inland water bodies.

Yang [20] applied four nature-inspired stochastic search methods, including teaching-learning-based optimization, sine cosine algorithm (SCA), water cycle algorithm (WCA), and Electromagnetic Field Optimization (EFO), to train an ANN to predict DO in the Klamath River, Oregon, US. All methods could train the MLPNN to reliably capture DO behavior by properly tuning each algorithm's parameters. Therefore, the models could correctly predict DO under new environmental conditions. To determine the best hybrid model, we compared the models in terms of accuracy, complexity, and computation time. During training, the EFO algorithm required 30 times more iterations but completed the process much faster than the other three algorithms. It also produced the best results in the testing phase, as measured by RMSE, R, and NSE. Another merit of this model is that it uses fewer search agents to determine the best response. WCA-MLPNN was the second-best performer. In the final part of the study, we proposed two DO predictive models, with weights and biases tuned by the WCA and EFO algorithms. The superior models herein also outperformed several hybrids and conventional models reported in other studies, demonstrating improved practical DO prediction. This development may also help address the common problem of poor water quality in the study area.

Abba et al. [21] developed, at three stations along the course of the Yamuna River (Nizamuddin, Palla, and Udi-Chambal), applications for the estimation of WQI using four different single models: BPNN, ANFIS, SVR, and multiple linear regression (MLR), through four various model combinations by implementing these single models. We assessed predictive accuracy using statistical criteria such as R^2 , RMSE, and R. Nonlinear ensemble techniques further improved model performance. The quantitative analysis showed that, during validation at Nizamuddin station, ANFIS-M3 improved absolute error by 30%, 41%, and 53% compared with BPNN, SVR, and MLR, respectively. At the Palla station, the BPNN-M4 model reduced the absolute error by up to 4% compared with the MLR model, whereas the remaining models (SVR and ANFIS) showed only marginal improvements. Meanwhile, for the Udi (Chambal) station, the absolute error increased by about 2% and 3% for the BPNN-M3 model compared to the SVR and MLR models, respectively. Overall, the results indicated that the NNE method provides efficient WQI predictions at the three stations along the Yamuna River and can be considered a reliable prediction approach. However, although the ensemble method outperformed any single model, a single model can still provide efficient predictions. Overall, the study showed that, in ensemble techniques, combining the best outputs from different models yields better performance than any single model. Further research on other data intelligence models is recommended to improve WQI prediction. Finally, an ensemble technique developed in their work can be applied to various rivers across different regions. While it outperforms standalone models, there is still room for improvement.

Wu and Wang [22] proposed the ANN-wavelet transform (WT)-LSTM model to improve water quality prediction accuracy. This model integrates ANN, WT, and LSTM. The model uses water quality data from the Jinjiang River basin in China to predict water quality and analyze trends. Because some data are missing due to instrument failure, an ANN fills the gaps using time-series information. WT then decomposed and reconstructed the time series, filtering out the short-term random noise and improving the model's prediction accuracy for both out-of-sample data and future dynamic trends. This approach allowed for more effective short-term and long-term water quality predictions. Consequently, the ANN-WT-LSTM model outperformed ANN-LSTM and nonlinear autoregressive (NAR) neural network models across all metrics and significantly improved water quality prediction accuracy, which is important for water environment protection. They provide essential data support for water quality management decisions and have theoretical and practical significance for the sustainable development and environmental protection of riverine and reservoir areas. Although time series analysis has successfully predicted future reservoir water quality trends, most prior studies have limitations: they can predict only single-point water quality and do not account for spatial variation. To establish a comprehensive reservoir early warning system, multiple water-quality monitoring points are recommended. Integrating water quality prediction with GIS technology would capture temporal and spatial changes in water quality and lay the basis for a robust water quality early warning system.

Asadollah et al. [23] compare the performance of three ML models for WQI forecasting in the Lam Tsuen River, Hong Kong, using monthly water-quality measurements. These are standalone SVR and decision tree regression (DTR); the novel ensemble extra tree regression (ETR) algorithm; and various combinations of input parameters, M1 to M10. The results show that the ETR model, by taking ten input parameters, offers the best performance in WQI prediction. SVR performed similarly but slightly worse, and the DTR model performed the poorest, though it still predicted WQI fairly well. Remarkably, the ETR model predicted equally well with three inputs of BOD, turbidity, and phosphate. This suggests that WQI can be estimated accurately with limited monitoring, provided these three parameters are measured. The uncertainty analysis showed that WQI prediction is more sensitive to the model type (DTR, SVR, and ETR) rather than the input variables. With this in mind, ensemble learning models that combine weak learners such as DTs have higher accuracy than stand-alone decision-tree models or traditional AI models, including SVR.

More precisely, the ETR-M3 model, along with other models proposed in this work, can enable more efficient WQI computation for water quality engineers and decision-makers and, therefore, support real-time monitoring and better water management practices. However, a limitation is selecting only certain chemical and physical parameters, since the chosen sample may prove unsatisfactory.

Lu and Ma [24] developed two hybrid DT-based models, namely complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)-XGB and CEEMDAN-RF, for water quality prediction. In these models, CEEMDAN decomposes raw data with significant fluctuations further to enhance predictive performance for both XGB and RF. This study examines water quality in Gales Creek in the Tualatin River, Oregon, USA. We used data generated from May 1st to July 20th, 2019. The data were split into training and test sets in a 9:1 ratio. The models predict water temperature, DO, pH, specific conductance, turbidity, and fluorescent dissolved organic matter (FDOM), and their results are compared with benchmark models. In addition, the standard deviation of errors (SDE) is used to compare and check the model stability. Thus, the results show that the temperature, DO, and specific conductance tasks yield excellent performance; for these three, CEEMDAN-RF achieved the best performance, with respective mean absolute percentage errors (MAPE) of 0.69%, 1.05%, and 0.90%. However, pH value, turbidity, and FDOM performed best in CEEMDAN-XGB, with MAPEs of 0.27%, 14.94%, and 1.59%, respectively. (2) The overall performances of the RF, XGB, CEEMDAN-RF, CEEMDAN-XGB, particle swarm optimization (PSO)-SVM, RBFNN, least squares SVM, and LSTM models have average MAPEs of 6.41%, 4.60%, 3.90%, 3.71%, 8.96%, 6.93%, 6.96%, and 7.35%, respectively. Altogether, CEEMDAN-XGB has the best predictive performance among the compared models, followed by CEEMDAN-RF. (3) For each water quality indicator, the SDE of both CEEMDAN-RF and CEEMDAN-XGB is relatively lower than that of other benchmark models, indicating that these new models have better stability.

Al-Mukhtar and Al-Yaseen [25] studied and compared three AI models: ANFIS, ANN, and MLR to simulate and forecast TDS and EC in the Abu-Ziriq marsh in southern Iraq. The study evaluated all models using three performance indices: R, RMSE, and NSE. The ANFIS model fit the observed data better than the other methods. This superior performance was attributed to ANFIS's unique structure, which combines the advantages of fuzzy logic for function simplification and neural networks for self-learning, enabling it to handle data noise effectively. The authors recommend ANFIS as a reliable predictive tool for assessing water quality parameters, TDS and EC, in Iraqi marshes. They suggest that the approaches used in this work may be applied to other marshes, rivers, and basins to test their broader applicability. In addition, the tools and results developed herein provide a sound basis for managers, engineers, and decision-makers to improve the design, management, and decision-making related to water resources in Iraq and other similar regions.

Barzegar et al. [26] used four measurable water quality parameters, such as EC, pH, oxidation-reduction potential (ORP), and water temperature, for the development of DL models comprising LSTM and convolutional neural network (CNN), with the proposition of a new hybrid model called CNN-LSTM, including some traditional ML models such as SVR and DT to predict DO and chlorophyll-a (Chl-a) concentrations in Small Prespa Lake, Greece. We collected data every 15 minutes from June 1st, 2012, to May 31st, 2013. We transformed the samples into a supervised format and divided them into training and testing sets. We trained optimal models with their hyperparameters and evaluated performance on the test set using statistical metrics: R, RMSE, MAE, and their normalized forms, RMSR and MAE. Although the CNN and LSTM models performed similarly in predicting Chl-a, LSTM outperformed CNN in predicting DO concentration. Similarly, the hybrid CNN-LSTM model, which combined the strengths of both DL models, reduced the RMAE in DO prediction by 0.24% and 1.45% over the LSTM and CNN models, respectively. For Chl-a prediction, it reduced the RMAE by 6.41% and 6.48% over the LSTM and CNN models, respectively. The CNN model performed better at lower-level predictions, while LSTM models performed better at higher-level predictions of DO and Chl-a concentrations during the testing period. By integrating these models, the hybrid CNN-LSTM can more effectively capture both low- and high-level water quality variables for predicting DO concentrations.

Baek et al. [27] combined two DL models, CNN and LSTM, to simulate the water level and three water quality parameters of total nitrogen (TN), total phosphorus (TP), and TOC in the Nakdong River basin. The CNN model was used to simulate water level, and the LSTM model was used to simulate pollutant concentrations. The following are the findings: 1. The CNN modeled the simulation of water level well; in fact, the value of NSE was 0.933, which is an acceptable performance. Water levels were higher during the rainy season and lower during the dry season. The LSTM model for all the pollutants showed an NSE greater than 0.75 during both training and validation periods, which falls under the category "very good." The model accurately captured temporal variations in each pollutant. (3) TOC and TP exhibited a similar pattern in temporal variation, highly fluctuating in the rainy season, while TN increased in the spring season. The research concludes that the combined CNN-LSTM approach may be promising for accurate simulations of water levels and quality, considering better water sustainability and management strategies. This study examined three pollutants, but most process-based models can simulate a wider suite of water quality parameters, such as chlorophyll, algae, DO, and fecal bacteria.

Uddin et al. [28] attempted to assess water quality in the Bay of Bengal using XGB and an advanced WQI method. They improved their proposed generalized model by assessing performance and sensitivity in marine water evaluation. A wide variety of statistical parameters were used for the performance assessment and comparison of the model against the observed discharges, including prediction performance metrics: RMSE, mean squared error (MSE), and MAE; model sensitivity: R^2 ; and model efficiency: NSE and MEF. From the findings of their research, a summary is extracted as follows: most of the water quality indicators from the various sampling sites in the Bay of Bengal fell within guideline values, except for DO and molybdate reactive phosphorus (MRP), where DO was below the threshold value, while MRP was above the guideline limit. Both weighted WQM-WQI and unweighted RMS-WQI models classified water quality as generally "good" or "fair", except for site A, which, from the WQM-WQI model, was consistently ranked as good. Although the models produced different WQI scores, the water-quality rankings did not

differ significantly. Performance evaluation showed that the models incorporating the XGB algorithm were efficient in performing an accurate assessment of marine water quality; furthermore, all performance indicators-RMSE, MSE, and MAE-showed reliability in reducing model uncertainty below 1% with no significant issues. Sensitivity analysis showed strong agreement between model input and output ($R^2 = 1$), but the results were sensitive to the spatiotemporal resolution of water bodies. In addition, NSE and MEF indicated that both models efficiently calculated the correct WQI score. This investigation further highlights the efficiency of both weighted WQM-WQI and unweighted RMS-WQI models for evaluating water quality in the Bay of Bengal. These findings help enhance monitoring programs and the assessment of marine water using widely adopted approaches. However, the research was limited to only ten cross-sectional sampling sites. Expanding the analysis to include additional areas alongshore and using time-series data will help indicate the models' response over larger areas. Sensitivity testing across a range of water quality attributes over an extended period would also help generalize the models for the region.

Melesse et al. [29] examined the Babol-Rood River, an important source of irrigation and drinking water in its watershed; therefore, deterioration of the water supply may lead to significant environmental and social impacts. Monitoring and predicting the river's future status requires precise models, which are considered the backbone of resource conservation and management planning. This study is among the first contributions to evaluate the performance of two independent algorithms and eight new hybrid algorithms in salinity prediction in rivers. These algorithms included M5P, RF, bagging-M5P, bagging-RF, random subspace (RS)-M5P, RS-RF, random committee (RC)-M5P, RC-RF, additive regression (AR)-M5P, and AR-RF. We used monthly data over a continuous 36-year period to develop different input variable combinations and evaluate their performance. The results showed that TDS is the most critical variable for estimating EC, with a correlation of 0.99. We tested 10 input combinations, and TDS alone proved very important. This study showed that model structure and input variables strongly influence prediction accuracy. As shown by the validation criteria, M5P had higher predictive power than RF; therefore, hybrid algorithms incorporating M5P outperformed algorithms that used RF. M5P and its hybrids, except RC-M5P, predicted EC accurately even under extreme conditions. Hybrid algorithms generally improved standalone performance, except for the RC model, which reduced predictive power when hybridized with M5P and RF algorithms. Although all algorithms performed very well, AR-M5P was the best. From most to least accurate, the ranking was: AR-M5P, RS-M5P, bagging-M5P, M5P, RC-M5P, AR-RF, bagging-RF, RF, RS-RF, and RC-RF. These results apply only to this study and cannot be generalized to other areas or hydrological data, but AR-M5P will likely be among the most accurate algorithms. Overall, the hybrid models performed better than the individual models, and it is possible to calculate river water quality using advanced ML algorithms and multiple readily available input variables.

Leggesse et al. [30] investigated the performance of six ML algorithms, including adaptive boosting (AdaBoost), RF, GB, SVM, XGB, and ANN, in modeling the retrieval of three water quality parameters, namely Chl-a, turbidity, and TDS, in data-scarce conditions at Lake Tana, where traditional monitoring methods are prohibitively expensive. They assumed that ML models trained on limited data and satellite imagery could accurately predict water quality parameters. Their results showed that ensemble ML methods, particularly XGB and RF, performed best despite limited data. Using the best model, they produced monthly water quality maps to assess spatial and temporal variability in Chl-a, turbidity, and TDS distribution. The key findings are that XGB or AdaBoost are recommended for Chl-a retrieval, while RF or XGB can retrieve TDS and turbidity from Landsat 8 OLI imagery. These ML approaches can produce fast, highly accurate water-quality maps needed for source identification, trend monitoring, and understanding influencing factors. These outputs have valuable implications for researchers and policymakers designing strategies for sustainable water management in tropical lakes and similar environments.

Nafsin and Li [31] evaluated several ML algorithms for BOD prediction in the Buriganga River system in Bangladesh. They applied four standalone ML models, such as ANN, SVM, RF, and GB, and six new hybrid models, such as RF-SVM, ANN-SVM, GB-SVM, RF-ANN, GB-ANN, and RF-GB. Feature importance analysis by RF showed that important predictors for BOD were COD, TDS, conductivity, TS, TSS, and turbidity. Among the models, RF-SVM had the highest predictive accuracy with an R^2 of 0.908.

Uddin et al. [32] presented a framework for evaluating WQI model performance in classifying water quality with precision in Cork Harbor, Ireland. The authors used four ML classifier algorithms to identify the best algorithm for predicting water quality classes. The main conclusions of the survey are as follows: (1) seven WQI models are evaluated, including four widely used and three newly proposed. (2) The XGB algorithm and K-nearest neighbor (KNN) surpassed the rest in terms of correct WQC. (3) XGB satisfactorily classified most of the WQCs, saving the class "poor." (4) weighted quadratic mean (WQM)-WQI and unweighted RMS-WQI models turned out to be good enough for the correct assessment of coastal water quality status. (5) By both models, the water quality in Cork Harbour was classified as "good" and "fair" during the period under study. The results of this study are important for developing WQI models that are more accurate, transparent, and reliable in classifying water quality along coastlines. One important limitation of the present research was the absence of consideration of temporal variability regarding water quality indicators in Cork Harbor.

Nanjappachetty et al. [33] estimated WQI by using a novel approach coupled with advanced optimization algorithms and deep capsule crystal edge graph neural networks (DCCEGNN). They applied the hybrid crested porcupine Genghis Khan shark optimization (CPGKSO) algorithm to ensure careful feature selection, using only the most relevant water quality indicators. The authors also applied the Greylag Goose optimization (GGO) algorithm to further refine the neural network's weight parameters and improve prediction accuracy. The results showed that the proposed model presents an MSE of 6.7 with 99% accuracy in WQI prediction.

Jahan and Pradhanang [34] proposed an ANN model to predict high chloride concentrations resulting from road salt and deicers

in a suburban watershed. The research used three years of data (2016-2019) from six sites to study seasonal chloride changes and differences between agricultural zones and impervious surfaces. Chloride concentrations during winter runoff were considerably higher than in other seasons. The ANN model used hydrometeorological data, stormwater runoff quality, and network parameters as input variables. We used data from 42 storm events across all four seasons for training. Sensitivity analyses identified the input variables that controlled model performance. Using the backpropagation algorithm and cross-validation, the ANN achieved about 91% prediction accuracy, with 95% confidence intervals for its outputs. These results show that storm characteristics such as duration, intensity, and runoff volume significantly impact chloride concentrations, particularly during the winter months, due to road salt application. During the non-winter seasons, chloride detection was poor and most likely due to fertilizer applications in agricultural areas. The spatial distribution analysis showed a higher percentage of chloride clustering around impervious parking lots (about 70%), while less than 50% was in agricultural zones, suggesting that road salt is the major source. ANN modeling proved a powerful tool for predicting water quality parameters like chloride and is practical and efficient for environmental management.

Overall, AI models in surface water quality modeling have shown strong potential to capture spatial and temporal variability across diverse regions. While ensemble and hybrid models such as CATBoost, GPR, and LSTM-based combinations have consistently outperformed traditional approaches in predictive accuracy, concerns about overfitting, limited interpretability, and limited generalizability persist. Notably, studies often lack details on data preprocessing, feature selection, and validation against unseen datasets, which are critical for real-world deployment. Furthermore, the lack of explainable AI (XAI) integration limits transparency in model decisions, especially in policy-making contexts. Future research should incorporate standardized benchmarking, uncertainty quantification, and cross-regional validation to strengthen the practical utility and trustworthiness of AI-driven water quality monitoring frameworks. Fig. 2 demonstrates the overall framework of surface water quality determination.

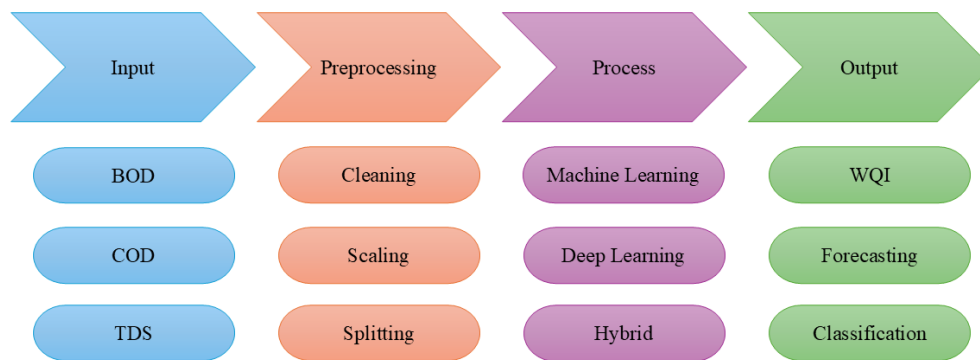


Fig. 2. Overall framework of surface water quality assessment.

Fig. 3 shows the distribution of AI models used to determine surface water quality. RF, SVM, XGB, and ANN are the most common models for determining surface water quality. LSTM, bagging-RF, ANFIS, SVR, MLR, RS-RF, and GB models rank next. Other models, including more than 30, account for 55% of all studies. Given that these models account for almost half of studies, they are considered more suitable for determining surface water quality. This may be because they can analyze large datasets with many features, where models such as XGB and RF are stronger, as well as the time-series forecasting ability of ANN and LSTM models.

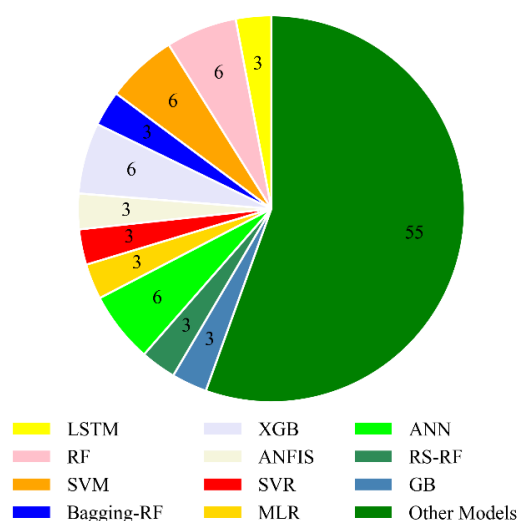


Fig. 3. Model distribution of surface water study by percent.

3.2. Groundwater

Goodarzi et al. [35] applied the world health organization (WHO) WQI and fuzzy hierarchical analysis process (FAHP)-WQI to evaluate the water quality of 96 wells in the Yazd-Ardakan plain and compared the results of both methods. In the WQI (WHO) results, 72 of the 96 wells were found to have good-quality water, and the rest were categorized as poor. In contrast, 31 wells showed good water quality when assessed with the FAHP-WQI method, 23 were poor, 22 were not usable, and 4 were excellent. In addition,

water quality classifications for agricultural purposes were also made using Wilcox and Schoeller's diagrams, respectively. From the Wilcox diagram, about 37.25% of the wells fell into the C3S2 and C3S1 classes, indicating low quality and suitability only for irrigation of specified crops in well-drained, large-textured fields. Schoeller's diagram divided drinking water quality into three classes: acceptable, bad, and inappropriate. Because calculating the WQI is time-consuming, this research estimated WQI using AI algorithms. Gene expression programming (GEP), M5P, and MARS were selected, and WQI calculations were carried out. In this study, models were evaluated using statistical indices such as Nash-Sutcliffe efficiency, RMSE, the agreement index, R, and MAE. These were then compared to identify the best model for WQI estimation. The GEP model data were relatively scattered, which is why its R was lower than that of the other models. The M5P and MARS model results are very similar. For the MARS model, during the training phase, the RMSE was 0.172, MAE 0.127, NSE 1, and R 1. During the test phase, the respective values were 0.212, 0.167, 1, 0.999, and 0.999. Although the MARS model had lower accuracy in the testing phase, it produced the best statistical indicator values among the three.

Kulisz et al. [36] carried out model studies using empirical data from 19 wells around a shale gas drilling rig in Poland. The developed ANN model used a two-layer feed-forward architecture with sigmoid hidden neurons and linear output neurons, fed by five input parameters: EC, pH, Ca, Mg, and K. Multiple regression results were used to optimize the input data. For the ANN model, an overall high R of 0.9992 and MSE of 0.0292, 0.2131, and 1.3244 for the training, testing, and validation datasets, respectively, were achieved. These were developed with dataset divisions of 70% training, 15% validation, and 15% testing using the NN fitting app and the Levenberg-Marquardt algorithm for training. The proposed ANN model provides a straightforward method for assessing groundwater pollution levels without the cumbersome calculations involved in conventional WQI methods.

Shiri et al. [37] applied eight soft computing techniques for modeling four critical groundwater quality parameters, namely EC, SAR, TDS, and SO₄, using data drawn from 90 observation wells in northwest Iran. Two modeling strategies were pursued: one predicted the targets based on other quality parameters, and the second solely based on geographical information (coordinates). The results show that all models simulated the target values appropriately, and the best among them was the coupled SVM-firefly algorithm (FFA). The models also perform well using only geographical inputs. One key takeaway is that the most appropriate k-fold testing strategy can deliver robust model performance, as obvious fluctuations were observed across different test stages. The results therefore suggest further studies using data from catchments under different climatic conditions and data availability to validate findings on comparative soft computing approaches.

Agrawal et al. [38] studied the estimation of drinking water quality using PSO and SVM. They collected groundwater samples from the Pindrawan tank command area and tested them against the developed model. They analyzed these samples for various water quality parameters, followed by WQI calculation. Some of the key inferences that can be made from the study are as follows: (1) The values of WQI indicated that 32.43% of water samples fall into the category of excellent quality, and 43.24% showed good quality. In contrast, 21.62% and 2.71% of water samples fell into the poor and very poor categories, respectively. The Raikheda pond area showed very poor quality due to high chromium levels, whereas parts of the villages of Deogaon, Dhansuli, Bangoli, Amlitalab, and Khauna showed poor quality. (2) Major cations and anions analysis revealed that all anions were within the acceptable limits except potassium, with 13% of the samples exceeding the limit. Future heavy metal pollution from mining activities is expected, as 48.6% of samples in this study exceeded chromium limits, potentially causing health problems such as auditory problems, blood diseases, high blood pressure, and even death. For the estimation and prediction of WQI values accurately, the study proposed ensemble ML algorithms. Specifically, an ensemble method with PSO combined with a Naive Bayes classifier (NBC) was the best algorithm, achieving 92.8% accuracy in predicting WQI indices. This will help estimate WQI and, through the user interface, save time and effort. Overall, the research demonstrates the benefits of ensemble ML techniques by combining outcomes from multiple algorithms to improve overall accuracy. The developed algorithm can be applied globally to estimate water quality using a user interface. This study used a synthetic dataset developed with PSO; more real data would improve accuracy. The authors proposed using larger datasets, which could be obtained from the field, but this is cumbersome due to sampling and laboratory analysis. In addition, the algorithm could be further improved by fine-tuning based on performance studies using different input parameters.

Mattas et al. [39] focused on predicting irrigation water salinity, represented by EC, in the aquifers of the Gallikos River basin, a catchment area of intensive agricultural activities and small-scale enterprises vulnerable to pollution. In this framework, we developed ANN and MLR predictive models, leveraging their simplicity, satisfactory performance, and low input-data requirements. We used data from 233 samples collected over three years (2004–2006) to calibrate, validate, and evaluate the models. ANN and MLR models both showed strong predictive accuracy with the large dataset, as shown by high evaluation indices: $R^2 > 0.927$, modeling efficiency > 0.93 , and MAPE $< 14.5\%$. For smaller verification datasets comprising 18 samples, both models remained statistically reliable: $R^2 > 0.75$, modeling efficiency > 0.976 , and MAPE $< 20\%$, although MLR performed slightly better. We selected input parameters for both models using factor analysis to improve accuracy and reliability. These results highlight the practical relevance of ANN and MLR models for cost-effective, efficient water quality parameter forecasting, such as EC, and reduce the need for expensive, time-consuming in situ measurements. The present study provided a methodological framework applicable to other regions facing similar environmental challenges, offering actionable insights for parameter selection and model evaluation in water resource management.

Singha et al. [40] proposed a DL algorithm. They compared it with three other traditional ML models, such as RF, XGB, and ANN, for estimating the groundwater quality index (EWQI) in India. These metrics were then used to assess the performance and prediction capability of the models, and the results showed that the deep learning model performed best with the highest validation accuracy: MSE = 1.537, MAE = 1.360, RMSE = 1.254, RSR = 0.034, NSE = 0.987, MAPE = 0.959, and $R^2 = 0.996$. In comparison,

XGB performed better than both the RF and ANN models. Based on validation R^2 values, the performance ranking was $DL > XGB > ANN > RF$, with values of 0.996, 0.927, 0.917, and 0.886, respectively. These results demonstrate the DL model's efficiency in selecting the most influential input variables to compute EWQI. Therefore, the proposed DL model is suitable for the study area to evaluate the quality of the groundwater and can be applied in different water studies for managing the quality prediction parameters, including BOD, DO, and COD, and mainly for the removal of pollutants and toxic trace elements such as Pb, Cd, and Mn from surface water and groundwater. Here, the study is limited to a single monsoon dataset. If the dataset included several seasons, the results could be more informative about groundwater quality.

Zamili et al. [41] estimated the WQI of the Yesilirmak River in Çorum, Turkey, in the monitoring period from 1995 to 2014. Eleven physicochemical parameters were subjected to computation for WQI. Results ranged between good and poor; it was once found unsuitable for drinking in June 2000. This research also applied several new hybrid models in which pre-processing data and the ANN approach were optimized by different metaheuristic methods, namely coefficient-based PSO and chaotic gravitational search algorithm (CPSOCSA), marine predator's optimization algorithm (MPA), and PSO for WQI prediction. Furthermore, to improve model precision, in addition to WQ parameters, some climate factors, mainly rainfall, were considered predictors. Two key points emerged from the results of this study. First, the pre-processing methodologies improved the dataset and provided an appropriate set of predictors. Second, the CPSOCSA-ANN model outperformed the other models, including MPA-ANN and PSO-ANN, with $R^2 = 0.965$, MAE = 0.01627, and RMSE = 0.0187.

Sinha et al. [42] investigated the suitability of ANNs for modeling WQI in the region affected by industrial wastewater contamination in Jodhpur district, India. They assumed that water quality, controlled by multiple parameters such as pH, TDS, and TS, could be described by a single composite index that captures the overall impact of contamination. ANN modeling supported this hypothesis by reliably predicting the relationships among individual water quality parameters and overall WQI. The results presented a new stochastic WQI protocol applied to characterize water quality in Jodhpur, India, and concluded that potable water quality was generally good (~60% or above) despite industrial contamination. The results proved the efficiency of ANNs in modeling complex nonlinear interactions among water quality parameters and predicting their impact on potability. Although the findings agree with previous studies, the authors acknowledge limitations such as the lack of alternative algorithms and architectural comparisons.

El Bilali et al. [43] conducted a study to verify whether ML models efficiently predict irrigation water quality (IWQ) parameters using physical variables as input data, and to assess the sensitivity of ML models to these inputs. The study area was the Berrechid aquifer in north-west Morocco, between the cities of Settat and Casablanca. We developed four models to predict TDS, potential salinity (PS), SAR, exchangeable sodium percentage (ESP), magnesium adsorption ratio (MAR), and residual sodium carbonate (RSC) from their physical features using the AdaBoost, RF, ANN, and SVR algorithms. The results showed that the AdaBoost model was the best for all parameters, while RF was the second most successful, predicting five parameters except MAR. It was observed that the ANN and SVR models could predict the four fitness parameters TDS, PS, SAR, and RSC, and three fitness parameters, PS, SAR, and RSC, respectively. Moreover, all ensembling models (AdaBoost and RF) performed better than the conventional ANN and SVR models in all validation tests. ANN and SVR generalized better than AdaBoost and RF. In addition, Sensitivity analysis showed that the models are relatively insensitive to input variables across the range of predicted parameters. In general, ML models can efficiently forecast IWQ parameters and provide real-time, low-cost monitoring of irrigation water quality.

Overall, AI techniques in groundwater quality assessment have shown considerable promise, particularly in data-scarce and geochemically complex regions. Studies reviewed in this section have shown that hybrid models such as CNN-LSTM and ensemble-based approaches like MARS and M5P can achieve high predictive accuracy across diverse groundwater settings. These models can effectively handle nonlinearity and temporal variability in hydrochemical parameters. However, several limitations persist. Many studies do not address missing data, measurement uncertainty, and groundwater-specific features such as aquifer heterogeneity or seasonal recharge impacts. Furthermore, while ANNs and DL architectures achieve strong statistical performance, their limited interpretability hinders transparency for policy-level decision-making. Future research should prioritize integrating explainable AI (XAI) methods, robust validation across hydrogeological zones, and improved uncertainty treatment through probabilistic frameworks. Moreover, coupling AI models with GIS-based spatial tools could enhance spatiotemporal monitoring and support the design of sustainable groundwater management strategies [44–51]. Fig. 4 demonstrates the overall framework of groundwater quality determination.

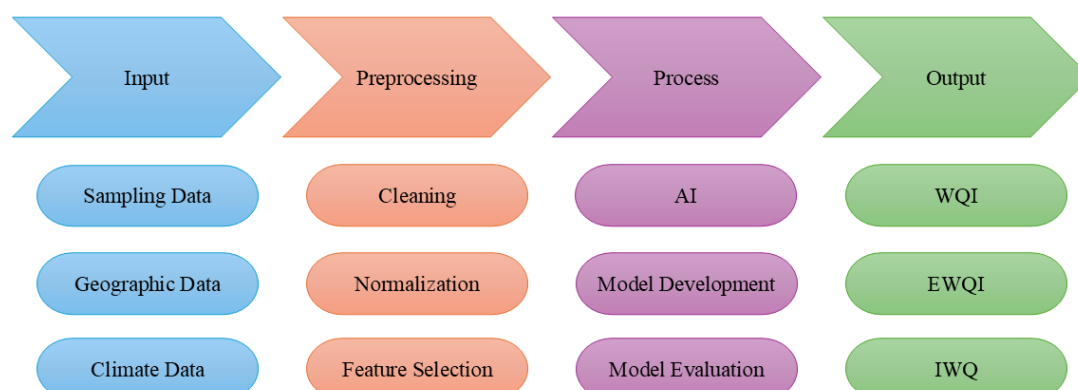
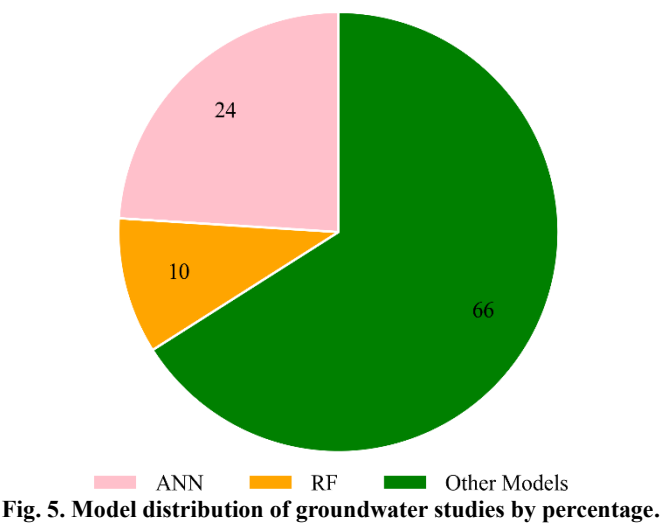


Fig. 4. Overall framework of groundwater quality assessment.

Fig. 5 shows the distribution of AI models used to determine groundwater quality. ANN and RF account for 24% and 10% of studies, respectively, and other models (more than 12) account for 66%. As mentioned, these models are preferred for groundwater quality because they effectively solve nonlinear relationships, handle missing data, and also provide high accuracy.



4. Comparative analysis and discussion

In this review, many AI methods have been used to determine water quality and have been published in recent years. The results indicate that different AI models have been successfully applied to predict water features across different geographical regions. Each model has its strengths and weaknesses. Fig. 6 provides a methodology with advantages and disadvantages. Table 2 illustrates the classification of AI by family.

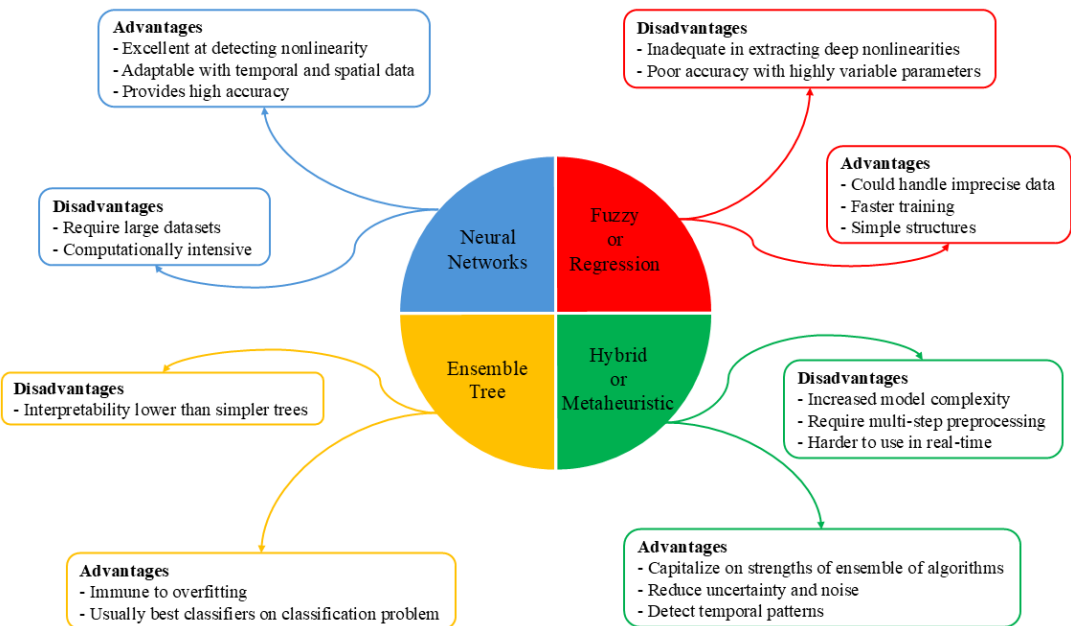


Fig. 6. Advantages and disadvantages of each methodology.

Table 2. Methodological families.

Method family	Artificial intelligence
Neural networks	ANN, MLP, CNN, LSTM, DL, CNN-LSTM, BPNN, MLPNN, RBFNN, ANN-WT-LSTM, DCCEGNN, NAR
Ensemble tree-based	RF, XGB, CatBoost, AdaBoost, GB, ETR, DT, RT, REPT, EN, bagging-RF, bagging-M5P, CVPS-REPT, RS-RF, RS-M5P, RC-RF, RC-M5P, AR-RF, AR-M5P
Hybrid/metaheuristic	CEEMDAN-XGB, CEEMDAN-RF, PSO, EFO, FFA, WCA, SCA, CPSOCGSA, MPA, CPGKSO, GGO, teaching-learning-based optimization
Fuzzy/regression-based	ANFIS, MLR, GEP, MARS, GPR, SLR, LR, SVM, KNN, SVR, DTR

As the tables and review show, no universal AI model performs best for all water quality parameters or under all conditions. Each algorithm is sensitive to the dataset's nature, the degree of nonlinearity, and the number of inputs. Therefore, hybrid and ensemble approaches have gained popularity in recent years, aiming to combine the strengths of multiple algorithms. Furthermore, studies show that preprocessing techniques such as decomposition (e.g., wavelet, CEEMDAN) and optimization (e.g., PSO, GWO) can significantly enhance model accuracy by reducing noise and selecting the most relevant features. These steps are essential in improving the overall modeling pipeline, not just the choice of algorithm.

Another general observation is the wide geographical and environmental diversity of study areas-from canals and lakes in South Asia to marshes in Iraq and groundwater aquifers in Europe. While surface water dominates the literature, AI use in groundwater quality modeling is also growing, particularly in regions where traditional monitoring is costly or logistically difficult. However, few studies address uncertainty quantification, real-time adaptability, or model explainability, limiting operational deployment in decision support systems. Future research should aim to address these gaps by integrating explainable AI methods, uncertainty analysis, and dynamic updating from sensor-based systems.

4.1. Critical factors affecting AI model performance

AI model performance in water quality prediction depends not only on algorithm structure but also on dataset characteristics, environmental conditions, input variables, and hydrological complexity. The reviewed studies demonstrate that hybrid and ensemble models generally outperform single algorithms because they combine different learning mechanisms and reduce uncertainty. For example, RF-SVM achieved superior BOD prediction performance by integrating feature selection capability with nonlinear classification strength, while CNN-LSTM models improved prediction by simultaneously capturing spatial and temporal patterns. Dataset size and quality are among the most important factors affecting model reliability. Studies using long-term monitoring datasets, such as the 36-year river salinity dataset analyzed by Melesse et al. [29], achieved higher generalization than studies based on limited seasonal datasets. However, large datasets alone do not guarantee better performance; preprocessing, noise reduction, feature selection, and appropriate validation procedures significantly influence the final accuracy. Hydrological characteristics also control model suitability. Surface water systems generally involve strong temporal variability caused by rainfall, runoff, and pollution events, making recurrent models such as LSTM more suitable for time-series prediction. Conversely, groundwater quality modeling is influenced by aquifer heterogeneity, recharge conditions, and geochemical interactions; therefore, models that handle nonlinear relationships, such as ANN, ANFIS, and hybrid approaches, have performed better.

The number and type of input parameters also affect prediction accuracy. Several studies demonstrated that feature selection can reduce unnecessary variables without reducing model performance. For instance, some WQI models achieved reliable predictions using only key parameters such as BOD, turbidity, and phosphate, rather than many measurements. Climatic conditions further influence AI model applicability. Models developed in humid, tropical, arid, or highly seasonal environments may not directly transfer to other regions because hydrological processes, pollutant transport mechanisms, and seasonal variability differ significantly. Therefore, regional validation and uncertainty analysis remain essential for developing transferable AI-based water quality prediction systems.

4.2. Bibliometric analysis of reviewed studies

To provide an additional perspective on the evolution and research trends of AI-based water quality prediction, we conducted a bibliometric analysis based on the 35 studies selected through the PRISMA procedure. The analysis focused on publication trends and keyword co-occurrence patterns to identify the field's development trajectory and dominant research themes.

Fig. 7 presents the annual distribution of the reviewed publications. The results indicate a continuous research interest in AI applications for water quality assessment, with a noticeable increase in publications after 2020. This growth corresponds with the rapid development of machine learning, deep learning, and hybrid modeling approaches in environmental and hydrological applications. The growing number of studies highlights the shift from conventional statistical approaches to data-driven prediction frameworks.

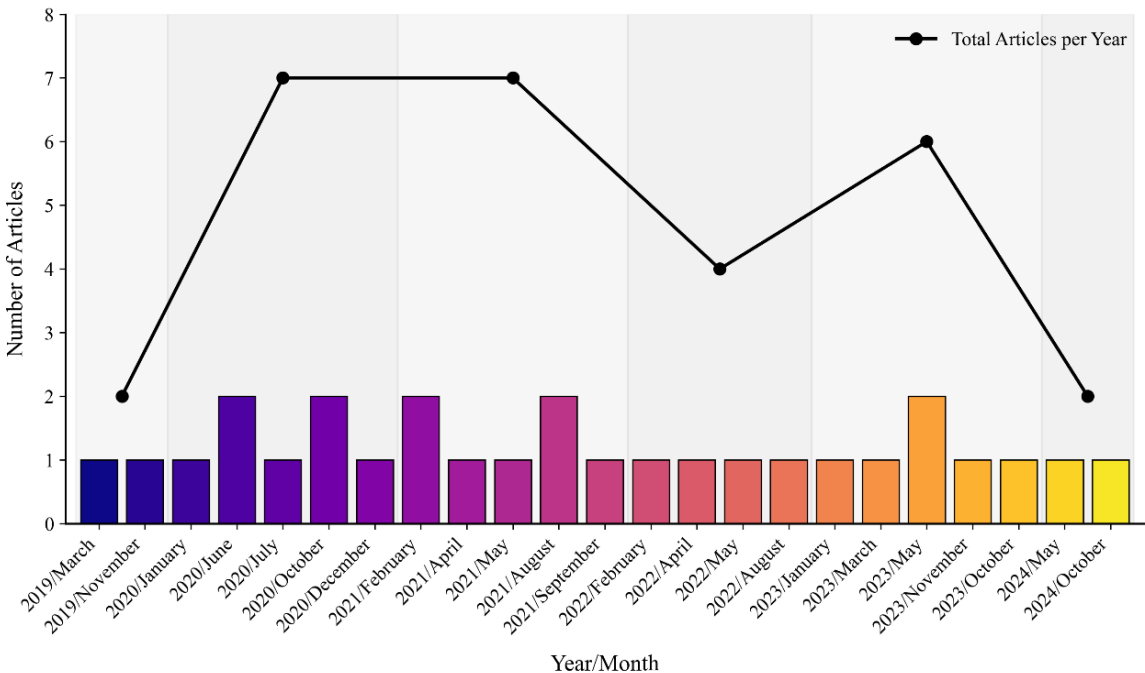


Fig. 7. Annual publication trend of AI-based water quality prediction studies included in the review.

We performed keyword co-occurrence analysis to identify major research themes and relationships among frequently used terms. As shown in Fig. 8, "water quality index," "machine learning," "artificial neural network," and "water quality" are central research themes with strong connections to various modeling approaches, prediction objectives, and environmental applications. The network shows the field's evolution from individual machine learning applications to integrated frameworks involving hybrid algorithms, ensemble learning, uncertainty analysis, and advanced optimization techniques.

The observed clusters indicate several major research directions. One cluster includes AI modeling approaches such as artificial neural networks, random forests, support vector machines, and deep learning methods. Another cluster focuses on water quality indicators and environmental applications, including groundwater quality, electrical conductivity, total dissolved solids, and water quality indices. A further emerging theme relates to hybrid modeling, feature selection, optimization algorithms, and explainable prediction frameworks.

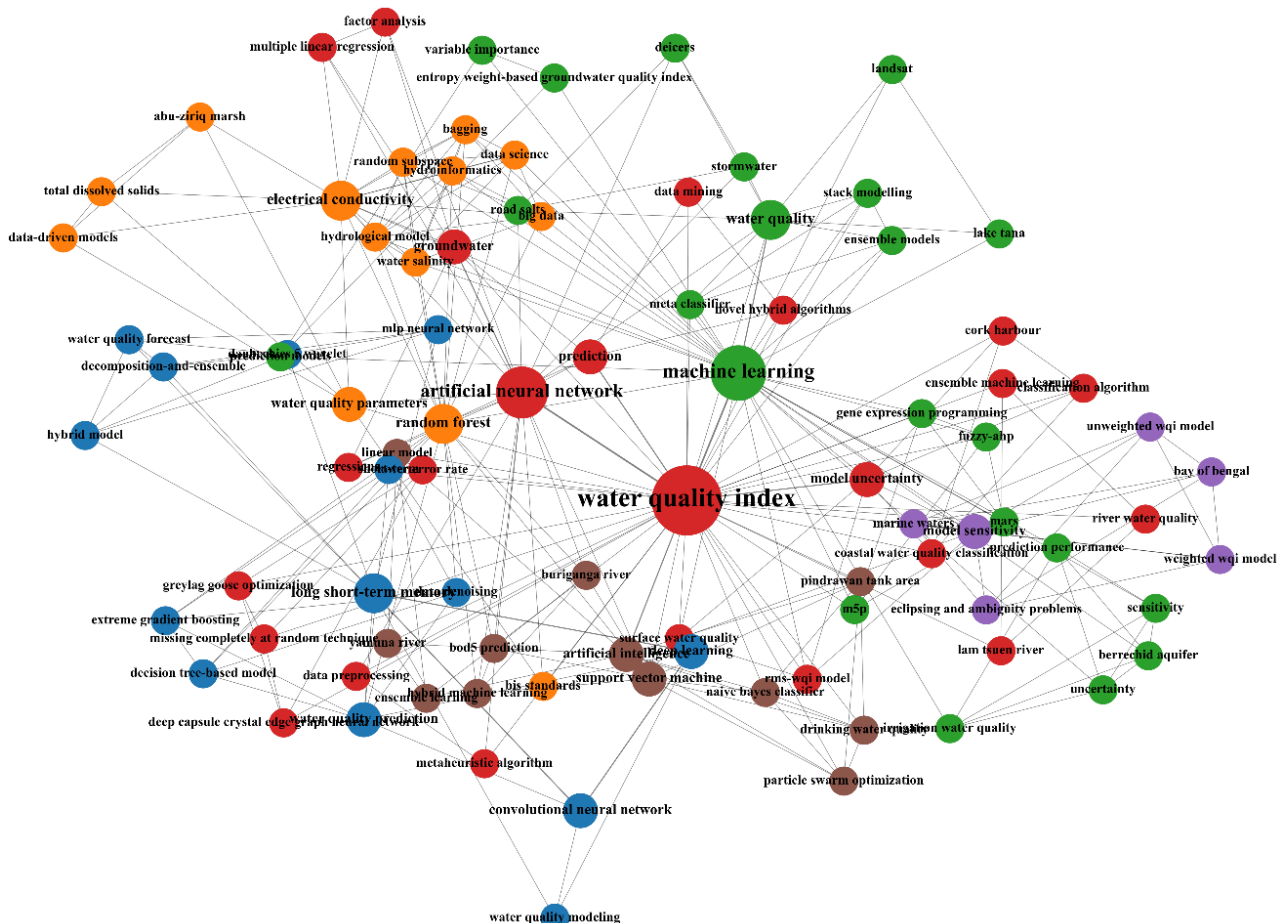


Fig. 8. Geographical distribution of reviewed studies.

Because the number of included studies is relatively small, we did not perform advanced citation-based network analyses such as institutional collaboration and co-citation mapping. Fig. 9 shows the geographical distribution of reviewed studies. Moreover, Fig. 10 demonstrates the distribution of reviewed studies by journal.

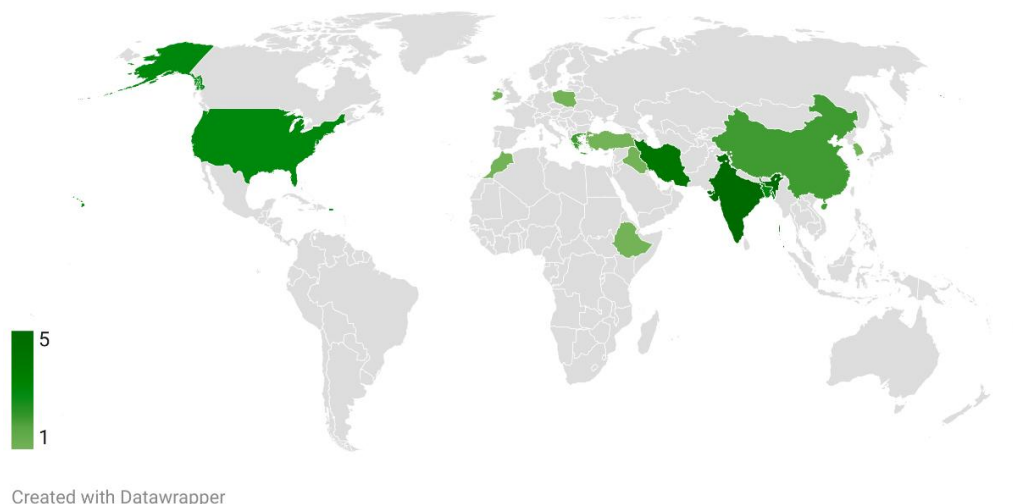


Fig. 9. Geographical distribution of reviewed studies.

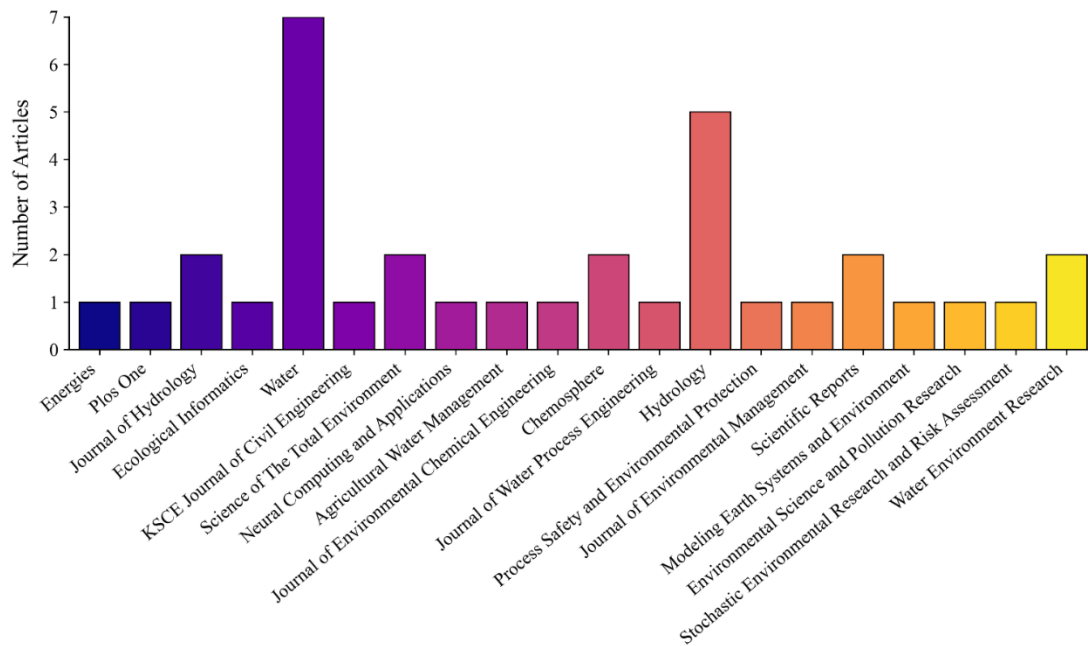


Fig. 10. Distribution of reviewed studies by journal.

4.3. Study distribution

This review showed that about 65% discussed surface water and 35% discussed groundwater. Methodologically, about 68% used neural networks, 46% used tree ensemble models, 29% used hybrid or metaheuristic approaches, and 50% used fuzzy logic or regression methods, with some overlap because each study used different methods. Fig. 11 indicates this statistic.

The distribution of groundwater studies and surface water studies is as follows: neural network models were used in 68% of surface water and 67% of groundwater studies; tree ensemble models were used in 53% of surface water and 33% of groundwater studies; hybrid or metaheuristic approach models were used in 26% of surface water and 33% of groundwater studies; fuzzy logic or regression methods were used in 47% of surface water and 56% of groundwater studies. Fig. 12 shows the distribution of method families in surface water and groundwater studies. Table 3 also lists each study and its method family.

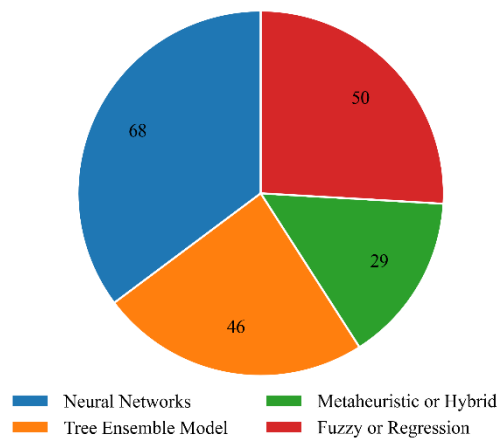


Fig. 11. Study distribution by methodological families by percent.

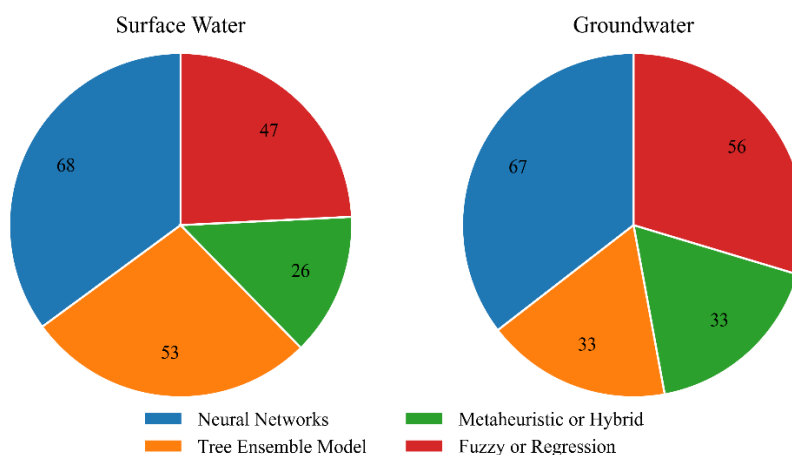


Fig. 12. Method family in surface water and groundwater by percent.

Table 3. Research distribution by each method family.

Method family	Surface water	Groundwater
Neural networks (ANN, MLP, CNN, LSTM, DL, CNN-LSTM, BPNN, MLPNN, RBFNN, ANN-WT-LSTM, DCCEGNN, NAR)	Wang et al. [16]	
	Nasir et al. [17]	
	Yang [20]	
	Abba et al. [21]	
	Wu and Wang [22]	Kulisz et al. [36]
	Lu and Ma [24]	Mattas et al. [39]
	Al-Mukhtar and Al-Yaseen [25]	Singha et al. [40]
	Barzegar et al. [26]	Zamili et al. [41]
	Baek et al. [27]	Sinha et al. [42]
	Leggesse et al. [30]	El Bilali et al. [43]
	Nafsin and Li [31]	
	Nanjappachetty et al. [33]	
	Jahan and Pradhanang [34]	
	Nasir et al. [17]	
Ensemble tree-based (RF, XGB, CatBoost, AdaBoost, GB, ETR, DT, RT, REPT, EN, bagging-RF, bagging-M5P, CVPS-REPT, RS-RF, RS-M5P, RC-RF, RC-M5P, AR-RF, AR-M5P)	Bui et al. [18]	
	Sajib et al. [19]	
	Asadollah et al. [23]	
	Lu and Ma [24]	Goodarzi et al. [35]
	Uddin et al. [28] - Bay of Bengal	Singha et al. [40]
	Melesse et al. [29]	El Bilali et al. [43]
	Leggesse et al. [30]	
	Nafsin and Li [31]	
	Uddin et al. [32] - Ireland	
	Nasir et al. [17]	
Hybrid/metaheuristic (CEEMDAN-XGB, CEEMDAN-RF, PSO, EFO, FFA, WCA, SCA, CPSOCGSA, MPA, CPGKSO, GGO, teaching-learning-based optimization)	Yang [20]	Shiri et al. [37]
	Lu and Ma [24]	Agrawal et al. [38]
	Nafsin and Li [31]	Zamili et al. [41]
	Nanjappachetty et al. [33]	
	Nasir et al. [17]	
	Sajib et al. [19]	
Fuzzy/regression-based (ANFIS, MLR, GEP, MARS, GPR, SLR, LR, SVM, KNN, SVR, DTR)	Abba et al. [21]	Goodarzi et al. [35]
	Asadollah et al. [23]	Shiri et al. [37]
	Lu and Ma [24]	Agrawal et al. [38]
	Al-Mukhtar and Al-Yaseen [25]	Mattas et al. [39]
	Leggesse et al. [30]	El Bilali et al. [43]
	Nafsin and Li [31]	
	Uddin et al. [32] - Ireland	

Table 3 shows that, by priority, popular methods for surface water are neural networks, ensemble tree-based methods, fuzzy/regression-based methods, and hybrid/metaheuristic methods. For groundwater, the trend is neural networks, fuzzy/regression, and metaheuristic/tree-based methods, respectively. This popularity demonstrates the power of neural networks in determining water quality in both surface water and groundwater because they can model nonlinear relationships. This classification is intended only to provide a general idea of which AI family is more commonly applied to computing surface water or groundwater parameters, not a specific water quality parameter. This is because, when computing a specific parameter such as DO or TDS, a different type of AI can be more efficient. This is because the degree of linearity or non-linearity in each water parameter's variation differs. Since each AI model specializes in a certain kind of computation, it provides more accurate results for that parameter.

4.4. Suitable AI methods by water quality parameter

- On WQI, different ML models have estimated surface and groundwater with high accuracy. In surface water, GPR had the lowest uncertainty ($< 2\%$) and the highest R^2 in inland canal studies, such as the Tongi canal, because it effectively estimates predictive uncertainty in ecological data. Additionally, the CEEMDAN-XGB model performed better at identifying both short-term and long-term trends by denoising and then boosting, with MAPEs as low as 0.27% for pH and 14.94% for turbidity. MARS and M5P hybrid models in groundwater analysis achieved near-perfect training and testing correlation ($R \approx 1.00$) using linear splines to model sudden changes in well chemistry. CPSOCGSA-ANN also outperformed other metaheuristic approaches, achieving $R^2 = 0.965$ when optimized for long-term river monitoring.
- RF-SVM hybrid ($R^2 = 0.908$) for Bangladesh's BOD was optimally combined using RF's strength and SVM's margin maximization of nonlinear relationships.
- In DO determination, the EFO-MLPNN hybrid required fewer iterations and agents but achieved improved RMSE and NSE by means of efficient network weight search space exploration. LSTM-hybrids based on ANN-WT-LSTM also improved short-term DO trend predictions using wavelet decomposition to eliminate noise.
- In pH and turbidity, CEEMDAN-XGB reduced noise effects to reach MAPEs $< 1\%$ in pH prediction and dominated extreme turbidity oscillations.
- In EC and TDS columns, ANFIS applied fuzzy logic to reduce input uncertainties in marsh and marshland conditions to surpass MLR and ANN in predicting TDS and EC.
- In Chl-a, TN, and TP assessment, CNN models retained high-resolution spatial patterns in spectral data, while hybrid CNN-

LSTM added temporal dynamics to reduce RMAE by up to 6.5% on Chl-a. Additionally, XGB and RF also performed effectively for Chl-a prediction.

4.5. Structured overview of AI applications in water quality modeling

To support this analysis, Table 4 categorizes each reviewed study by AI model, key advantages and limitations, water resource type, application area, accuracy level, usage location, and suitability.

Table 4. Summary of AI-based water quality prediction models with comparative advantages, limitations, and suitability across various hydrological contexts.

Artificial intelligence (AI) model	Advantages	Disadvantages	Water resource	Application area	Accuracy	Usage location	Suitability	Ref. No.
LSTM	Predicts industrial pollution impact	Limited to a single-point prediction	Surface water	River water quality management	High accuracy in predicting water quality trends	Industrial river basins, China	Pollution management	Wang et al. [16]
CATBoost, RF, MLP, SVM, DT	High accuracy, improved predictions	Model complexity	Surface water	WQC	CATBoost: 94.51%, Stacking: 100%	India	Complex analyses	Nasir et al. [17]
BA-RT, bagging-RF, CVPS-REPT	Better performance than standalone models	Poor performance with extreme WQI values	Surface water	WQI prediction	BA-RT: Highest performance	Mazandaran and Semnan provinces in Iran	Groundwater management	Bui et al. [18]
GPR, XGB, SVM, NN	Best performance for WQI	Limited to a single water body	Surface water	WQI prediction	Best in RMSE, MAE, R ²	Tongi canal, Bangladesh	Urban water quality analysis	Sajib et al. [19]
EFO, WCA-MLPNN, SCA	High accuracy, faster computation	Requires many iterations for EFO	Surface water	DO prediction	Superior in RMSE, RP, NSE	Klamath River, Oregon, USA	DO prediction	Yang [20]
BPNN, ANFIS, SVR, MLR	Improved accuracy with ensemble models	Performance varies by location	Surface water	WQI prediction	ANFIS: 30-53% improvement in absolute error	Yamuna River, India	WQI prediction in rivers	Abba et al. [21]
ANN-WT-LSTM, ANN-LSTM	High accuracy, long-term, and short-term prediction	Limited spatial variation predictions	Surface water	Water quality prediction	Outperforms ANN-LSTM	Jinjiang River, China	Suitable for early warning systems	Wu and Wang [22]
ETR, SVR	Higher accuracy than SVR and DTR	Sensitive to input variables	Surface water	WQI prediction	Best with three inputs	Lam Tsuen River, Hong Kong	Real-time WQI monitoring	Asadollah et al. [23]
CEEMDAN-XGB, CEEMDAN-RF, PSO-SVM, RBFNN	High accuracy, model stability	Limited to the water quality of the specific site (Gales Creek)	Surface water	Water quality prediction	MAPE: 0.69%-14.94%	Gales Creek, Oregon, the USA	Water quality monitoring	Lu and Ma [24]
ANFIS, ANN, MLR	Provides robust predictive frameworks for TDS and EC	Higher complexity of ANFIS structure	Surface water	Water quality (TDS, EC) in marshes	Best for ANFIS outperforms others	Abu-Ziriq Marsh, Iraq	Designing and managing water resource systems	Al-Mukhtar and Al-Yaseen [25]
CNN-LSTM, DL	Combines strengths of CNN and LSTM	Limited in high-level performance	Surface water	DO and Chl-a prediction	RMAE: 0.24% improvement in DO	Small Prespa Lake, Greece	Multi-level prediction	Barzegar et al. [26]
CNN, LSTM	Accurate temporal modeling	Limited to three pollutants	Surface water	Water level and pollutant simulation	NSE > 0.75	Nakdong River basin, South Korea	Pollutant and water level simulation	Baek et al. [27]
XGB	Efficient marine water assessment	Limited spatial coverage	Surface water	Coastal water quality assessment	RMSE, MAE < 1%, no ambiguity	Bay of Bengal	Marine water quality monitoring	Uddin et al. [28]
M5P, RF, bagging-M5P, bagging-RF, RS-M5P, RS-RF, RC-M5P, RC-RF, AR-M5P, AR-RF	Better than standalone models	Inconsistent performance when hybridized	Surface water	River salinity prediction	R = 0.99	Babol-Rood River, Iran	River salinity prediction	Melesse et al. [29]
AdaBoost, RF, GB, SVM, XGB, ANN	High performance in data-scarce environments	Limited to satellite imagery	Surface water	Modeling Chl-a, turbidity, and TDS	High (XGB and RF recommended)	Lake Tana, Ethiopia	Quick, reasonably accurate water quality monitoring in tropical lakes	Leggesse et al. [30]

ANN, SVM, RF, GB, RF-SVM, ANN-SVM, GB-SVM, RF-ANN, GB-ANN, RF-GB	High prediction accuracy, cost-effective, systematic data management	Requires significant computational resources	Surface water	BOD prediction	$R^2 = 0.908$	Buriganga River, Bangladesh	Developing countries with pollution control needs	Nafsin and Li [31]
XGB, KNN	High accuracy in classification	Poor performance in the "poor" category	Surface water	Water quality classification	XGB is best in most classes	Cork Harbor, Ireland	Coastal water quality	Uddin et al. [32]
DCCEGNN, CPGKSO, GGO	Extremely high accuracy, robust optimization, precise feature selection	Complex model requiring advanced expertise	Surface water	WQI prediction	99% (MSE = 6.7)	Global freshwater monitoring	Environmental authorities managing river quality	Nanjappachetty et al. [33]
ANN	High prediction accuracy	Requires significant hydrometeorological data	Surface water	Chloride prediction in watersheds	Prediction accuracy: 91%	Suburban watershed, the USA	Evaluating the impact of road salt and seasonal runoff on chloride levels	Jahan and Pradhanang [34]
GEP, MSP, MARS	High accuracy during training	Lower accuracy in the testing phase	Groundwater	Groundwater quality prediction	MARS best in statistical indicators	Yazd-Ardakan Plain, Iran	Groundwater WQI estimation	Goodarzi et al. [35]
ANN	High R	Limited dataset	Groundwater	Groundwater pollution assessment	$R = 0.9992$	Shale gas site, Poland	Groundwater pollution monitoring	Kulisz et al. [36]
SVM-FFA	Accurate predictions with geographical inputs	Fluctuations in test stages	Groundwater	Groundwater quality modeling	Best with SVM-FFA	Northwest Iran	Catchment area quality analysis	Shiri et al. [37]
PSO, SVM	High prediction accuracy	Limited real dataset	Groundwater	WQI estimation	Accuracy: 92.8%	Pindrawan, India	Drinking water quality estimation	Agrawal et al. [38]
ANN, MLR	Cost-effective and time-saving	Results depend on data availability	Groundwater	Irrigation water salinity prediction	$R^2 > 0.927$	Gallikos River basin, Greece	Forecasting water salinity efficiently with minimal resources	Mattas et al. [39]
DL, XGB, RF, ANN	Highest validation accuracy	Limited dataset (monsoon season only)	Groundwater	Groundwater quality prediction	DL accuracy: $R^2=0.996$	India	Pollutant and toxic element prediction	Singha et al. [40]
CPSOCGSA-ANN, MPA-ANN, PSO-ANN	Enhanced dataset preprocessing	Limited consideration of climate factors	Groundwater	WQI prediction	CPSOCGSA-ANN: $R^2 = 0.965$	Yesilirmak River, Turkey	WQI prediction using metaheuristics	Zamili et al. [41]
ANN	Introduced a stochastic WQI protocol	Limited exploration of the alternative model	Groundwater	Industrial wastewater regions	Satisfactory for WQI	Jodhpur, India	Assessing water potability despite contamination from industrial sources	Sinha et al. [42]
AdaBoost, RF, ANN, SVR	Efficient in IWQ prediction	ANN and SVR generalize less effectively	Groundwater	IWQ prediction	AdaBoost outperforms others	Irrigation systems in Morocco	Real-time IWQ monitoring	El Bilali et al. [43]

By analyzing Table 4, it can be concluded that:

- LSTM model: From the results, this model appears to provide accuracy in predicting industrial water pollution impacts, particularly for river water quality management. This makes it well suited for managing or controlling pollution in industrial river basins, since it can precisely predict water-quality trends.
- CATBoost RF stacking ensemble: Despite the complexity of this ensemble method, it has reached an incredible accuracy level regarding WQC tasks. For example, this stacking model reached 100% accuracy in India. Therefore, this model may prove extremely effective for complex water-quality analysis.
- Hybrid models - BA-RT: Bagging-RF - hybrid algorithms are better independent models, especially useful for WQI predictions in areas where routine monitoring is scarce. However, scores tend to drop when dealing with extreme WQI values, which is challenging.
- GPR: The model performed best in predicting WQI for urban areas, as in the case of the Tongi canal in Bangladesh, with the highest accuracy for different statistical metrics such as RMSE, MAE, and R^2 . This model has proved highly reliable for assessing urban water quality.

- Results showed that AI and ML models have successfully predicted water quality and related environmental parameters beyond expectations. However, each model had its respective limitations. For instance, LSTM models were bound to single-point prediction, making them inappropriate for large-scale or multi-point water monitoring systems. On the other hand, hybrid models such as BA-RT may perform well in areas with no regular monitoring; however, they perform poorly for extreme WQI values.
- Advanced models such as CATBoost, RF, and stacking ensemble perform much better in the classification task, especially regarding water quality studies in such complicated environments. Yet this advantage often comes with increased computational complexity, which can raise costs and resource demands.
- Generally speaking, the region, characteristics of a dataset, and set goals will determine which model is most appropriate. For instance, CNN-LSTM models are well-suited for multi-level predictions. In contrast, models such as XGB and RMS-WQI are sufficiently powerful for accurate assessment of marine water quality in coastal regions.
- The analysis pinpointed that model selection should be carried out concerning unique characteristics pertinent to the task of monitoring water quality.

In summary, model selection for water quality monitoring should always consider the unique dynamics of the target parameter, the spatial-temporal scale, and the availability of data and computational resources.

4.6. Toward a unified analytical framework

To enhance water quality modeling, we propose a hybrid explainable AI-spatiotemporal decomposition (XAI-STD) framework that integrates data decomposition, feature selection, model ensemble, interpretability, and real-time adaptability. The framework begins with data preprocessing and multiscale decomposition, utilizing methods such as WT or CEEMDAN to remove noise and extract intrinsic signal patterns across temporal and spatial resolutions.

It then selects the most significant features using metaheuristic optimizers such as PSO or GWO to optimize the pool of physicochemical variables and remote sensing indicators chosen for the model's input stage. This feature selection optimizes the input variable set. In the modeling stage, a hybrid model approach is taken, using an ensemble of three algorithmic approaches with different specialties:

An uncertainty estimator probabilistic regressor model (e.g., Gaussian process regression, GPR), which models both the predictive value and the confidence intervals for this prediction; an ensemble tree-based machine learning model (e.g., XGBoost) with nonlinear boosting capabilities; and a recurrent neural network (e.g., LSTM) to account for potential temporal dependencies or temporal data patterns.

Finally, real-time responsiveness and adaptability are sought through learning algorithms with incremental or adaptive retraining methods, using a continuous data stream from IoT sensors to keep model performance up to date. In summary, the proposed XAI-STD modeling framework is expected to be able to:

Represent the nonlinear spatiotemporal dependencies of water quality variables in an open water reservoir, explain its predictions transparently and intuitively for all three input feature types, and remain adaptive and responsive for real-time updates in water quality monitoring.

This modeling framework is computationally intensive, requires domain and machine learning expertise, and involves implementation costs. Nonetheless, it may suit high-stakes water resource management and optimization use cases where a compromise between prediction accuracy, explainability, and adaptability is not possible.

5. Conclusions

In this literature review, we analyzed recent works that used artificial intelligence (AI) techniques for water quality prediction. The papers selected for this analysis were published in recent years, as the focus on implementing AI methods for hydrological and environmental modeling has increased with the recent development of new algorithms. The insights extracted from the analysis of the selected literature can be summarized as follows:

Most analyzed studies used hybrid approaches that combined multiple models to better perform the task at hand. Although success rates varied by model, target parameter, and scenario, most approaches achieved satisfactory, reliable results, indicating the area's maturity. Some AI algorithms also consistently outperformed others for certain water quality indicators. Based on our analysis, the most commonly modeled parameters were total dissolved solids (TDS), Biochemical Oxygen Demand (BOD), and pH, all significant indicators of water quality and likely the most popular "benchmark variables".

Among the studies included in this review, neural-network approaches were the most frequently represented methodological family. Neural networks, in general, are among the most robust at capturing the nonlinear and multivariate relations, but tend to be less interpretable and more computationally expensive. Ensemble tree-based approaches (RF, XGB, etc.) were the second most used AI method, used in 53% of surface water studies and 33% of groundwater studies, offering a balanced approach in terms of accuracy and overfitting resistance, as well as explainability through feature importance.

Metaheuristic and hybrid models appear more effective for noisy time series datasets, and optimization-guided feature selection

was confirmed as a strong performance booster. Fuzzy logic and regression-based AI approaches were less popular than the other techniques. However, they still showed applicability for low-complexity, fast-processing use cases (e.g., TDS and EC modeling in a wetland environment).

Among the reviewed studies, GPR demonstrated particularly strong performance for WQI estimation, with prediction uncertainty below 2% in the Tongi canal study. CEEMDAN-XGB hybrids offered the best noise-filtering capability in pH and turbidity. RF-SVM hybrids performed well in BOD modeling, and EFO-optimized MLPs could offer better DO estimation with faster convergence and lower errors. Fuzzy-neural hybrids (ANFIS) are very useful in TDS and EC estimation in groundwater.

Several open research issues remain, the main one being model interpretability. As most AI models are “black boxes”, the transparency of AI systems is significantly lower, leading to lower stakeholder trust. Another data-side challenge is data sparsity in many areas, which reduces the generality and transferability of developed models to other geographical areas. As a possible step forward in improving on this, this work introduces the XAI-STD (explainable AI-spatiotemporal decomposition) framework that is based on four key components:

- Signal decomposition
- Metaheuristic-based feature selection
- Stacked probabilistic modeling
- Interpretability modules

Such a framework could be more transparent, adaptive, and robust, offering higher-quality water-quality assessments that, with the spread of IoT-based sensor networks, will be more deployable, as the models can be retrained and redeployed in real time.

Future research should move beyond conventional supervised learning toward foundation models and large language models (LLMs), which may facilitate knowledge integration, multimodal data interpretation, and decision-support applications. However, their reliability and domain-specific validation remain important challenges. Integration with Digital Twin technologies could enable continuous coupling between physical water systems, sensor observations, and dynamic AI predictions. Physics-informed AI may improve physical consistency and generalizability by incorporating hydrological and hydrochemical constraints into model development. Federated Learning may provide a pathway for collaborative model development across geographically distributed monitoring networks while reducing the need to centralize sensitive or institutionally restricted data. Together, these approaches may support more transferable, interpretable, physically consistent, and operationally adaptive water-quality prediction systems.

Statements & declarations

Author contributions

Mohammad Khoshnam: Conceptualization, Methodology, Data curation, Visualization, Writing - Original Draft, Writing - Review & Editing.

Hojat Karami: Investigation, Supervision.

Amin Shams: Project administration, Methodology, Validation, Writing - Review & Editing.

Ozgur Kisi: Writing - Review & Editing, Validation.

Funding

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

Data availability

Data sharing not applicable: No new data were created or analyzed in this study.

Declarations

The authors declare no conflict of interest.

References

- [1] Elkiran, G., Nourani, V., Abba, S. I. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *Journal of Hydrology*, 2019; 577: 123962. doi:10.1016/j.jhydrol.2019.123962.
- [2] Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A., Elshafie, A. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications*, 2017; 28: 893–905. doi:10.1007/s00521-016-2404-7.
- [3] Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., Elshafie, A. Machine learning methods for better water quality prediction. *Journal of Hydrology*, 2019; 578: 124084. doi:10.1016/j.jhydrol.2019.124084.

- [4] Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., García-Nieto, J. Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 2019; 11: 2210. doi:10.3390/w11112210.
- [5] Azad, A., Karami, H., Farzin, S., Saeedian, A., Kashi, H., Sayyahi, F. Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River). *KSCE Journal of Civil Engineering*, 2018; 22: 2206–2213. doi:10.1007/s12205-017-1703-6.
- [6] Wang, X., Zhang, F., Ding, J. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China. *Scientific Reports*, 2017; 7: 12858. doi:10.1038/s41598-017-12853-y.
- [7] Granata, F., Papirio, S., Esposito, G., Gargano, R., De Marinis, G. Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators. *Water*, 2017; 9: 105. doi:10.3390/w9020105.
- [8] Hridoy, M. A. A. M., Shawkat, A. I., Bordin, C., Acharjee, M. R., Masood, A., Baki, A. O., Al Mamun, M. A. Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management. *Science of The Total Environment*, 2025; 1008: 180965. doi:10.1016/j.scitotenv.2025.180965.
- [9] Li, Y., Jiang, S., Wu, Y., Cai, H., Jiang, C., Xiong, J., Hu, H. Auto-optimized interpretable machine learning for terminal water quality prediction of the ultrapure water production system in the semiconductor industry: demonstration in a full-scale system. *Journal of Cleaner Production*, 2026; 575: 149242. doi:10.1016/j.jclepro.2026.149242.
- [10] Mena, A., Bergillos, R. J., Paredes-Arquiola, J., Corzo, G., Solera, A., Andreu, J. A deep learning-based probabilistic framework for streamflow prediction and water quality assessment. *Journal of Hydrology*, 2026; 677: 135934. doi:10.1016/j.jhydrol.2026.135934.
- [11] Oh, H., Baek, S.-S., Hur, J. Explainable machine learning for water quality prediction across freshwater systems and water infrastructure: A comprehensive review. *Environmental Technology & Innovation*, 2026; 43: 105166. doi:10.1016/j.eti.2026.105166.
- [12] Amin, G., Pourret, O., Dupin, V., Guérin-Rechdaoui, S., Dujany, A. Comparing AI-driven approaches for predicting river water quality: a systematic review of water quality indices and remote sensing methods. *Water Research*, 2026; 301: 126047. doi:10.1016/j.watres.2026.126047.
- [13] Said, I. Machine learning for water quality management: a systematic review of prediction models, data challenges, and future directions. *Environmental Sciences Europe*, 2026; 38: 136. doi:10.1186/s12302-026-01430-8.
- [14] Pandey, K., Shukla, P., Bhakar, S., Tekchandani, H., Mittal, R. A systematic review on machine learning deep learning and IoT for water quality prediction. *Discover Sustainability*, 2026; 7: 1279. doi:10.1007/s43621-026-03756-y.
- [15] Khoshnam, M., Karami, H., Mousavi, S.-F. Chapter 23 - Applications of the grey wolf optimizer in water engineering and management. In: S. Mirjalili editors. *Grey Wolf Optimizer*. Morgan Kaufmann; 2026. p. 447–458. doi:10.1016/B978-0-443-36624-6.00033-0.
- [16] Wang, P., Yao, J., Wang, G., Hao, F., Shrestha, S., Xue, B., Xie, G., Peng, Y. Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. *Science of The Total Environment*, 2019; 693: 133440. doi:10.1016/j.scitotenv.2019.07.246.
- [17] Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., Al-Shamma'a, A. Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, 2022; 48: 102920. doi:10.1016/j.jwpe.2022.102920.
- [18] Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., Kazakis, N. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of The Total Environment*, 2020; 721: 137612. doi:10.1016/j.scitotenv.2020.137612.
- [19] Sajib, A. M., Diganta, M. T. M., Moniruzzaman, M., Rahman, A., Dabrowski, T., Uddin, M. G., Olbert, A. I. Assessing water quality of an ecologically critical urban canal incorporating machine learning approaches. *Ecological Informatics*, 2024; 80: 102514. doi:10.1016/j.ecoinf.2024.102514.
- [20] Yang, J. Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence. *Scientific Reports*, 2023; 13: 20370. doi:10.1038/s41598-023-47060-5.
- [21] Abba, S. I., Pham, Q. B., Saini, G., Linh, N. T. T., Ahmed, A. N., Mohajane, M., Khaledian, M., Abdulkadir, R. A., Bach, Q.-V. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environmental Science and Pollution Research*, 2020; 27: 41524–41539. doi:10.1007/s11356-020-09689-x.
- [22] Wu, J., Wang, Z. A Hybrid Model for Water Quality Prediction Based on an Artificial Neural Network, Wavelet Transform, and Long Short-Term Memory. *Water*, 2022; 14: 610. doi:10.3390/w14040610.
- [23] Asadollah, S. B. H. S., Sharafati, A., Motta, D., Yaseen, Z. M. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*, 2021; 9: 104599. doi:10.1016/j.jece.2020.104599.
- [24] Lu, H., Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 2020; 249: 126169. doi:10.1016/j.chemosphere.2020.126169.
- [25] Al-Mukhtar, M., Al-Yaseen, F. Modeling Water Quality Parameters Using Data-Driven Models, a Case Study Abu-Ziriq Marsh in South of Iraq. *Hydrology*, 2019; 6: 24. doi:10.3390/hydrology6010024.

- [26] Barzegar, R., Aalami, M. T., Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, 2020; 34: 415–433. doi:10.1007/s00477-020-01776-2.
- [27] Baek, S.-S., Pyo, J., Chun, J. A. Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach. *Water*, 2020; 12: 3399.
- [28] Uddin, M. G., Rahman, A., Nash, S., Diganta, M. T. M., Sajib, A. M., Moniruzzaman, M., Olbert, A. I. Marine waters assessment using improved water quality model incorporating machine learning approaches. *Journal of Environmental Management*, 2023; 344: 118368. doi:10.1016/j.jenvman.2023.118368.
- [29] Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddad, S., Kim, S., Mosavi, A., Pham, B. T. River Water Salinity Prediction Using Hybrid Machine Learning Models. *Water*, 2020; 12: 2951. doi:10.3390/w12102951.
- [30] Leggesse, E. S., Zimale, F. A., Sultan, D., Enku, T., Srinivasan, R., Tilahun, S. A. Predicting Optical Water Quality Indicators from Remote Sensing Using Machine Learning Algorithms in Tropical Highlands of Ethiopia. *Hydrology*, 2023; 10: 110. doi:10.3390/hydrology10050110.
- [31] Nafsin, N., Li, J. Prediction of 5-day biochemical oxygen demand in the Buriganga River of Bangladesh using novel hybrid machine learning algorithms. *Water Environment Research*, 2022; 94: e10718. doi:10.1002/wer.10718.
- [32] Uddin, M. G., Nash, S., Rahman, A., Olbert, A. I. Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection*, 2023; 169: 808–828. doi:10.1016/j.psep.2022.11.073.
- [33] Nanjappachetty, A., Sundar, S., Vankadari, N., Bathey Ramesh Babu, T. B., Shanmugam, P. An efficient water quality index forecasting and categorization using optimized Deep Capsule Crystal Edge Graph neural network. *Water Environment Research*, 2024; 96: e11138. doi:10.1002/wer.11138.
- [34] Jahan, K., Pradhanang, S. M. Predicting Runoff Chloride Concentrations in Suburban Watersheds Using an Artificial Neural Network (ANN). *Hydrology*, 2020; 7: 80. doi:10.3390/hydrology7040080.
- [35] Goodarzi, M. R., Niknam, A. R. R., Barzkar, A., Niazkar, M., Zare Mehrjerdi, Y., Abedi, M. J., Heydari Pour, M. Water Quality Index Estimations Using Machine Learning Algorithms: A Case Study of Yazd-Ardakan Plain, Iran. *Water*, 2023; 15: 1876. doi:10.3390/w15101876.
- [36] Kulisz, M., Kujawska, J., Przysucha, B., Cel, W. Forecasting Water Quality Index in Groundwater Using Artificial Neural Network. *Energies*, 2021; 14: 5875. doi:10.3390/en14185875.
- [37] Shiri, N., Shiri, J., Yaseen, Z. M., Kim, S., Chung, I.-M., Nourani, V., Zounemat-Kermani, M. Development of artificial intelligence models for well groundwater quality simulation: Different modeling scenarios. *PLoS ONE*, 2021; 16: e0251510. doi:10.1371/journal.pone.0251510.
- [38] Agrawal, P., Sinha, A., Kumar, S., Agarwal, A., Banerjee, A., Villuri, V. G., Annavarapu, C. S., Dwivedi, R., Dera, V. V., Sinha, J., Pasupuleti, S. Exploring Artificial Intelligence Techniques for Groundwater Quality Assessment. *Water*, 2021; 13: 1172. doi:10.3390/w13091172.
- [39] Mattas, C., Dimitraki, L., Georgiou, P., Venetsanou, P. Use of Factor Analysis (FA), Artificial Neural Networks (ANNs), and Multiple Linear Regression (MLR) for Electrical Conductivity Prediction in Aquifers in the Gallikos River Basin, Northern Greece. *Hydrology*, 2021; 8: 127. doi:10.3390/hydrology8030127.
- [40] Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., Kumar, S. Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, 2021; 276: 130265. doi:10.1016/j.chemosphere.2021.130265.
- [41] Zamili, H., Bakan, G., Zubaidi, S. L., Alawsi, M. A. Water quality index forecast using artificial neural network techniques optimized with different metaheuristic algorithms. *Modeling Earth Systems and Environment*, 2023; 9: 4323–4333. doi:10.1007/s40808-023-01750-1. doi:10.1007/s40808-023-01750-1.
- [42] Sinha, K. K., Gupta, M. K., Banerjee, M. K., Meraj, G., Singh, S. K., Kanga, S., Farooq, M., Kumar, P., Sahu, N. Neural Network-Based Modeling of Water Quality in Jodhpur, India. *Hydrology*, 2022; 9: 92. doi:10.3390/hydrology9050092.
- [43] El Bilali, A., Taleb, A., Brouziyne, Y. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, 2021; 245: 106625. doi:10.1016/j.agwat.2020.106625.
- [44] Bagheri, M., Biglarijoo, N., Shams, A., Sharifan, H., Bagheri, K., Farshforoush, N., Velazquez, A., Moaveni, M. Chapter 2 - Artificial intelligence for the water sector/industry. In: S. Jafarinejad, A. Shahirinia editors. *Artificial Intelligence for the Water-Energy-Food Nexus*. Amsterdam: Elsevier; 2026. p. 45–79. doi:10.1016/B978-0-443-34019-2.00005-5.
- [45] Biglarijoo, N., Shams, A. Multivariate modeling of the Fenton process for enhanced COD removal and low sludge generation in landfill leachate treatment. *Journal of Applied Research in Water and Wastewater*, 2024; 11: 110–115. doi:10.22126/arww.2025.11294.1346.
- [46] Biglarijoo, N., Shams, A., Bagheri, M., Gill, L., Keypour, S. From data to sustainable clean water: AI/ML contributions to the UN Sustainable Development goals. *Physics and Chemistry of the Earth, Parts A/B/C*, 2026; 144: 104427. doi:10.1016/j.pce.2026.104427.

- [47] Biglarijoo, N., Shams, A., Hadipour, V. The use of machine learning techniques in water and wastewater treatment processes: opportunities and challenges. *Numerical Methods in Civil Engineering*, 2025; 10: 86–101. doi:10.66224/NMCE.2511.1106.
- [48] Izadi, A., Yazdandoost, F. Iran's Agricultural Exports Policy from the Perspective of Virtual Water and Its Economic Value. *Civil Engineering and Applied Solutions*, 2025; 1: 53–67. doi:10.22080/ceas.2025.29230.1013.
- [49] Mohammadrezapour Tabari, M., Kabiri Samani, M. Groundwater Quality Assessment Using Entropy Weighted Osculating Value and Set Pair Analysis Methods (Case study, Sarayan plain). *Journal of Environmental Science and Technology*, 2019; 21: 99–112.
- [50] Elsayed, A., Levison, J., Binns, A., Larocque, M., Goel, P. A review of machine learning applications in the prediction of selected groundwater quality parameters: Key lessons, knowledge gaps, and future directions. *Science of The Total Environment*, 2026; 1027: 181693. doi:10.1016/j.scitotenv.2026.181693.
- [51] Ghaderi Khorasgani, M., Zarif Sanayei, H. R., Morsali, M., Toghyani, S. Multi-Parameter Assessment and Qualitative Zoning of Groundwater Resources to Identify Critical Quality Areas Using a GIS Approach. *Civil Engineering and Applied Solutions*, 2025; 1: 74–87. doi:10.22080/ceas.2025.29622.1026.