

APPLICATION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN THE PREDICTION OF WATER QUALITY PARAMETERS

Ochiagha Kate Ekwutosi^{*1}, Aghanwa Charles Ifeanyi², Obiefuna Joy Ngozika³, Onyeije Ugomma Chibuzo⁴, Oli Christain Chukwuemeka⁵, Ezeokoye Genevieve Nwanneka⁶

***corresponding Author:**

email: ke.ochiagha@unizik.edu.ng

address: Department of Pure and Industrial Chemistry, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Abstract

Reliable water-quality prediction is increasingly important because laboratory monitoring alone cannot provide the temporal coverage needed for rapid pollution detection, process control, and water-resource planning. This article critically examines how artificial intelligence (AI) and machine learning (ML) are being applied to predict physicochemical and ecological water-quality variables, with attention to dissolved oxygen (DO), pH, electrical conductivity (EC), total dissolved solids (TDS), nutrients, chlorophyll-A, and composite water quality indices (WQIs). We conducted a structured integrative evidence synthesis using 28 DOI-verified peer-reviewed studies published from 2020 to 2026; 25 were empirical modelling studies, and three were recent reviews. The evidence was coded by target variable, data setting, model family, feature strategy, validation approach, and reported performance. The study shows a clear movement from standalone artificial neural networks toward tree-based ensembles, optimized hybrids, explainable AI, and attention/Transformer architectures. Across the empirical set, tree/boosting ensembles and ANN-based models were the most frequently used families, while long short-term memory networks and Transformers were concentrated in time-series and long-horizon forecasting. Results also show that model accuracy depends less on algorithm novelty alone than on data completeness, temporal alignment, feature selection, and validation design. Random Forest and XGBoost perform strongly for WQI classification and reduced-input prediction, while optimized ANN models remain competitive for DO forecasting; self-attentive LSTM and transfer-learning Transformers are particularly useful when temporal dependence or sparse observations dominate. The article proposes an integrated prediction framework combining data-quality control, feature selection, model benchmarking, uncertainty-aware validation, and explainable outputs. The study also emphasizes that benchmark comparisons and data-leakage controls are essential before interpreting high published accuracy as evidence of deployment readiness. It concludes that AI/ML can substantially reduce monitoring cost and improve early warning, but operational deployment requires cross-site validation, transparent uncertainty reporting, and safeguards against data leakage and black-box decision-making.

Keywords: Artificial intelligence; machine learning; water quality prediction; water quality index; dissolved oxygen; XGBoost; LSTM; explainable AI

1. Introduction

Water-quality management has traditionally depended on periodic sampling followed by laboratory analysis. That approach remains indispensable for regulatory confirmation, yet it has an obvious

temporal limitation: a sample describes conditions at a particular place and time, whereas aquatic systems can change within hours because of rainfall, wastewater discharge, algal activity, temperature shifts, or operational changes in treatment plants. For river, reservoir, groundwater and drinking-water systems, the management problem is therefore no longer simply how to measure quality, but how to anticipate deterioration early enough for a useful response. Recent advances in sensor networks, hydrological databases and automated treatment systems have made data-driven forecasting increasingly practical (Jung et al., 2021; Pang et al., 2025).

Artificial intelligence (AI) and machine learning (ML) offer an alternative to purely mechanistic simulation. Rather than explicitly representing every physical, chemical, and biological process, supervised learning estimates relationships between observed predictors and a target such as dissolved oxygen (DO), total dissolved solids (TDS), electrical conductivity (EC), chlorophyll-a, or a water quality index (WQI). This flexibility is valuable when interactions are nonlinear or when governing processes are only partly observed. Early work in the present review period showed that long short-term memory (LSTM) networks could reproduce the forecast capacity of complex water-quality models and that transfer learning could improve river forecasts when inputs were missing (Liang et al., 2020; Zhou, 2020). In parallel, hybrid learning machines were used to estimate groundwater EC, salinity and TDS from long time series, with later wavelet-based comparisons extending the same idea to coupled qualitative and quantitative groundwater variables (Poursaeid et al., 2020, 2021). A subsequent survey by Poursaeid et al. (2022) documented the broader shift toward AI-assisted estimation across water-resources applications.

The field has since broadened considerably. ANN models have been coupled with weather forecasts, wavelet transforms, and metaheuristic optimizers; tree ensembles have been used for WQI regression and classification; feature selection has reduced the number of laboratory variables needed; and explainable AI (XAI) has been introduced to show which parameters are driving a prediction (Park et al., 2022; Kushwaha et al., 2023). More recent studies extend the same logic to attention-based LSTM networks, multi-stage ensembles and Transformer architectures that can capture long temporal dependencies and transfer knowledge across monitoring sites (Huang et al., 2024; Heydari et al., 2024; Sun et al., 2026).

However, high reported accuracy should not be accepted uncritically. Water-quality datasets often contain repeated seasonal patterns, strong collinearity, imbalanced quality classes, and measurements collected at unequal intervals. A random train-test split can therefore expose a model to near-duplicates of the test conditions during training. Similarly, a model may perform well at one river station but fail after transfer to a new catchment with different hydrochemistry. The practical question is not merely which algorithm produces the largest R^2 or classification accuracy, but whether the model generalizes across time, location, and changing pollution regimes. Recent comparative research reinforces this point: carefully selected training subsets can match or outperform larger redundant datasets, and feature-engineered conventional ML can equal or exceed deeper architectures in full-scale treatment forecasting (Fernández del Castillo et al., 2024; Pang et al., 2025).

This article therefore examines the application of AI and ML to water-quality parameter prediction from a methodological rather than promotional perspective. It synthesizes recent evidence on model

families, inputs, target variables, validation choices, explainability, and operational constraints.

1.1 Objectives of the study

The specific objectives are to:

- (1) clarify the main water-quality prediction concepts and model mechanisms;
- (2) compare recent empirical evidence from 2020–2026;
- (3) identify the methodological conditions under which particular models perform well;
- (4) formulate a transparent modelling framework with equations and evaluation criteria; and
- (5) identify research gaps relevant to real-world deployment.

2. Literature Review

2.1 Conceptual Review

Water quality is a multivariate state, not a single measurement. Common physicochemical indicators include temperature, pH, DO, EC, TDS, turbidity, salinity, oxidation-reduction potential, biochemical oxygen demand (BOD), chemical oxygen demand (COD), nitrogen species, and phosphorus. Biological or ecological proxies such as chlorophyll-a can add information about eutrophication. Their usefulness depends on the management objective: drinking-water safety, irrigation suitability, aquatic ecosystem health and treatment-process control do not necessarily prioritize the same variables. This is one reason WQIs are widely used: they compress several measurements into a single interpretable score or quality class. The compression is convenient, but it can also hide parameter-specific deterioration and introduces uncertainty through weighting and aggregation choices (Uddin et al., 2023).

A generic weighted arithmetic WQI can be represented as:

$$WQI = \frac{[\sum(i=1..m)w_iq_i]}{[\sum(i=1..m)w_i]} \quad (1)$$

where q_i is the quality rating for parameter i , and w_i is its assigned weight. AI/ML can be used in two distinct ways around this index. The first predicts WQI directly from a reduced set of measured parameters. The second predicts individual parameters before calculating or interpreting quality. The first approach is useful when monitoring cost is the main constraint; the second preserves more process information and is often preferable for early warning. Recent groundwater and coastal studies show that both approaches can work well, but their error structure and management implications differ (Hussein et al., 2024; Mohseni et al., 2024).

For supervised learning, the core prediction problem can be written as:

$$y = f_{\theta}(X) + \varepsilon \quad (2)$$

where X is a vector or sequence of measured predictors, f_{θ} is the learned model parameterized by θ , y is the target water-quality variable, and ε captures noise and unobserved processes. For a static WQI task, X may contain simultaneous pH, EC, TDS, hardness, and ion concentrations. For a forecasting problem, X is better treated as a temporal sequence containing lagged water-quality values, meteorology, flow and operational variables. This distinction matters because tree ensembles are strong on structured tabular data, whereas LSTM and Transformer models are designed to retain information across time.

Artificial neural networks (ANNs) remain common because they can approximate nonlinear relationships without requiring a predetermined functional form. A single neuron is expressed as:

$$h_j = \varphi(\sum_i w_{ij} x_i + b_j) \quad (3)$$

where x_i are inputs, w_{ij} are learned weights, b_j is a bias, and φ is an activation function. ANN models have successfully predicted irrigation water parameters, surface-water EC/TDS and DO, particularly when coupled with optimization, hyperparameter search, or decomposition methods (Ubah et al., 2021; Shah et al., 2021a, 2021b; Azma et al., 2023). Their limitation is that performance can be sensitive to architecture choice and training conditions, while direct interpretability is weak.

Tree-based ensembles address a different part of the problem. A Random Forest regression estimate is the average of B independently grown trees:

$$\hat{y}_{RF}(x) = (1/B) \sum_{b=1..B} T_b(x) \quad (4)$$

while boosting builds trees sequentially to reduce residual error. Random Forest, gradient boosting, and XGBoost have been particularly effective for WQI prediction because water-quality tables often contain nonlinear thresholds, interactions, and mixed-scale variables. In recent studies, XGBoost produced near-perfect WQI classification under specific datasets, while RF has provided strong groundwater and potability predictions (Uddin et al., 2023; Apogba et al., 2024; Nallakuruppan et al., 2024). These results are useful but should be interpreted relative to class balance, spatial independence, and the way WQI was generated. Support vector regression/classification uses a kernel mapping to find a function with controlled complexity. In its regression form:

$$f(x) = w^T \varphi(x) + b \quad (5)$$

with optimization balancing model flatness against errors outside an ε -insensitive margin. SVM models remain competitive when datasets are moderate in size and well scaled. They are frequently included as benchmarks in groundwater and river-quality studies, although ensembles or ANN variants often win the final comparison (Khan & Ayaz, 2024; Fernández del Castillo et al., 2024). Time-series forecasting introduces memory. The LSTM architecture uses gated state updates, commonly summarized as:

$$i_t = \sigma(W_i[x_t, h(t-1)] + b_i), f_t = \sigma(W_f[x_t, h_{t-1}] + b_f) \quad (6)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c[x_t, h_{t-1}] + b_c) \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

This mechanism helps a model learn delayed responses, seasonality, and persistence. LSTM models have been used for river forecasting under missing data, sparse pollution-load observations, and mechanism-guided simulation (Zhou, 2020; Liang et al., 2020; Huang et al., 2024). Transformers extend the idea through attention, assigning context-dependent weights to different positions in a sequence. Recent water-quality applications show growing interest in this architecture for long-horizon forecasting and small-data transfer, although it requires greater computational resources and validation (Huang et al., 2025; Sun et al., 2026).

A final concept is explainability. Prediction alone is insufficient in high-consequence water management if users cannot see why an alert has been generated. SHAP-based explanations and feature-importance analysis can identify which variables contribute most strongly to a prediction and whether the direction of influence is scientifically plausible. Park et al. (2022) showed that SHAP-informed input selection improved the stability and interpretation of XGBoost chlorophyll-a models.

Nallakaruppan et al. (2024) similarly used XAI to explain potability classification. Explainability does not prove causal correctness, but it can expose implausible dependencies and support more defensible monitoring decisions.

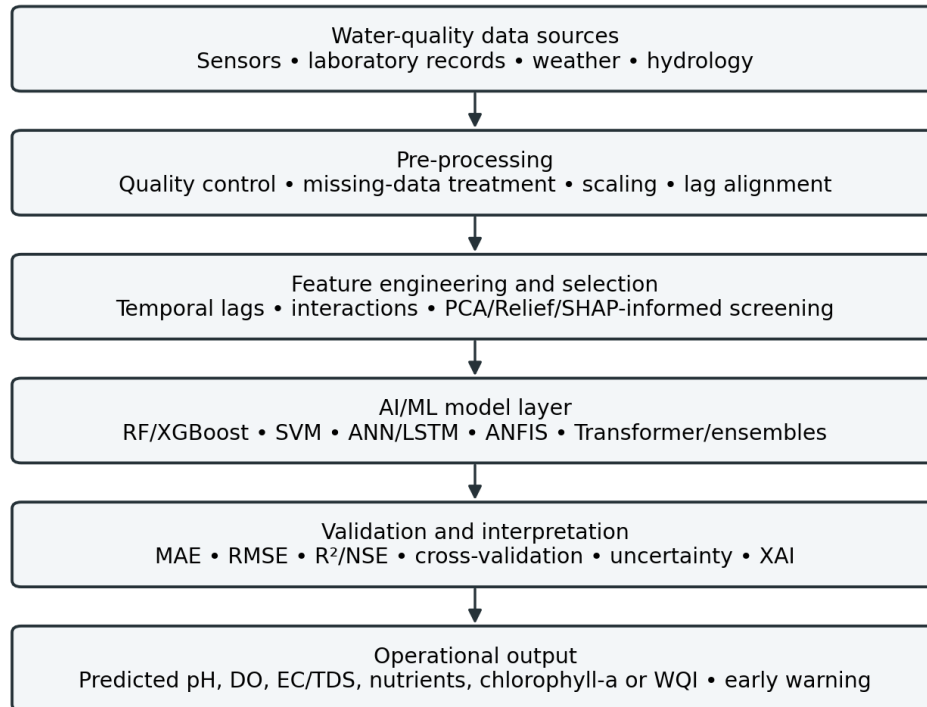


Figure 1. Integrated AI/ML pipeline for water-quality parameter prediction.

2.2 Theoretical Framework

The theoretical framework adopted here combines statistical learning theory, temporal-dependence learning, and explainable prediction. Statistical learning theory treats model development as a generalization problem: the objective is not to minimize training error alone, but to learn a function that performs well on unseen observations. The practical implication is the bias-variance trade-off. Simple linear models may underfit nonlinear hydrochemical relationships; highly flexible models may fit noise or site-specific patterns. Cross-validation, temporal holdout, and external-site testing are therefore part of the model's theory, not merely optional reporting steps.

Temporal-dependence learning adds the assumption that the present state of a water body contains information from previous states and external drivers. This is why lagged variables, moving statistics, LSTM memory cells, and attention mechanisms are relevant. The importance of temporal structure is visible in studies where transfer learning helped recover forecasts under missing inputs and where a 12-hour lag captured delayed treatment-chemical effects (Zhou, 2020; Pang et al., 2025). The framework also recognizes data drift: a function estimated under one season, pollution regime, or treatment policy may lose accuracy when that regime changes.

The third element is interpretability. A useful operational model should return a prediction, an estimate of predictive error or uncertainty, and an explanation of influential features. SHAP can be represented in additive form as:

$$g(z') = \varphi_0 + \sum_{i=1..M} \varphi_i z'_i \quad (9)$$

where φ_0 is the baseline output, and φ_i is the contribution assigned to feature i . Within water-quality prediction, the theoretical role of XAI is diagnostic rather than decorative: it allows analysts to test whether model reliance corresponds with known water chemistry and process behavior. Figure 2 summarizes the framework as an iterative cycle linking observed state, representation, learning, generalization testing, and interpretable decision support.

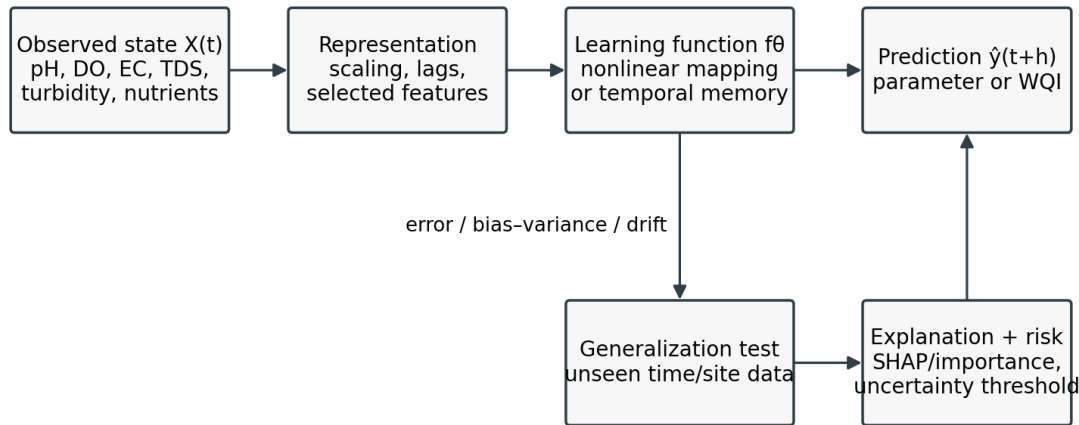


Figure 2. Theoretical framework linking function learning, generalization, temporal dependence and explainability.

2.3 Empirical Review

Recent empirical work falls into four broad streams: predicting individual physicochemical variables; WQI regression or quality-class prediction; time-series forecasting with recurrent or attention architectures; and model interpretation/feature reduction. Table 2.1 presents representative studies. The table is deliberately comparative: it highlights the target, model strategy, principal result, and methodological implication rather than reproducing every metric reported in each paper.

Table 2.1. Comparative empirical review of selected AI/ML water-quality prediction studies

Study	Target/context	Model(s)	Reported result	Relevance to present synthesis
Liang et al. (2020)	Mechanism-guided water-quality forecasting	LSTM; RF benchmark	LSTM reproduced the forecasting behavior of a complex EFDC model.	Shows the value of temporal memory and hybrid physics-data learning.

Jung et al. (2021)	Surface-water quality with probabilistic weather inputs	ANN	Weather-informed ANN supported prospective surface-water prediction.	Demonstrates coupling of hydrometeorology and water-quality learning.
Ubah et al. (2021)	Irrigation water-quality forecasting, Nigeria	ANN	ANN was used to forecast multiple water-quality parameters and WQI.	Provides African evidence for low-cost predictive monitoring.
Park et al. (2022)	Chlorophyll-a prediction and interpretation	XGBoost + SHAP	SHAP-based variable prioritization produced stable and interpretable models.	Links XAI with monitoring-variable selection.
Uddin et al. (2023)	Coastal WQI classification	SVM, NB, RF, KNN, XGBoost	XGBoost achieved accuracy 1.00 with precision and F1 around 0.99; KNN also classified strongly.	Shows high performance but raises need for independent validation.
Azma et al. (2023)	Daily dissolved oxygen	BBO-ANN; ASO-ANN	BBO-ANN testing MAPE 2.3848%; correlation 0.99186.	Supports optimized ANN for nonlinear DO forecasting.
Kushwaha et al. (2023)	Groundwater WQI and irrigation indices	M5P, ANN, RF, Relief feature selection	M5P WQI NSE 0.988 and R 0.999 with reduced inputs.	Feature selection can reduce monitoring burden.
Huang et al. (2024)	Sparse CODMn and NH ₃ -N forecasting	SA-LSTM + LOADEST	RMSE reduced by 24.6% and 21.3% versus standalone SA-LSTM.	Addresses data scarcity and irregular monitoring.
Apogba et al. (2024)	Groundwater WQI mapping, Ghana	Random Forest	RF achieved R ² 0.82 for spatial groundwater-quality prediction.	Demonstrates spatial ML under heterogeneous basin conditions.
Fernández del Castillo et al. (2024)	Highly polluted river monitoring network	ANFIS, ANN, SVM + clustering/time-series analysis	Reduced representative training subset matched or	Evidence against “more data is always better”.

			improved full-data performance.	
Heydari et al. (2024)	Six-hour DO and chlorophyll-a forecasts	XGB, GBR, LightGBM, MLP + GWO/NSGA-II	Optimization improved KGE; ensembles improved further over single models.	Supports tuned ensembles for short-term operational forecasts.
Pang et al. (2025)	Full-scale drinking-water treatment	XGBoost and DL time-series models + SHAP	XGBoost with 12-h lag outperformed naive baseline by about 3–4% MAPE.	Feature engineering can rival deeper architectures.
Sun et al. (2026)	Long-sequence smart-water forecasting	Transformer + phase-space reconstruction + transfer learning	Best in 41/45 scenarios; error reduction about 10.08–39.52%.	Shows emerging potential for long-horizon multi-site forecasting.

2.4 Gap in Literature

The recent literature demonstrates strong predictive capability, but several connected gaps remain. First, many high-accuracy WQI studies validate models within the same dataset from which the index is calculated; this can overstate generalization when training and test observations share locations, seasons, or repeated quality patterns (Uddin et al., 2023; Shams et al., 2024). Second, studies that reduce input parameters often emphasize predictive accuracy without fully quantifying whether the reduced set remains stable across catchments or changing pollution regimes (Kushwaha et al., 2023; Mohseni et al., 2024). Third, deep-learning and Transformer studies address temporal complexity, yet external-site validation and uncertainty calibration remain less common than point-forecast comparisons (Huang et al., 2024; Sun et al., 2026). Finally, explainability is growing but is still frequently treated as post-hoc visualization rather than part of model acceptance. This review responds by integrating model selection, feature sufficiency, temporal validation, uncertainty, and explainability in one operational framework.

3. Materials and Methods

This study used a structured integrative evidence-synthesis design. Candidate studies were identified through targeted searches of recent peer-reviewed literature on artificial intelligence, machine learning, water quality prediction, water quality index, dissolved oxygen, groundwater quality, LSTM, XGBoost, and explainable AI. Publisher pages and bibliographic records were used to verify authorship, publication year, journal details, and DOI. Eligibility was restricted to publications from 2020 through August 2026 that directly addressed prediction, forecasting, classification, or AI-assisted estimation of water-quality variables or indices. We retained review papers only to support conceptual synthesis; quantitative counts of model adoption were based on empirical studies.

The final evidence set comprised 28 verified publications: 25 empirical studies and three recent reviews. Each empirical study was coded for year, water setting, target variable, model family, use

of feature selection or optimization, interpretability method, and reported validation approach. Because performance metrics differ by target scale and study design, we did not calculate a pooled meta-analytic effect size. Instead, the synthesis reports within-study best-performing models and uses frequency counts to identify methodological trends. This avoids the invalid practice of directly averaging an R^2 from WQI regression with an F1-score from quality classification or an NSE from DO forecasting.

The model-comparison framework used in this article recommends a common data split that respects temporal or spatial dependence. For a chronological forecasting problem, observations should be ordered and divided into training, validation, and test periods without random shuffling. For multi-site datasets, include at least one spatial holdout. Tune hyperparameters only on training/validation data, then leave the test set untouched until final evaluation. When classes are imbalanced, include stratified metrics and confusion matrices alongside headline accuracy.

Four core regression metrics are recommended:

$$MAE = \left(\frac{1}{n}\right) \sum |y_i - \hat{y}_i| \quad (10)$$

$$RMSE = \text{sqrt} \left[\left(\frac{1}{n}\right) \sum (y_i - \hat{y}_i)^2 \right] \quad (11)$$

$$R^2 = 1 - \left[\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \right] \quad (12)$$

$$NSE = 1 - \left[\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_{obs})^2} \right] \quad (13)$$

MAE is easy to interpret in the original units; RMSE penalizes larger errors; R^2 measures explained variance; and NSE is widely used in hydrological forecasting. Classification tasks should report precision, recall, F1-score, and balanced accuracy rather than accuracy alone. The evidence reviewed here shows why several metrics are necessary: a model can maintain high correlation while producing operationally important bias, and a high overall accuracy can conceal poor performance for rare degraded-water classes.

Figure 1 shows the proposed implementation pathway. Raw sensor, laboratory, meteorological, and hydrological inputs are first subjected to quality checks, missing-value treatment, and temporal alignment. Feature engineering then creates lags, moving summaries and relevant interactions. Candidate models should include at least one interpretable baseline, one tree ensemble and, for genuine time-series problems, one recurrent or attention-based model. Feature selection can use embedded tree importance, Relief-type methods or SHAP-informed screening, but selection must occur inside the training process to avoid leakage. The final model is chosen for generalization performance, stability, and interpretability rather than a single best metric.

For operational acceptance, an uncertainty or alert rule should accompany the point estimate. One simple threshold formulation is:

$$\text{Alert}_t = 1, \text{ if } P(y(t+h) > L | X_t) \geq \tau; \text{ otherwise } 0 \quad (14)$$

where L is a regulatory or management limit, and τ is the minimum probability required for an alert. Even when a model is deterministic, uncertainty can be approximated using ensembles, bootstrap resampling or quantile models. The purpose is to move water-quality AI from retrospective curve fitting toward risk-aware decision support.

4. Results

The evidence coding reveals a rapidly expanding and diversifying field. Figure 3 shows that only three of the coded empirical studies were published in 2020, whereas nine appeared in 2024. This increase reflects more than a rise in paper volume. The methodological emphasis also changed: earlier studies concentrated on ANN, LSTM, and hybrid optimization, while recent work increasingly combines ensembles, feature selection, XAI, and attention mechanisms.

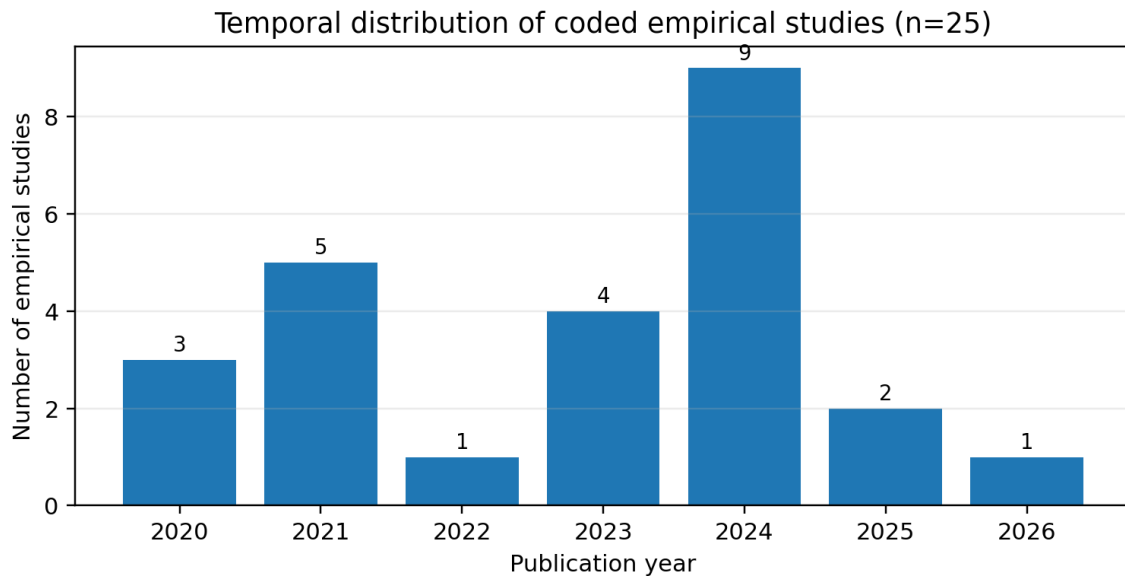


Figure 3. Publication-year distribution of the 25 empirical studies coded in the study.

ANN/MLP models were the most frequently represented family in the coded empirical literature, followed by tree and boosting ensembles (Figure 4). SVM remained a common benchmark, particularly for groundwater and WQI tasks, whereas LSTM and Transformer methods were concentrated in sequence forecasting. The pattern suggests a practical division of labor. Structured water-quality tables favor RF, XGBoost, and related ensembles because these algorithms handle nonlinear interactions with limited preprocessing. Long temporal sequences favor recurrent or attention architectures because persistence and delayed response carry predictive information. Hybrid/metaheuristic methods occur in both settings, mainly to tune weights or hyperparameters.

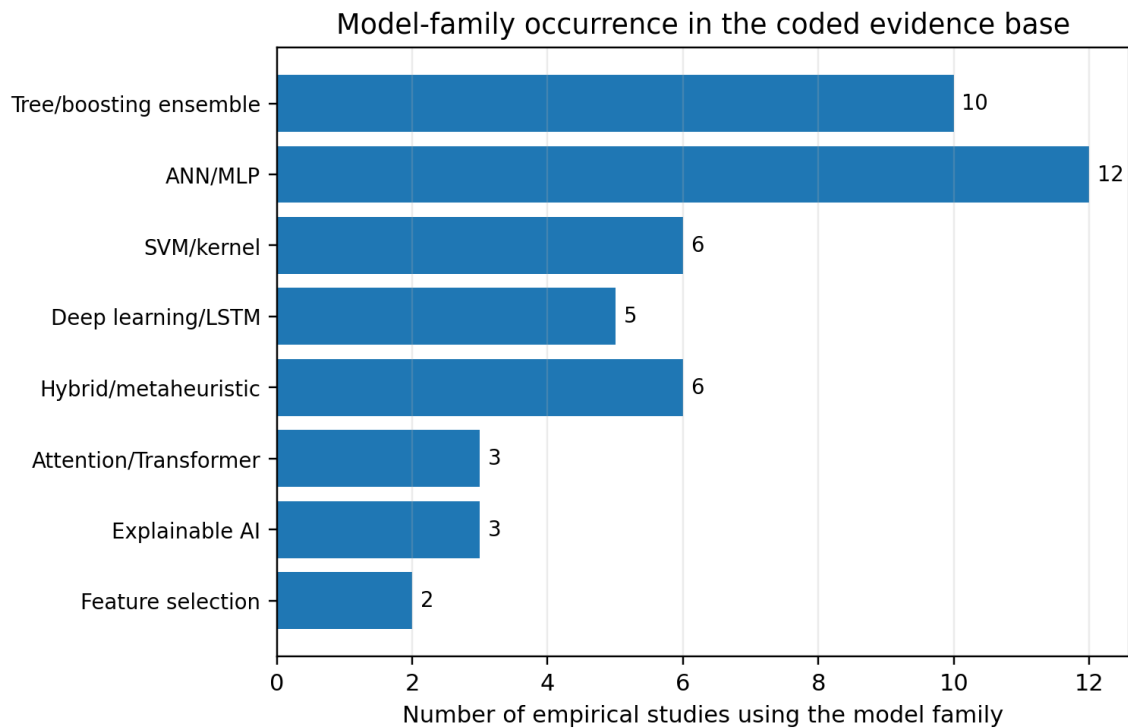


Figure 4. Frequency of major model families in the empirical evidence set; studies may contribute to more than one family.

The target distribution in Figure 5 shows that WQI or water-quality classification is the single largest application group. DO-centered forecasting is also prominent because DO responds dynamically to temperature, flow, biological activity, and treatment conditions and is difficult to measure continuously in some settings. Multi-parameter forecasting has become more visible with treatment-plant and Transformer studies. This shift matters because operational systems rarely need only one variable; they need coordinated forecasts that can be converted into alarms, treatment decisions, or monitoring priorities.

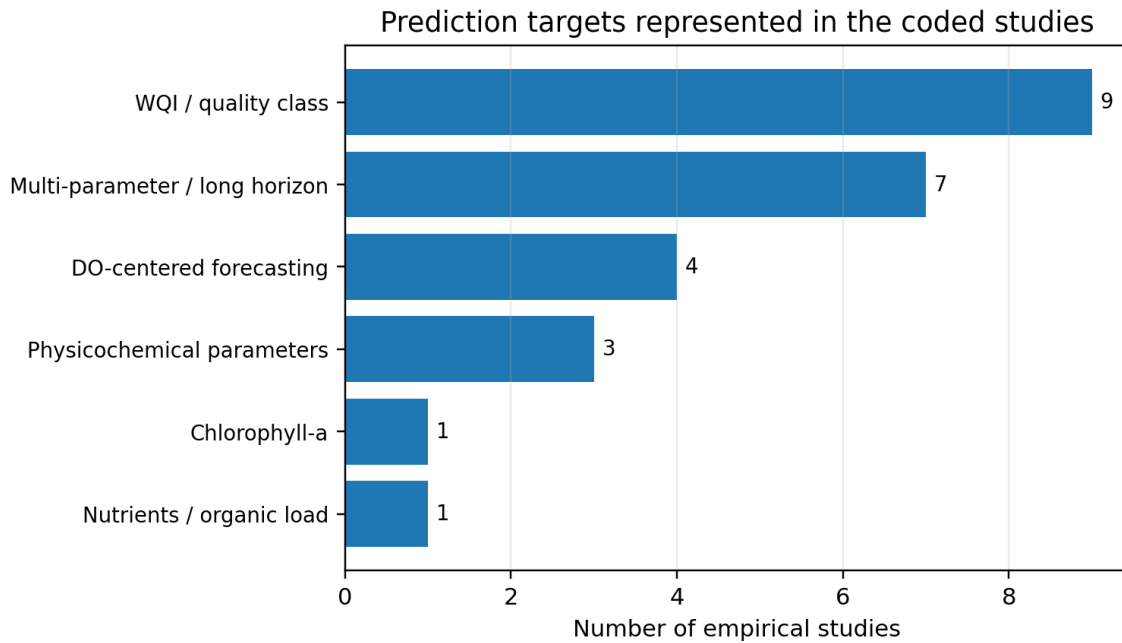


Figure 5. Broad prediction targets represented in the empirical evidence set.

Table 4.1 summarizes selected reported outcomes without treating unlike metrics as directly comparable. Several patterns stand out. Uddin et al. (2023) reported very high classification performance for XGBoost and KNN across WQI formulations, demonstrating the capacity of ensemble and instance-based methods when the class boundaries are learnable. Kushwaha et al. (2023) found M5P particularly effective for WQI, with an NSE of 0.988 and a correlation of 0.999 on their groundwater dataset. Azma et al. (2023) achieved a testing MAPE of 2.3848% and a correlation of 0.99186 with a BBO-ANN model for DO. Huang et al. (2024) showed a different type of gain: the SA-LSTM-LOADEST framework reduced RMSE by 24.6% for CODMn and 21.3% for NH₃-N relative to standalone SA-LSTM while operating under sparse monitoring.

Spatial groundwater prediction also produced useful but less extreme performance. Apogba et al. (2024) reported $R^2 = 0.82$ for Random Forest prediction of groundwater WQI in the Nabogo Basin, a value that is more plausible for a spatially heterogeneous problem than the near-perfect scores sometimes observed in internally derived WQI datasets. In full-scale drinking-water treatment, Pang et al. (2025) found that an XGBoost model with time-feature engineering outperformed a naive baseline by about 3–4% in MAPE at a 12-hour lag and could be interpreted through SHAP. The most recent long-horizon study in the evidence set, Sun et al. (2026), reported that Aquaformer achieved the best result in 41 of 45 scenarios and reduced prediction error by approximately 10.08–39.52% relative to baseline models.

These results indicate that the strongest development is not a universal winning algorithm. Performance improves when the model architecture matches the data-generating structure: ensembles for tabular nonlinear relations, recurrent/attention models for temporal memory, and hybrid or transfer methods when the dataset is sparse or shifts across sites. Feature selection and XAI also appear increasingly important because operational water monitoring costs money. A model that retains accuracy with fewer measurements can be more valuable than a marginally more accurate model requiring an expensive full laboratory panel.

Table 4.1. Selected published performance results used for comparative interpretation

Study	Target	Best / highlighted model	Reported performance	Study context
Uddin et al. (2023)	WQI class	XGBoost	Accuracy 1.00; precision 0.99; F1 0.99	Coastal dataset / multiple WQI formulations
Kushwaha et al. (2023)	WQI	M5P	NSE 0.988; R 0.999; RMSE 10.769	Groundwater; Relief-selected predictors
Yang (2023)	DO	EFO/TLBO-MLPNN variants	Best testing RMSE $\approx$ 1.29; Pearson R $\approx$ 0.88	Klamath River; daily data
Azma et al. (2023)	DO	BBO-ANN	MAPE 2.3848%; PCC 0.99186	Rock Creek; five-year daily series
Huang et al. (2024)	CODMn / NH3-N	SA-LSTM-LOADEST	NSE 0.71 / 0.57; RMSE reduction 24.6% / 21.3%	Sparse river-lake monitoring
Apogba et al. (2024)	Groundwater WQI	Random Forest	R ² 0.82; RMSE 23.03	Nabogo Basin spatial prediction
Nallakaruppan et al. (2024)	Potability class	Random Forest	Accuracy/F1 0.9999; precision 0.9997; recall 0.998	Public potability dataset; SHAP interpretation
Pang et al. (2025)	Four treatment outputs	XGBoost	3–4% lower MAPE than naive baseline at 12-h lag	Full-scale treatment plant
Sun et al. (2026)	Long-horizon water quality	Aquaformer	Best in 41/45 scenarios; 10.08–39.52% error reduction	Five monitoring datasets; 144-step horizon

5. Discussion

The study supports three main findings. First, AI/ML is now mature enough to predict several classes of water-quality outcomes with useful accuracy, but reported performance is conditional on data design. The strong results of Random Forest, XGBoost, and M5P for WQI agree with the broader review by Yan et al. (2024), which documented the rapid expansion of ML for both indicator and index prediction. At the same time, the near-perfect classification reported in some studies should be interpreted as evidence of learnable dataset structure, not proof that the same model will classify every new river or aquifer with comparable accuracy. WQI is calculated from the same parameters used as predictors in many studies, so the learning task may partly reproduce a deterministic aggregation rule.

Second, the evidence favors parsimonious, well-engineered inputs over indiscriminate data accumulation. Kushwaha et al. (2023) used Relief-based feature selection to reduce the predictor set while retaining high performance. Fernández del Castillo et al. (2024) reached a similar conclusion from another direction: a best training subset representing the most variable and representative river conditions performed as well as, or better than, the full dataset. This challenges a common assumption that more rows or more parameters automatically improve prediction. Redundant records can increase computational cost, reinforce dominant regimes, and contribute to overtraining without adding new information.

The same principle appears in groundwater and urban WQI studies. Mohseni et al. (2024) identified EC, alkalinity, hardness, and TDS as important drivers of weighted arithmetic WQI, while Khan and Ayaz (2024) found EC highly influential in their ANN-based integrated groundwater quality model. These results help design lower-cost monitoring programs, but parameter importance should not be treated as universal. Hydrogeology, pollution source, and the chosen WQI formula alter the relationship. A feature set optimized in one basin therefore requires revalidation before transfer.

Third, temporal models are valuable when water quality dynamics are delayed or persistent. Zhou (2020) demonstrated that transfer learning can mitigate forecast degradation when inputs are missing. Huang et al. (2024) extended this idea to sparse datasets through self-attention, LSTM, and LOADEST, while Pang et al. (2025) showed that explicit lag engineering can allow conventional XGBoost to compete with deeper models. The lesson is important: choose deep learning because the prediction problem requires sequence representation, not because it is assumed to be intrinsically superior. Helaly et al. (2025) reached a similar conclusion in their practical review and comparative implementation, where model performance depended on dataset characteristics and preparation.

Recent Transformer work pushes the horizon further. Huang et al. (2025) combined Transformer components with bidirectional sequence processing and KAN/Mamba elements for nonlinear water-quality series. Sun et al. (2026) used phase-space reconstruction and multi-source transfer learning to address long-sequence forecasting and data scarcity. These architectures are promising for smart-water networks with high-frequency sensor streams, yet their operational credibility depends on external validation. A complex model can appear excellent when the test period resembles the training regime but deteriorate during extreme events, sensor replacement, or abrupt changes in pollutant loading.

Explainability offers a partial safeguard. Park et al. (2022) showed that SHAP can do more than produce a post-hoc plot; it can guide variable selection and improve model stability. Nallakuruppan et al. (2024) used SHAP outputs to interpret potability classification, and Pang et al. (2025) used it to verify that treatment predictions followed plausible process relationships. Explainability nevertheless remains associative. If a model assigns high importance to a feature, it means the feature contributes to the model's predictions within the observed data; it does not establish that manipulating that feature will causally change water quality.

For practical use, model validation should therefore include temporal holdout, spatial transfer where feasible, sensitivity to missing values, and calibration of uncertainty. Optimized ANN models for DO illustrate why this matters. Yang (2023) and Azma et al. (2023) achieved strong results using metaheuristic optimization, but they developed their models at specific USGS stations. They

demonstrate technical feasibility rather than universal transferability. Likewise, ANN-based surface-water models using probabilistic weather inputs show that hydrometeorological covariates can improve forecasting, but the model still inherits the local relationships embedded in its training data (Jung et al., 2021).

The overall evidence therefore supports a layered view of water-quality AI: prediction accuracy is necessary, data representativeness determines generalization, feature strategy determines monitoring cost, and explainability/uncertainty determine whether the output is defensible in practice. Research that reports only a best accuracy score addresses only the first layer.

6. Practical Implications and Recommendations

Water agencies and treatment operators can use the evidence in this review to design AI systems incrementally. A reasonable starting point is a robust tabular baseline such as Random Forest or XGBoost using routinely measured variables. If the prediction horizon is short and the dataset is modest, this may outperform a deep network while being easier to maintain. Recurrent or Transformer models become more appropriate when high-frequency histories, long temporal dependence, or multi-step forecasts are central to the decision problem.

Monitoring design should link to feature importance but not be dictated by a single fitted model. Variables repeatedly shown to be useful, such as DO, pH, EC, TDS, turbidity, temperature, and selected nutrient indicators, can form a core sensor set, with periodic laboratory measurements used for recalibration. Feature reduction should be re-tested after seasonal change, sensor replacement, or transfer to a new catchment.

Four safeguards are recommended for deployment:

- (1) preserve a truly unseen temporal or spatial test set;
- (2) report MAE/RMSE or class-sensitive metrics alongside headline accuracy;
- (3) provide feature explanations and an uncertainty or confidence indicator for each alert; and
- (4) establish a retraining trigger based on data drift or sustained error.

These safeguards convert AI from a static research model into a monitored decision-support component.

Future studies should prioritize multi-site datasets, uncertainty calibration, missing-data stress tests, and prospective evaluation during unusual events. Hybrid physics-ML approaches are also promising because they can constrain purely statistical predictions with known mass-balance or process behavior. For low-resource settings, sparse-data methods and transfer learning deserve particular attention because they can extend prediction to locations where continuous laboratory monitoring is unrealistic.

7. Conclusion

Artificial intelligence and machine learning have moved water-quality prediction from a niche modelling exercise toward a practical component of smart monitoring. Evidence published since 2020 shows successful applications to DO, EC, TDS, nutrients, chlorophyll-a, groundwater quality, and composite WQIs. Tree ensembles are particularly effective for structured tabular data and quality classification, while ANN variants remain strong for nonlinear parameter estimation. LSTM and Transformer architectures add value when temporal memory, sparse observations, or long forecasting horizons are central to the task.

The most important conclusion, however, is methodological rather than algorithmic. No model family is consistently superior across all water systems. Strong performance emerges when the architecture matches the data, the input set is informative rather than merely large, and evaluation protects against temporal or spatial leakage. Feature selection can lower monitoring cost, and XAI can make predictions more transparent, but both require revalidation when the environmental context changes.

A defensible operational framework therefore combines data-quality control, carefully separated training/validation/test periods, benchmark comparison, uncertainty reporting, and explainable outputs. Used in this way, AI/ML can complement, not replace, laboratory analysis and professional water-quality judgment. Its greatest value lies in extending the reach of monitoring: filling temporal gaps, anticipating deterioration, prioritizing sampling and supporting faster responses. The next stage of research should focus less on isolated record-breaking accuracy and more on transferable, uncertainty-aware models that continue to perform when deployed outside the dataset on which they were originally trained.

References

- Apogba, J. N., Anornu, G. K., Koon, A. B., Dekongmen, B. W., Sunkari, E. D., Fynn, O. F., & Kpiebaya, P. (2024). Application of machine learning techniques to predict groundwater quality in the Nabogo Basin, Northern Ghana. *Heliyon*, 10(7), e28527. <https://doi.org/10.1016/j.heliyon.2024.e28527>
- Azma, A., Liu, Y., Azma, M., Saadat, M., Zhang, D., Cho, J., & Rezaia, S. (2023). Hybrid machine learning models for prediction of daily dissolved oxygen. *Journal of Water Process Engineering*, 54, 103957. <https://doi.org/10.1016/j.jwpe.2023.103957>
- Fernández del Castillo, A., Verduzco Garibay, M., Díaz-Vázquez, D., Yebra-Montes, C., Brown, L. E., Johnson, A., Garcia-Gonzalez, A., & Gradilla-Hernández, M. S. (2024). Improving river water quality prediction with hybrid machine learning and temporal analysis. *Ecological Informatics*, 82, 102655. <https://doi.org/10.1016/j.ecoinf.2024.102655>
- Helaly, M. A., Rady, S., Mabrouk, M., Aref, M. M., Villarroja, S., Cotos, J. M., & Mera, D. (2025). Advancements in water quality prediction: A practical review of machine learning and deep learning approaches. *Cluster Computing*, 28, 598. <https://doi.org/10.1007/s10586-025-05221-3>
- Heydari, S., Nikoo, M. R., Mohammadi, A., & Barzegar, R. (2024). Two-stage meta-ensembling machine learning model for enhanced water quality forecasting. *Journal of Hydrology*, 641, 131767. <https://doi.org/10.1016/j.jhydrol.2024.131767>
- Huang, S., Xia, J., Wang, Y., Lei, J., & Wang, G. (2024). Water quality prediction based on sparse dataset using enhanced machine learning. *Environmental Science and Ecotechnology*, 20, 100402. <https://doi.org/10.1016/j.es.2024.100402>
- Huang, T., Jiang, Y., Gan, R., & Wang, F. (2025). A novel water quality prediction model based on BiMKANsDformer. *Environmental Science: Water Research & Technology*, 11, 590–603. <https://doi.org/10.1039/D4EW00883A>
- Hussein, E. E., Derdour, A., Zerouali, B., Almaliki, A., Wong, Y. J., Ballesta-de los Santos, M., Ngoc, P. M., Hashim, M. A., & Elbeltagi, A. (2024). Groundwater quality assessment and

- irrigation water quality index prediction using machine learning algorithms. *Water*, 16(2), 264. <https://doi.org/10.3390/w16020264>
- Jung, W. S., Kim, S. E., & Kim, Y. D. (2021). Prediction of surface water quality by artificial neural network model using probabilistic weather forecasting. *Water*, 13(17), 2392. <https://doi.org/10.3390/w13172392>
- Khan, I., & Ayaz, M. (2024). Sensitivity analysis-driven machine learning approach for groundwater quality prediction: Insights from integrating ENTROPY and CRITIC methods. *Groundwater for Sustainable Development*, 26, 101309. <https://doi.org/10.1016/j.gsd.2024.101309>
- Kushwaha, N. L., Rajput, J., Suna, T., Sena, D. R., Singh, D. K., Mishra, A. K., Sharma, P. K., & Mani, I. (2023). Metaheuristic approaches for prediction of water quality indices with relief algorithm-based feature selection. *Ecological Informatics*, 75, 102122. <https://doi.org/10.1016/j.ecoinf.2023.102122>
- Liang, Z., Zou, R., Chen, X., Ren, T., Su, H., & Liu, Y. (2020). Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *Journal of Hydrology*, 581, 124432. <https://doi.org/10.1016/j.jhydrol.2019.124432>
- Mohseni, U., Pande, C. B., Pal, S. C., & Alshehri, F. (2024). Prediction of weighted arithmetic water quality index for urban water quality using ensemble machine learning model. *Chemosphere*, 352, 141393. <https://doi.org/10.1016/j.chemosphere.2024.141393>
- Nallakuruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports*, 14, 7520. <https://doi.org/10.1038/s41598-024-56775-y>
- Pang, H., Ben, Y., Cao, Y., Qu, S., & Hu, C. (2025). Time series-based machine learning for forecasting multivariate water quality in full-scale drinking water treatment with various reagent dosages. *Water Research*, 268, 122777. <https://doi.org/10.1016/j.watres.2024.122777>
- Park, J., Lee, W. H., Kim, K. T., Park, C. Y., Lee, S., & Heo, T.-Y. (2022). Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Science of the Total Environment*, 832, 155070. <https://doi.org/10.1016/j.scitotenv.2022.155070>
- Poursaeid, M., Mastouri, R., Shabanlou, S., & Najarchi, M. (2020). Estimation of total dissolved solids, electrical conductivity, salinity and groundwater levels using novel learning machines. *Environmental Earth Sciences*, 79, 453. <https://doi.org/10.1007/s12665-020-09190-1>
- Poursaeid, M., Mastouri, R., Shabanlou, S., & Najarchi, M. (2021). Modelling qualitative and quantitative parameters of groundwater using a new wavelet conjunction heuristic method: Wavelet extreme learning machine versus wavelet neural networks. *Water and Environment Journal*, 35, 67–83. <https://doi.org/10.1111/wej.12595>
- Poursaeid, M., Poursaeed, A. H., & Shabanlou, S. (2022). Study of water resources parameters using artificial intelligence techniques and learning algorithms: A survey. *Applied Water Science*, 12, 156. <https://doi.org/10.1007/s13201-022-01675-7>
- Shah, M. I., Alaloul, W. S., Alqahtani, A., & Aldrees, A. (2021a). Predictive modeling approach for

- surface water quality: Development and comparison of machine learning models. *Sustainability*, 13(14), 7515. <https://doi.org/10.3390/su13147515>
- Shah, M. I., Javed, M. F., Alqahtani, A., & Aldrees, A. (2021b). Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. *Process Safety and Environmental Protection*, 151, 324–340. <https://doi.org/10.1016/j.psep.2021.05.026>
- Shams, M. Y., Elshewey, A. M., El-Kenawy, E.-S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83, 35307–35334. <https://doi.org/10.1007/s11042-023-16737-4>
- Sun, M., Xu, C., Jia, Q., & Jia, H. (2026). Aquaformer: Multi-source transfer learning model based on Transformer and phase space reconstruction for long sequence water quality forecasting. *Journal of Hydrology*, 664, 134372. <https://doi.org/10.1016/j.jhydrol.2025.134372>
- Ubah, J. I., Orakwe, L. C., Ogbu, K. N., Awu, J. I., Ahaneku, I. E., & Chukwuma, E. C. (2021). Forecasting water quality parameters using artificial neural network for irrigation purposes. *Scientific Reports*, 11, 24438. <https://doi.org/10.1038/s41598-021-04062-5>
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection*, 169, 808–828. <https://doi.org/10.1016/j.psep.2022.11.073>
- Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X., & Zhao, X. (2024). A comprehensive review of machine learning for water quality prediction over the past five years. *Journal of Marine Science and Engineering*, 12(1), 159. <https://doi.org/10.3390/jmse12010159>
- Yang, J. (2023). Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence. *Scientific Reports*, 13, 20370. <https://doi.org/10.1038/s41598-023-47060-5>
- Zhou, Y. (2020). Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. *Journal of Hydrology*, 589, 125164. <https://doi.org/10.1016/j.jhydrol.2020.125164>