Transpublika International Research in Exact Sciences
https://ojs.transpublika.com/index.php/TIRES
Online ISSN 2828-5921 | Vol. 5 No. 3, 111–126

https://doi.org/10.55047/tires.v5i3.2476

Literature Review

A Literature Review of Machine Learning Algorithm Performance in Water Quality Parameter Prediction

Asep Saefurohman*

UIN Sultan Maulana Hasanuddin Banten, Indonesia
Email: asepsaefurohman@uinbanten.ac.id

Received : 16 October - 2026
Revised : 13 November - 2026
Accepted : 18 November - 2026
Published online : 19 November - 2026

Abstract

Water quality is essential for environmental sustainability and public health, yet conventional laboratory-based monitoring is costly and slow. Machine learning (ML) has been proposed as an alternative, but algorithm performance is inconsistent across studies and seldom compared on identical target parameters. This study synthesizes ML algorithm use and performance in predicting water quality parameters. A structured narrative literature review searched Scopus for English-language publications from 2015 to 2026 using keywords combining machine learning, deep learning, or artificial intelligence with water quality and prediction, forecasting, or classification. Titles, abstracts, and full texts were screened against predefined inclusion and exclusion criteria; 37 publications were analyzed. Algorithms, target parameters, validation strategies, and metrics (R^2, RMSE, MAE) were extracted and synthesized narratively, with performance grouped by target parameter, not compared across targets. Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbors, neural networks, and boosting algorithms (XGBoost, AdaBoost, CatBoost, LightGBM) were most frequently used, while the Water Quality Index (WQI) was the most frequently predicted target, followed by dissolved oxygen (DO), pH, and biochemical oxygen demand (BOD). Ensemble and boosting models generally performed strongly; for example, XGBoost reached R^2 = 0.93 and RMSE = 0.047 for the Irrigation Water Quality Index, and an attention-based LSTM reached R^2 = 0.985 for daily DO. However, no algorithm was universally superior, and external validation remained rare. Algorithm selection should consider data characteristics, target parameter, evaluation metrics, and interpretability.

Keywords: Algorithm Performance, Literature Review, Machine Learning, Water Quality Index, Water Quality.

1. Introduction

Water is an essential resource for human life, the sustainability of aquatic ecosystems, and socioeconomic development. However, access to clean and safe water remains a global challenge. The WHO/UNICEF Joint Monitoring Programme (JMP) report recorded that in 2022, 2.2 billion people still lacked access to safely managed drinking water, including 115 million people who still relied on untreated surface water as their main source of drinking water (WHO/UNICEF, 2023). WHO (2023) also reported that at least 1.7 billion people use drinking water sources contaminated with feces. The degradation of water quality in both surface water bodies and groundwater has a direct impact on public health, the sustainability of aquatic ecosystems, and the effectiveness of water resource management.

Conventional water quality monitoring is generally carried out by collecting field samples that are analyzed in laboratories. This approach is accurate but has practical limitations. The analysis of parameters such as biochemical oxygen demand (BOD), chemical

Copyright: © 2026 by the authors.
This is an open access article distributed under the terms and conditions of the CC BY 4.0.

oxygen demand (COD), or heavy metals requires expensive reagents and equipment; analysis results can take several days, making it difficult to detect sudden pollution early; limited measurement frequency cannot capture rapid dynamics such as daily fluctuations in dissolved oxygen (DO); and the limited coverage of monitoring points makes it less representative of conditions at a broad spatial scale. These limitations drive the need for more efficient and sustainable monitoring and prediction systems. This need is reflected in studies from varied settings, such as dissolved oxygen prediction in surface water discharged into the Mar Menor coastal lagoon, Spain (García del Toro et al., 2022), classification of the Water Quality Index in Laguna de Bay (Catajan Jr et al., 2023), monitoring of an industrial waste-polluted stream (Ejaz et al., 2024), and modeling of BOD and COD in aquaculture sites of the southeastern coastal region of Bangladesh (Chakraborty et al., 2026).

Advances in computing and the availability of data have encouraged the widespread application of ML in environmental modeling, including water quality prediction in general (Aldhyani et al., 2020), in rivers and surface water (Alqahtani et al., 2022; Shah et al., 2021), groundwater (Singha et al., 2021), and coastal waters (García del Toro et al., 2022). Unlike linear regression, which assumes linear relationships between variables, algorithms such as Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) are able to capture nonlinear relationships and complex interactions among parameters without strict assumptions about data distribution.

The target parameters for prediction are highly diverse, including pH, DO, BOD, COD, turbidity, total dissolved solids (TDS), electrical conductivity (EC), nutrients, and composite indices such as the Water Quality Index (WQI). The frequency of use of each parameter is uneven across studies; the actual distribution in the reviewed studies is reported in the Results section.

In terms of algorithms, approaches have moved from classical models such as linear regression and Decision Tree, to ensemble models such as RF, Gradient Boosting, and XGBoost, to deep learning such as ANN, LSTM, Convolutional Neural Network (CNN), and hybrid models. Jena et al. (2022) reported that RF reached R^2 = 90.8% for TDS and 86.08% for pH on river water data. Sakaa et al. (2022) applied RF and a hybrid sequential minimal optimization-SVM (SMO-SVM) model for WQI in Wadi Saf-Saf, Algeria, with RF reaching a testing R^2 of 0.82 using 13 input parameters. In deep learning, He et al. (2025) reported that an attention-based bidirectional LSTM (AT-BiLSTM) reached R^2 = 0.985 and RMSE = 0.054 mg/L for daily DO in the Neches River, United States, surpassing MLPNN, which reached only R^2 = 0.771.

A number of reviews have summarized the application of ML in water quality prediction. Yan et al. (2024) reviewed more than 170 studies from the last five years and grouped them into single or multiple indicator prediction and water quality index prediction. Lokman et al. (2025) reviewed ML models and statistical methods for surface water quality forecasting and classification, while Ab Karim et al. (2026) focused on WQI formulation, modeling approaches, and explainable artificial intelligence (XAI). These reviews emphasize methodological and technical trends. Based on the author's search, there is no review that quantitatively compares reported performance metrics per target parameter with a transparently reported literature selection procedure.

Several research gaps can be identified. First, the performance of algorithms is inconsistent across studies because of differences in datasets, target parameters, locations, and sample sizes. Second, evaluation metrics are not uniform: some studies report only R^2, while others include RMSE, MAE, or MAPE, making direct comparison difficult. Third,

validation strategies vary, from simple hold-out to k-fold cross-validation, which affects the level of generalization. Fourth, most studies perform only internal validation at a single location without external validation at different locations or periods. Fifth, ensemble and deep learning models are often treated as black boxes, although some studies have begun to use XAI such as SHapley Additive exPlanations (SHAP) (Hridoy et al., 2025). These gaps matter in practice because ML applications have been reported for different water bodies, such as rivers (He et al., 2025; Najafabadi et al., 2025), irrigation water in the Bouhamdane Basin, Algeria (Mohamed et al., 2025), and lakes and reservoirs (Prasad et al., 2021; Walczak & Walczak, 2025), each with different datasets, target parameters, and validation schemes. As a result, water quality managers lack a consolidated, parameter-specific basis for choosing an algorithm for their own monitoring context, which underlies the urgency of this review.

This study makes three contributions. First, it presents a literature synthesis with a transparently reported selection procedure. Second, it groups performance comparisons by the same target parameter, such as WQI with WQI and DO with DO, rather than comparing metric values across different targets. Third, it qualitatively discusses the validation practices, the completeness of metric reporting, and the use of explainable artificial intelligence (XAI) in the reviewed studies.

The study aims to answer four research questions. The first asks which machine learning algorithms are most widely used in water quality prediction. The second asks which water quality parameters are most frequently predicted using machine learning. The third asks what performance, based on R^2 and RMSE, is reported for the same target parameters, as a descriptive summary of the reported values without ranking the algorithms. The fourth asks how performance compares across algorithms, in terms of relative patterns within the same studies, including RF, SVM, ANN, XGBoost, Gradient Boosting, LSTM, and others, and what validation and interpretability practices these algorithms follow.

2. Literature Review

2.1. Machine learning in water quality prediction

Commonly used ML algorithms can be grouped into several families. Decision Tree builds hierarchical decision rules that are easy to interpret. Random Forest combines many trees trained on random samples (bagging), making it more stable than a single tree. Boosting algorithms, such as Gradient Boosting, XGBoost, AdaBoost, CatBoost, and LightGBM, build models sequentially by correcting the errors of the previous model. SVM separates or regresses data through kernel functions; Sakaa et al. (2022), for example, used the hybrid SMO-SVM variant for WQI. Artificial neural networks model nonlinear relationships through layers of neurons, and LSTM adds memory cells and gates so that it can retain long-term information in time series data, as in daily DO prediction (He et al., 2025).

Comparative studies are widely conducted because no algorithm has been proven superior in all conditions. Kularbphettong et al. (2025), for example, compared SVM, RF, Decision Tree, XGBoost, and AdaBoost for predicting WQI of rivers and canals in Thailand. Variation in datasets and target parameters such as this is the reason a systematic synthesis is needed.

2.2. Water quality parameters and indices

The physical, chemical, and biological parameters that are predicted include pH, DO, BOD, COD, turbidity, temperature, EC, nitrate, iron, and total suspended solids (TSS). pH indicates the level of acidity or alkalinity of water, DO is related to the survival of aquatic

organisms, BOD reflects organic pollution, and COD describes the content of pollutants that require oxygen for chemical oxidation. WQI integrates several parameters into a single value. Ab Karim et al. (2026) noted that the deterministic structure of WQI can produce high predictive performance because the model reflects the predetermined index formulation rather than independently learned patterns. This note is important in interpreting high R^2 values in WQI prediction.

2.3. Evaluation metrics and validation strategies

The performance of regression models is generally reported with R^2, RMSE, and MAE, while classification models are reported with accuracy and related metrics (Table 1). The choice of metrics affects the interpretation of results, so reporting more than one metric is recommended (Khoshvaght et al., 2025). In terms of validation, hold-out splits the data once so the results may depend on the particular split, k-fold cross-validation reduces this dependence, and external validation tests generalization at locations or periods that were not used during training.

Table 1. Main evaluation metrics

Metric	Description	Interpretation
R^2	Proportion of variance in actual values explained by the model	Closer to 1 indicates a better fit
RMSE	Root mean square error; same unit as the target variable; sensitive to large errors	Lower is better
MAE	Mean absolute error; more robust to outliers	Lower is better
MAPE	Mean absolute percentage error	Lower is better; problematic when actual values are close to zero
NSE	Nash–Sutcliffe Efficiency; compares model error with the mean of observations	Closer to 1 indicates better performance
Accuracy	Proportion of correct class predictions (classification)	Higher is better; needs to be supplemented with other metrics when classes are imbalanced

Source: Author's compilation based on standard metric definitions

2.4. Model interpretability

Ensemble and deep learning models can be accurate but difficult to explain. XAI approaches such as SHAP help explain the contribution of each variable to the prediction (Cabri & Rahimi, 2025; Hridoy et al., 2025). In the reviewed studies, SHAP-based explainability was applied by Cabri and Rahimi (2025) and Hridoy et al. (2025), but it has not become a common practice in the reviewed literature. By showing which variables drive a prediction, XAI makes model outputs more transparent and traceable, so that managers can judge whether a prediction rests on plausible water quality drivers before acting on it. This is especially relevant for index-based targets: because WQI is calculated with a deterministic formula, a high R^2 may reflect the model's ability to imitate that formula (Ab Karim et al., 2026), and XAI can help show which input variables the model relies on. In water quality management, knowing the determinants of a prediction can be as important as accuracy, because such explanations support decisions such as which parameters to prioritize for monitoring and how to justify management actions based on model outputs.

Ensemble and deep learning models can be accurate but difficult to explain. XAI approaches such as SHAP help explain the contribution of each variable to the prediction

(Cabri & Rahimi, 2025; Hridoy et al., 2025). In water quality management, knowing the determinants of a prediction can be as important as accuracy.

3. Methods

3.1. Research design

This study is a structured narrative literature review. Literature selection was carried out with explicitly stated criteria so that the process is transparent and traceable, but it did not follow a formal systematic review protocol: the protocol was not registered in a registry and a formal risk of bias assessment was not conducted.

3.2. Data sources and literature search

Literature was searched in the Scopus database for English-language publications from 2015 to 2026. The keywords combined the terms machine learning, deep learning, or artificial intelligence with water quality and prediction, forecasting, or classification, in titles, abstracts, and keywords. This search is structured but is not intended to cover all of the literature (see Limitations).

3.3. Literature selection criteria

Table 2. Inclusion and Exclusion Criteria

Code	Criterion
IC1	Published in 2015–2026 and written in English.
IC2	Indexed in Scopus; a peer-reviewed journal article, conference paper, or book chapter.
IC3	Empirical study applying ML to predict or classify water quality parameters or indices (surface water, groundwater, or treated water).
IC4	Reports the algorithms used and at least one quantitative performance metric.
EC1	Review articles or secondary research (used only as supporting literature, not as synthesis data).
EC2	Prediction target is not a water quality parameter (for example, water level, river discharge, or process performance).
EC3	Does not use ML, or is a duplicate.
EC4	Full text is not accessible or the information is insufficient for data extraction.

Source: Inclusion and exclusion criteria developed by the author for this review

3.4. Study selection

Titles and abstracts were reviewed against the criteria in Table 2, then the full texts of candidates were assessed for eligibility. The selection stages are summarized in Table 3.

3.5. Data extraction

Data were extracted using a structured form containing study identity, publication type, type of water body, target parameter, algorithms used, validation strategy, and the metrics R^2, RMSE, MAE, and other metrics where reported. A formal risk of bias assessment was not conducted.

3.6. Data synthesis

Synthesis was carried out in a structured narrative manner because the heterogeneity of datasets, target parameters, and validation methods did not allow a meta-analysis. First, the frequencies of algorithms and parameters were calculated by counting one study once per category (RQ1 and RQ2). Second, metric values were tabulated and grouped per target parameter; values across targets were not combined (RQ3). Third, comparisons between

algorithms were read primarily within the same study and summarized as qualitative patterns, because within-study comparisons use the same data and validation scheme (RQ4). Validation practices and the use of XAI were recorded qualitatively and their frequencies were not counted. The synthesis results are reported without declaring any single algorithm as universally the best.

4. Research Results

4.1. Research Results

4.1.1. Selection results and study characteristics

The literature selection stages are summarized in Table 3 and resulted in 37 publications being analyzed. The characteristics of the publications are summarized in Table 4. Distribution by publication year: 0 (0.0%) publications in 2015–2019, 10 (27.0%) in 2020–2022, and 27 (73.0%) in 2023–2026, so the reviewed literature is dominated by recent publications. The publications consisted of 23 (62.2%) journal articles, 13 (35.1%) conference papers, and 1 (2.7%) book chapter. Data extraction per publication recorded algorithms for 24 and target parameters for 23 of the 37 publications, while water body type, region, and performance metric values were recorded for only four publications (Table 7). Validation strategies and the use of XAI have not been systematically extracted, so they are discussed qualitatively.

Table 3. Literature selection stages

Stage	Procedure	Output
Search	Scopus search: publications from 2015–2026, in English, with the keywords machine learning/deep learning/artificial intelligence, water quality, and prediction/forecasting/classification.	Set of candidate publications
Screening	Titles and abstracts reviewed against the inclusion and exclusion criteria (IC1–IC4, EC1–EC4; Table 2).	Candidates for full-text assessment
Eligibility	Full texts assessed; publications that did not meet the criteria, were duplicates, or were inaccessible were excluded (EC1–EC4).	Publications that met the criteria
Final corpus	Publications analyzed and synthesized.	37 publications (23 journal articles, 13 conference papers, 1 book chapter)

Source: Author's selection process applying the criteria in Table 2

Table 4. Characteristics of the analyzed publications

Category	Subcategory	n (%)
Publication year	2015–2019	0 (0.0%)
Publication year	2020–2022	10 (27.0%)
Publication year	2023–2026	27 (73.0%)
Publication type	Journal article	23 (62.2%)
Publication type	Conference paper	13 (35.1%)
Publication type	Book chapter	1 (2.7%)

Source: Author's analysis of the 37 included publications

4.1.2. Most widely used algorithms (RQ1)

Table 5 summarizes the algorithms found in 24 of the 37 reviewed publications (those whose algorithms have been extracted). The ensemble and boosting families appeared most often: RF appeared in 18 studies and boosting algorithms in 14 studies, followed by SVM (12),

Decision Tree (8), ANN/DNN (8), and KNN (8). LSTM, Logistic Regression, Naive Bayes, and ELM appeared in a much smaller number of studies. Classical algorithms such as Decision Tree, SVM, KNN, and neural networks remain widely used, both as single models and as comparators or components of stacking and optimization approaches.

Table 5. Algorithms used in water quality prediction

Algorithm	n studies*	Studies
Random Forest (RF)	18	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Catajan Jr et al., 2023; Costa et al., 2024; Ehmimed et al., 2025; Elmotawakkil et al., 2025; Hridoy et al., 2025; Husain et al., 2025; Jena et al., 2022; Krishna & Sudha, 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Najafabadi et al., 2025; Prasad et al., 2021; Rawat et al., 2023; Sakaa et al., 2022; Sharma et al., 2024; Walczak & Walczak, 2025)
Boosting (Gradient Boosting, XGBoost, LightGBM, AdaBoost, CatBoost)	14	(Al-Sulttani et al., 2021; Catajan Jr et al., 2023; Costa et al., 2024; Ehmimed et al., 2025; Elmotawakkil et al., 2025; Hridoy et al., 2025; Husain et al., 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Mohamed et al., 2025; Najafabadi et al., 2025; Pisal et al., 2025; Rawat et al., 2023; Sharma et al., 2024)
Support Vector Machine (SVM)	12	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Ehmimed et al., 2025; Hridoy et al., 2025; Husain et al., 2025; Krishna & Sudha, 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Najafabadi et al., 2025; Prasad et al., 2021)
Decision Tree (DT)	8	(Catajan Jr et al., 2023; Costa et al., 2024; Daniels & Koutsougeras, 2021; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Prasad et al., 2021; Rawat et al., 2023; Wang, 2025)
Artificial/Deep Neural Network (ANN/DNN)	8	(Cabri & Rahimi, 2025; Daniels & Koutsougeras, 2021; Ehmimed et al., 2025; Elmotawakkil et al., 2025; García del Toro et al., 2022; He et al., 2025; Krushna & Sasikala, 2024; Walczak & Walczak, 2025)
K-Nearest Neighbors (KNN)	8	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Daniels & Koutsougeras, 2021; Ehmimed et al., 2025; Elmotawakkil et al., 2025; Husain et al., 2025; Krushna & Sasikala, 2024; Rawat et al., 2023)
LSTM and recurrent variants	3	(Cabri & Rahimi, 2025; Ehmimed et al., 2025; He et al., 2025)
Logistic Regression	2	(Prasad et al., 2021; Rawat et al., 2023)
Naive Bayes	2	(Prasad et al., 2021; Rawat et al., 2023)
Extreme Learning Machine (ELM)	1	(Wang, 2025)

*Number of the 24 publications (out of the 37 reviewed) whose algorithms have been extracted; one study may use more than one algorithm. Algorithms are recorded as explicitly mentioned in each study.

Source: Compiled by the author from the 24 publications whose algorithms were identified

4.1.3. Most frequently predicted parameters (RQ2)

WQI (including IWQI) was the most frequently predicted target, in 14 studies, followed by DO (5 studies), pH (4 studies), and BOD (3 studies) (Table 6). Other parameters such as COD, temperature, nitrate, turbidity, EC, TSS, and TDS appeared in one to two studies. This tendency shows that ML research on water quality is concentrated on composite indices and

on indicators of dissolved oxygen and organic pollution, while other parameters are relatively underexplored. These counts are based on the 23 publications whose target parameters have been extracted.

Table 6. Predicted water quality parameters

Parameter	Role	n*	Studies
Water Quality Index (WQI/IWQI)	Composite index	14	(Al-Mukhtar et al., 2024; Catajan Jr et al., 2023; de Lima et al., 2025; Ejaz et al., 2024; Elmotawakkil et al., 2025; Hridoy et al., 2025; Kularbphettong et al., 2025; Lokman, Wan Ismail, et al., 2025; Mohamed et al., 2025; Prasad et al., 2021; Sakaa et al., 2022; Sharma et al., 2024; Walczak & Walczak, 2025; Yan et al., 2024)
Dissolved oxygen (DO)	Essential for aquatic organisms	5	(Al-Mukhtar et al., 2024; Chakraborty et al., 2026; García del Toro et al., 2022; He et al., 2025; Voza et al., 2025)
pH	Acidity/alkalinity	4	(Daniels & Koutsougeras, 2021; Jena et al., 2022; Vaishali et al., 2025; Yaseen et al., 2024)
Biochemical oxygen demand (BOD)	Organic pollution	3	(Al-Mukhtar et al., 2024; Chakraborty et al., 2026; Walczak & Walczak, 2025)
Chemical oxygen demand (COD)	Pollutants that are chemically oxidized	2	(Chakraborty et al., 2026; Walczak & Walczak, 2025)
Temperature	Affects chemical/biological processes	2	(Chakraborty et al., 2026; Vaishali et al., 2025)
Nitrate, iron, and other chemicals	Nutrients and contaminants	2	(Bazai et al., 2024; Walczak & Walczak, 2025)
Drinkability (potability)	Suitability for drinking water	2	(Rawat et al., 2023; Vaishali et al., 2025)
Total dissolved solids (TDS)	Dissolved solids	1	(Jena et al., 2022)
Turbidity	Water clarity	1	(Vaishali et al., 2025)
Electrical conductivity (EC)	Ion concentration	1	(Chakraborty et al., 2026)
Total suspended solids (TSS)	Particulates	1	(Walczak & Walczak, 2025)

*Number of the 23 publications (out of the 37 reviewed) whose target parameters have been extracted; one study may predict more than one parameter.

Source: Compiled by the author from the 23 publications whose target parameters were identified

4.1.4. Performance based on R^2 and RMSE (RQ3)

Table 7 groups the reported metric values by target parameter. For IWQI in the Bouhamdane Basin, Algeria, XGBoost reached R^2 0.93 and RMSE 0.047, while AdaBoost reached R^2 0.79 and RMSE 0.085, and CatBoost was reported to have performance close to AdaBoost (Mohamed et al., 2025). For WQI in Wadi Saf-Saf, RF reached a testing R^2 of 0.82 (Sakaa et al., 2022). For daily DO in the Neches River, AT-BiLSTM reached R^2 0.985 and

RMSE 0.054 mg/L, compared with MLPNN with R^2 0.771 (He et al., 2025). RF was also reported to reach R^2 0.908 for TDS and 0.8608 for pH (Jena et al., 2023).

Table 7. Examples of reported performance, grouped by target parameter

Target	Algorithm	R^2	RMSE	Study and context
IWQI	XGBoost	0.93	0.047	Mohamed et al. (2025); Bouhamdane Basin, Algeria
IWQI	AdaBoost	0.79	0.085	Mohamed et al. (2025)
WQI	RF	0.82 (test)	–	Sakaa et al. (2022); Wadi Saf-Saf; 13 input parameters
Daily DO	AT-BiLSTM	0.985	0.054 mg/L	He et al. (2025); Neches River, US
Daily DO	MLPNN	0.771	–	He et al. (2025)
TDS	RF	0.908	–	Jena et al. (2023); river water
pH	RF	0.8608	–	Jena et al. (2023); river water

Source: Compiled by the author from the cited studies, with values as reported in each source

4.1.5. Performance comparison across algorithms (RQ4)

Table 8 summarizes the relative performance patterns across algorithms in the reviewed publications. RF and XGBoost were consistently reported to perform strongly in both prediction and classification tasks; Gradient Boosting was also strong although slightly below the two in some studies. LSTM showed superiority on time series data, while SVM and ANN/DNN were competitive, especially with feature selection. KNN and Decision Tree generally had lower accuracy than ensemble models but were easy to interpret. These patterns come from comparisons within each study and do not mean that any one algorithm is the best in all conditions.

Table 8. Relative performance patterns across algorithms

Algorithm	n*	Reported performance pattern	Interpretation notes	Studies
Random Forest	11	Strong performance in regression and classification; relatively more resistant to overfitting than a single tree; often among the best models.	Feature importance provides some interpretation, but RF is still an ensemble model so XAI is needed.	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Catajan Jr et al., 2023; Costa et al., 2024; Ehmimed et al., 2025; Husain et al., 2025; Krishna & Sudha, 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Najafabadi et al., 2025; Rawat et al., 2023)
XGBoost	10	High and efficient accuracy; often equal to or slightly better than RF. Example: R^2 0.93 on IWQI.	Superiority over RF is not consistent across studies.	(Al-Sulttani et al., 2021; Catajan Jr et al., 2023; Costa et al., 2024; Ehmimed et al., 2025; Husain et al., 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Mohamed et al., 2025; Najafabadi et

Algorithm	n*	Reported performance pattern	Interpretation notes	Studies
				al., 2025; Rawat et al., 2023)
Gradient Boosting	5	Strong predictive power, slightly below XGBoost or RF in some studies.	–	(Al-Sulttani et al., 2021; Costa et al., 2024; Kularbphettong et al., 2025; Najafabadi et al., 2025; Sharma et al., 2024)
AdaBoost	3	Good, but often outperformed by RF, XGBoost, or GB (e.g., R^2 0.79 vs 0.93 on IWQI).	Small number of supporting studies.	(Husain et al., 2025; Kularbphettong et al., 2025; Mohamed et al., 2025)
SVM	9	Competitive, especially with feature selection; below RF/XGBoost in some studies.	Sensitive to kernel and hyperparameter selection.	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Ehmimed et al., 2025; Husain et al., 2025; Krishna & Sudha, 2025; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Najafabadi et al., 2025; Rawat et al., 2023)
ANN/DNN	3	Effective for complex nonlinear relationships; comparable to or slightly better than tree-based models on time series tasks.	Small number of supporting studies; requires more data.	(Cabri & Rahimi, 2025; Ehmimed et al., 2025; He et al., 2025)
LSTM	3	Superior on time series data. Example: AT-BiLSTM R^2 0.985 vs MLPNN 0.771 on daily DO.	The advantage is related to the temporal structure of the data, not a universal advantage.	(Cabri & Rahimi, 2025; Ehmimed et al., 2025; He et al., 2025)
KNN	6	Simple and easy to explain, but accuracy is generally below ensemble and deep learning.	Useful as a baseline comparator.	(Al-Sulttani et al., 2021; Cabri & Rahimi, 2025; Ehmimed et al., 2025; Husain et al., 2025; Krushna & Sasikala, 2024; Rawat et al., 2023)
Decision Tree	5	Easy to interpret, but accuracy is generally lower and prone to overfitting.	Useful when transparency is prioritized over accuracy.	(Catajan Jr et al., 2023; Costa et al., 2024; Krushna & Sasikala, 2024; Kularbphettong et al., 2025; Rawat et al., 2023)

*Number of publications supporting the pattern in the "Reported performance pattern" column among the 37 reviewed publications; the figure may be smaller than in Table 5 because Table 5 counts all uses of an algorithm, while Table 8 only counts publications that compare it with other algorithms.

Source: Compiled by the author from within-study comparisons reported in the cited publications

4.2. Discussion

The dominance of RF and boosting algorithms in water quality prediction is in line with the nature of the data: tabular, medium-sized, and involving nonlinear relationships and interactions among physical, chemical, and environmental parameters. Tree ensembles handle these characteristics without strict distributional assumptions. Their strong performance can also be explained by how they work: RF combines many trees trained on random samples, which makes it more stable than a single tree, while boosting builds models sequentially by correcting the errors of the previous model, so both can fit nonlinear interactions flexibly without the strict assumptions of linear models. However, frequency of use does not automatically reflect superiority: popular algorithms are tried more often, and studies tend to report models that succeed, so there is potential for publication bias.

Quantitative examples show why cross-study comparisons must be made carefully. XGBoost outperformed AdaBoost on IWQI (R^2 0.93 vs 0.79; Mohamed et al., 2025), but this difference applies to a single dataset, location, and validation scheme. Therefore, comparisons in this study are grouped per target parameter and read as patterns, not universal rankings. Consistent with this, no algorithm is superior across all parameters and datasets. This is consistent with the view that performance depends on how well a model's structure fits the data, so algorithm rankings should be treated as conditional on the dataset and target. Whereas earlier reviews mainly mapped methodological and technical trends (Yan et al., 2024; Lokman et al., 2025; Ab Karim et al., 2026), the present synthesis adds a parameter-specific reading of reported performance, which helps prevent high values in one setting from being mistaken for general superiority.

WQI is the most frequently predicted target because it summarizes many parameters into a single value that is easy to use for assessment and management. However, Ab Karim et al. (2026) showed that because WQI is calculated with a deterministic formula, a very high R^2 value can reflect the model's ability to imitate the formula, especially when the parameters that make up the index are used as inputs. Readers need to interpret WQI prediction performance by considering the composition of input variables, and future studies should report scenarios without the parameters that make up the index. This implies that the practical value of ML for WQI lies less in reproducing the index and more in estimating water quality from fewer or cheaper inputs, a direction examined for reservoir water by Walczak and Walczak (2025).

The concentration of studies on DO, pH, BOD, and COD shows that ML is widely applied to indicators directly related to physical-chemical conditions and organic pollution. A plausible reason is that DO can be measured continuously with sensors, which produces the time series that data-driven models need (He et al., 2025), while BOD and COD require expensive reagents and equipment and take days to analyze, so predictive models offer the greatest savings there, as shown for aquaculture sites by Chakraborty et al. (2026). Parameters such as nutrients, heavy metals, and biological indicators are relatively rare, even though their laboratory measurement is expensive; this is a research opportunity. LSTM is relevant for sensor and time series data, for example daily DO (He et al., 2025), while for tabular data without a strong temporal structure, tree-based models are generally sufficient. The advantage of LSTM on time series data is related to its memory cells and gates, which retain information over long periods and suit the temporal dependence of daily measurements.

The heterogeneity of metrics and validation is the main obstacle to synthesis. Some studies report only R^2, so the magnitude of error in original units cannot be assessed, and external validation at different locations or periods remains rare. As a result, the generalization ability of models under new conditions has not been proven. Because most

models are evaluated on data from the same site and period used for training, high reported accuracy mainly shows that a model fits a known system, not that it can be transferred to a new water body, which is the property that matters for real monitoring applications. These results support the recommendation to report multiple metrics (Khoshvaght et al., 2025) and to use validation that tests generalization.

In terms of interpretability, XAI such as SHAP has begun to be used (Cabri & Rahimi, 2025; Hridoy et al., 2025), but it has not become a common practice. For water quality management, understanding the variables that drive predictions is as important as accuracy, so model selection needs to weigh accuracy, interpretability, and data requirements.

For water quality managers, these findings suggest: (1) for medium-sized tabular data, start with RF or boosting with a simple baseline; (2) for sequential sensor data, consider LSTM or its variants; (3) when decisions must be explained, supplement the model with SHAP; and (4) avoid choosing a model based only on R^2 on internal test data.

The contribution of this review to the development of water quality prediction can be summarized in three points. First, it shows which algorithms and target parameters dominate the literature, and why. Second, it reads performance per target parameter, so that high values are not misinterpreted as general superiority of an algorithm. Third, it identifies reporting and validation practices, namely uniform metrics, external validation, WQI prediction without its constituent parameters, and routine use of XAI, that would make future studies comparable and more useful for water quality management.

This study has several limitations. First, this study is a structured narrative literature review, not a formal systematic review: the search used only Scopus and English-language publications and the protocol was not registered, so studies in other databases or languages are not covered and the coverage of all literature cannot be guaranteed. Second, conference papers and book chapters were included, so the quality of reporting varies across studies, and a formal risk of bias assessment was not conducted. Third, the extraction of algorithms and parameters follows what is explicitly mentioned in each study and has only been recorded for 24 (algorithms) and 23 (target parameters) of the 37 publications; water body type, region, and metric values (R^2, RMSE, and MAE) have only been extracted for four publications (Table 7), so the frequencies and the synthesis of quantitative performance do not yet cover all studies. Fourth, no meta-analysis was conducted because of the heterogeneity of datasets and validation methods. Fifth, there are no publications from the 2015–2019 period in the set of reviewed publications, so the findings reflect more recent literature even though the search range covers 2015–2026. Sixth, validation practices and the use of XAI were recorded qualitatively and their frequencies were not counted. Finally, selective reporting by primary studies may affect the observed patterns.

Future research needs to use external validation at different locations and periods, report uniform metrics (R^2, RMSE, MAE) along with hyperparameters, test WQI prediction without the parameters that make up the index, expand target parameters to nutrients and priority contaminants, and integrate XAI routinely.

5. Conclusion

This literature review of 37 publications answers four research questions. First (RQ1), RF and boosting algorithms are the most widely used algorithms, followed by SVM, Decision Tree, neural networks, and KNN. Second (RQ2), WQI is the most frequently predicted target, followed by DO, pH, and BOD. Third (RQ3), high performance is reported across various targets, for example XGBoost with R^2 0.93 and RMSE 0.047 on IWQI and AT-BiLSTM with

R^2 0.985 on daily DO, but values can only be compared within the same target. Fourth (RQ4), RF and XGBoost consistently perform strongly, LSTM excels on time series data, and no algorithm is universally superior. External validation and uniform metric reporting remain limited. The main implication is that ML, particularly tree-based ensembles and, for sensor time series, LSTM, is a promising alternative to costly and slow laboratory-based monitoring, but high reported accuracy, especially for WQI, should not be taken as proof of general superiority or of transferability to new water bodies. The main contribution of this review is therefore a parameter-specific reading of reported performance that clarifies what the current evidence supports and where it remains limited.

Based on these findings, water quality managers are advised to start with RF or boosting for medium-sized tabular data, consider LSTM for sequential sensor data, and supplement models with SHAP when decisions must be explained. Algorithm selection should be based on data characteristics, target parameter, evaluation metrics, and interpretability requirements. Future research is advised to use external validation at different locations and periods, report uniform metrics (R^2, RMSE, MAE) along with hyperparameters, test WQI prediction without the parameters that make up the index, expand target parameters to nutrients and priority contaminants, and integrate XAI routinely. Future reviews are also advised to extract performance metrics for all included studies and to search additional databases beyond Scopus.

5.1. Conflict of Interest Statement

The author declares that there is no conflict of interest regarding the publication of this paper.

5.2. Author Contributions

The sole author was responsible for Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing.

5.3. Funding Statement

No funding was received for this research.

5.4. Informed Consent

Informed consent was not required, as this study is a literature review and did not involve human subjects.

5.5. Ethical Approval

This article does not contain any studies with human participants performed by the author.

5.6. Data Availability

The data that support the findings of this study are available from the corresponding author [A.S.] upon reasonable request.

5.7. Declaration of Generative AI and AI-assisted Technologies

During the preparation of this manuscript, the author utilized Claude (Sonnet 5.5, Anthropic) for proofreading, limited to checking grammar, spelling, and clarity of expression. All outputs generated by the AI tool underwent rigorous critical review and thorough editing by the author to ensure factual accuracy, enhance clarity of expression, and confirm adherence to established academic standards. The author assumes full responsibility for the integrity and content of this manuscript.

6. References

Ab Karim, M. A., Wan Ismail, W. Z., Mohd Shuib, F. M., Ab Aziz, N. A., & Ghazali, A. K. (2026). Water Quality Monitoring and Assessment Using Machine Learning: A Review of Formulation, Modeling Approaches, and Explainable Artificial Intelligence. *Environments*, *13*(5), 267. https://doi.org/10.3390/environments13050267

Al-Mukhtar, M., Srivastava, A., Khadke, L., Al-Musawi, T., & Elbeltagi, A. (2024). Prediction of Irrigation Water Quality Indices Using Random Committee, Discretization Regression, REPTree, and Additive Regression. *Water Resources Management*, *38*(1), 343–368. https://doi.org/10.1007/s11269-023-03674-y

Al-Sulttani, A. O., Al-Mukhtar, M., Roomi, A. B., Farooque, A. A., Khedher, K. M., & Yaseen, Z. M. (2021). Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction. *IEEE Access*, *9*,108527–108541. https://doi.org/10.1109/access.2021.3100490

Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water Quality Prediction Using Artificial Intelligence Algorithms. *Applied Bionics and Biomechanics*, *2020*, 1–12. https://doi.org/10.1155/2020/6659314

Alqahtani, A., Shah, M. I., Aldrees, A., & Javed, M. F. (2022). Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality. *Sustainability*, *14*(3), 1183. https://doi.org/10.3390/su14031183

Cabri, G., & Rahimi, A. (2025). A Comparative Study of Machine Learning Algorithms for Water Quality Prediction Using SHAP-based Explainability. In *2025 33rd International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 1–6). IEEE. https://doi.org/10.1109/wetice67341.2025.11091841

Catajan Jr, A. L., Fajardo, A. C., & Limbago, J. S. (2023). Classification of Water Quality Index in Laguna de Bay using XGBoost. In *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 403–408). IEEE. https://doi.org/10.1109/jcsse58229.2023.10202029

Chakraborty, M., Ismail, M., Basak, J. K., Mohinuzzaman, M., Paudel, B., & Halim, A. (2026). Modeling of BOD and COD Concentrations in Aquaculture Sites of the Southeastern Coastal Region of Bangladesh Using Machine Learning. *Water Environment Research*, *98*(7). https://doi.org/10.1002/wer.70480

Costa, D., Bayissa, Y., Barbosa, K. V., Villas-Boas, M. D., Bawa, A., Junior, J. L., Neto, A. J. S., & Srinivasan, R. (2024). Water quality estimates using machine learning techniques in an experimental watershed. *Journal of Hydroinformatics*, *26*(11), 2798–2814. https://doi.org/10.2166/hydro.2024.132

Daniels, A., & Koutsougeras, C. (2021). Predicting Water Quality Parameters in Lake Pontchartrain using Machine Learning. In *2021 the 5th International Conference on Information System and Data Mining* (pp. 28–33). ACM. https://doi.org/10.1145/3471287.3471308

de Lima, L., Capriles, P., & Oliveira, N. (2025). Comparison of Machine Learning Algorithms in Water Quality Index Prediction: A Case Study in Juiz De Fora, Brazil. In *Lecture Notes in Networks and Systems* (pp. 43–54). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-6438-2_3

Ehmimed, N., Chkouri, M. Y., & Touhafi, A. (2025). Temporal SMOTE and Model Diversity: Boosting Accuracy in Water Quality Forecasting. In *Lecture Notes in Networks and Systems* (pp. 103–114). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-07915-2_9

Ejaz, U., Khan, S. M., Jehangir, S., Ahmad, Z., Abdullah, A., Iqbal, M., Khalid, N., Nazir, A., & Svenning, J.-C. (2024). Monitoring the Industrial waste polluted stream - Integrated analytics and machine learning for water quality index assessment. *Journal of Cleaner Production*, *450*, 141877. https://doi.org/10.1016/j.jclepro.2024.141877

Elmotawakkil, A., Enneya, N., Bhagat, S. K., Ouda, M. M., & Kumar, V. (2025). Advanced machine learning models for robust prediction of water quality index and classification. *Journal of Hydroinformatics*, *27*(2), 299–319. https://doi.org/10.2166/hydro.2025.290

García del Toro, E. M., Mateo, L. F., García-Salgado, S., Más-López, M. I., & Quijano, M. Á. (2022). Use of Artificial Neural Networks as a Predictive Tool of Dissolved Oxygen Present in Surface Water Discharged in the Coastal Lagoon of the Mar Menor (Murcia, Spain). *International Journal of Environmental Research and Public Health*, *19*(8), 4531. https://doi.org/10.3390/ijerph19084531

He, M., Qian, Q., Liu, X., & Zhang, J. (2025). River Dissolved Oxygen Prediction Using Machine Learning Models and Wireless Sensor Measurements. *Journal of Hydrologic Engineering*, *30*(5). https://doi.org/10.1061/jhyeff.heeng-6522

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

Husain, M. S., Fatima, K., Hussain, M. S., Viswambharan, A., & Idrisi, A. H. (2025). Machine Learning Based Water Quality Classification in Geothermally Active Regions. In *2025 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1–6). IEEE. https://doi.org/10.1109/aset66891.2025.11428147

Jena, P. K., Rahaman, S. M., Das Mohapatra, P. K., Barik, D. P., & Patra, D. S. (2022). Surface water quality assessment by Random Forest. *Water Practice and Technology*, *18*(1), 201–214. https://doi.org/10.2166/wpt.2022.156

Khoshvaght, H., Permala, R. R., Razmjou, A., & Khiadani, M. (2025). A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction. *Journal of Environmental Chemical Engineering*, *13*(6), 119675. https://doi.org/10.1016/j.jece.2025.119675

Krishna, S., & Sudha, I. (2025). Prediction of water quality using novel support vector machine algorithm based on feature selection over random forest for improved accuracy. In *AIP Conference Proceedings* (Vol. 3318, p. 20108). AIP Publishing. https://doi.org/10.1063/5.0277261

Krushna, B. V., & Sasikala, D. (2024). Comparative Analysis of Machine Learning Models for Water Quality Prediction. In *2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1–6). IEEE. https://doi.org/10.1109/icaect60202.2024.10468812

Kularbphettong, K., Raksuntorn, N., & Boonseng, C. (2025). Prediction of Water Quality Index (WQI) Using Machine Learning. *International Journal of Environmental Science and Development*, *16*(1), 34–40. https://doi.org/10.18178/ijesd.2025.16.1.1507

Lokman, A., Ismail, W. Z. W., & Aziz, N. A. A. (2025). A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis. *Water*, *17*(15), 2243. https://doi.org/10.3390/w17152243

Lokman, A., Wan Ismail, W. Z., & Ab Aziz, N. A. (2025). Enhancing water quality index prediction accuracy in Mranti lake and rivers in Malaysia using regression forest model. *Applied Water Science*, *16*(2). https://doi.org/10.1007/s13201-025-02705-w

Mohamed, G., Yassine, G., Hamza, B., Tachi, S. E., Hasnaoui, Y., & Ammar, M. (2025). Assessment and prediction of irrigation water quality using machine learning techniques: a case study from the Bouhamdane Basin, Guelma Region, Eastern Algeria. *Water Supply*, *25*(11), 1465–1481. https://doi.org/10.2166/ws.2025.087

Najafabadi, E. F., Shojaei, P., & Askarizadeh, M. (2025). Comparative analysis of machine learning models for predicting river water quality: a case study of the Zayandeh Rood River. *Results in Engineering*, *27*, 106665. https://doi.org/10.1016/j.rineng.2025.106665

Pisal, P. S., Varshney, N., Chamoli, A. S., & Singh, A. (2025). Prediction of Water Quality using Machine Learning and Artificial Intelligence Techniques. In *2025 7th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1–6). IEEE. https://doi.org/10.1109/iscon65210.2025.11341082

Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Soumya, K., & Poornema, A. J. (2021). Prediction on water quality of a lake in Chennai, India using machine learning algorithms. *Desalination and Water Treatment*, *218*, 44–51. https://doi.org/10.5004/dwt.2021.26970

Rawat, P., Bajaj, M., Sharma, V., & Vats, S. (2023). A Comprehensive Analysis of the Effectiveness of Machine Learning Algorithms for Predicting Water Quality. In *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* (pp. 1108–1114). IEEE. https://doi.org/10.1109/icidca56705.2023.10099968

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

Shah, M. I., Alaloul, W. S., Alqahtani, A., Aldrees, A., Musarat, M. A., & Javed, M. F. (2021). Predictive Modeling Approach for Surface Water Quality: Development and Comparison of Machine Learning Models. *Sustainability*, *13*(14), 7515. https://doi.org/10.3390/su13147515

Sharma, A., Sharma, R., Rana, R., & Kalia, A. (2024). Water quality prediction using Machine Learning Models. *E3S Web of Conferences*, *596*, 1025. https://doi.org/10.1051/e3sconf/202459601025

Vaishali, M., Monika, M., Atish, M., & Lalitha. (2025). A Water Quality Prediction and Assessment Model using Machine Learning Classifiers. In *AI and ML Solutions Driving Modern Farming and Urban Innovation* (pp. 153–168). BENTHAM SCIENCE PUBLISHERS. https://doi.org/10.2174/9798898812102125030014

Voza, D., Babić, G., Vuković, M., & Ranisavljević, I. M. (2025). Comparative analysis of machine learning algorithms efficacy in water resources management. *International Journal of Environment and Waste Management*, *38*(3), 295–323. https://doi.org/10.1504/ijewm.2025.149522

Walczak, N., & Walczak, Z. (2025). Assessing the feasibility of using Machine learning algorithms to determine reservoir water quality based on a reduced set of predictors. *Ecological Indicators*, *175*, 113556. https://doi.org/10.1016/j.ecolind.2025.113556

Wang, M. (2025). Enhancing water quality parameter prediction accuracy using a hybrid machine learning and optimization approach. *Signal, Image and Video Processing*, *19*(13). https://doi.org/10.1007/s11760-025-04635-1

WHO/UNICEF. (2023). *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2022: Special Focus on Gender*. WHO/UNICEF Joint Monitoring Programme for Water Supply Sanitation and Hygiene. https://data.unicef.org/resources/jmp-report-2023/

WHO. (2023). *Drinking-Water*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/drinking-water

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

Yaseen, Bazai, S. U., Ghafoor, M. I., Shahreen, M., & Marjan, S. (2024). Water Quality Prediction Using Random Forest Classifier: An Analysis of Chemical Attributes and Their Feature Importance. In *2024 5th International Conference on Innovative Computing (ICIC)* (pp. 1–7). IEEE. https://doi.org/10.1109/icic63915.2024.11116234