Results in Engineering 26 (2025) 105182

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Review article

A review of machine learning and internet-of-things on the water quality assessment: Methods, applications and future trends

Gangani Dharmarathne [a], A.M.S.R. Abekoon [b], Madhusha Bogahawaththa [c], Janaka Alawatugoda [d], D.P.P. Meddage [c,e,*]

[a] Australian Laboratory Services Global, Water and Hydrographic, Hume ACT 2620, Australia
[b] Department of Aquaculture & fisheries, University of Wayamba, Kuliyapitiya, Sri Lanka
[c] School of Engineering and Technology, The University of New South Wales, Canberra ACT 2600, Australia
[d] Research and Innovations Centers Division, Rabdan Academy, Abu Dhabi, United Arab Emirates
[e] Ceylon Institute for Artificial Intelligence Research (CIAIR), Colombo, Sri Lanka

ARTICLE INFO

Keywords:
Internet of things
Machine learning
Review
Prediction
Water pollution
Water quality

ABSTRACT

Clean and safe water is fundamental to human health and environmental sustainability, yet increasing pollution due to urbanisation, industrialisation, and climate change poses significant risks. Traditional water quality monitoring relies on manual sampling and laboratory analysis, which are often costly, time-intensive, and lack real-time insights. This review critically examines the integration of machine learning (ML) and the internet of things (IoT) for real-time water quality monitoring and predictive analytics. The study evaluates peer-reviewed research published between 2016 and 2024, focusing on advancements, limitations, and future trends in automated water quality assessment. ML models, including random forest, extreme gradient boosting, support vector machines, and neural networks, have been more frequently used in water quality research and have achieved high accuracies (R^2=0.99 in regression and 0.99 accuracy metric in classification). Explainable AI (XAI) which can explain the decision making process of ML, is underutilised, appearing in only a few recent studies. While IoT significantly improves real-time contamination detection, persistent challenges remain in sensor fouling, data continuity, data privacy, network reliability, and cybersecurity. Such challenges can hinder the scalability and effectiveness of long-term IoT implementations. Integrating IoT with machine learning enhances water quality monitoring by enabling real-time data collection, remote tracking, and predictive analytics. This synergy improves efficiency, addresses monitoring challenges, and supports sustainable water management.

1. Introduction

Clean and safe water is essential for human health, agricultural productivity and ecological stability. However, water pollution has become one of the most critical environmental challenges which is governed by rapid urbanisation, industrialisation and climate change [1–5]. Globally, water resources are increasingly at risk from pollutants such as agricultural runoff, industrial discharge and untreated domestic wastewater [6–8]. The impacts of water contamination are far-reaching, with approximately 250 million cases of waterborne diseases reported annually [9], alongside significant economic losses and biodiversity decline.

Safe and good-quality water is essential for public health, regardless of its intended use. Fig. 1 shows the ambient water quality from 2017 to 2020, based on data obtained from the world health organisation (WHO). As shown, many countries lack data on water quality, while others struggle with access to clean water. Contaminated water and poor sanitation are often linked to many transmissible diseases, such as

Abbreviations: BLE, Bluetooth Low Energy; CNN, Convolutional Neural Networks; CRWU, Creating Resilient Water Utilities; DO, Dissolved Oxygen; EPA, Environmental Protection Agency; GIS, Geographic Information System; GSM, Global System for Mobile Communications; IoT, Internet of Things; KNN, K-Nearest Neighbors; LoRaWAN, Long Range Wide Area Network; LPWAN, Low Power Wide Area Network; LSTM, Long Short-Term Memory; ML, Machine Learning; NB-IoT, Narrowband IoT; PCA, Principal Component Analysis; PUB, Public Utilities Board; RMSE, Root Mean Squared Error; RNN, Recurrent Neural Networks; SDG, Sustainable Development Goal; SHAP, Shapley Additive Explanation; SVM, Support Vector Machine; TOC, Total Organic Carbon; WHO, World Health Organisation; WQI, Water Quality Index; WSN, Wireless Sensor Network; XAI, Explainable Artificial Intelligence.

* Corresponding author at: School of Engineering and Information Technology, University of New South Wales, Canberra, Australia.
E-mail address: p.meddage@unsw.edu.au (D.P.P. Meddage).

https://doi.org/10.1016/j.rineng.2025.105182
Received 29 January 2025; Received in revised form 27 April 2025; Accepted 30 April 2025
Available online 1 May 2025

diarrhea and cholera. Therefore, evaluating water quality is one of the most important tasks to ensure the health, safety and well-being of the community.

Water quality comprises the physical, chemical and biological attributes of water. Assessing and monitoring water quality is crucial, as it allows the timely detection of environmental issues caused by pollutants from human activities [11]. This process is important both in the short and long term. It also helps to identify and hold accountable those who violate regulations [12].

Water quality is often measured using indices such as the water quality index (WQI), which consolidates multiple parameters into a single dimensionless score [13]. Nine commonly used water quality indices are demonstrated in Fig. 2, including dissolved oxygen, oxidation–reduction potential, chemical oxygen demand and biological oxygen demand, temperature, electrical conductivity, total dissolved solids, turbidity and acidity.

Traditional methods of water quality monitoring have primarily relied on manual sampling and laboratory testing. Even though these approaches deliver accurate and detailed data, they are mainly constrained by high costs, labour-intensive processes and an inability to provide real-time insights. Moreover, manual sampling often fails to capture short-term fluctuations or to monitor remote or inaccessible areas effectively. These limitations emphasise the urgent need for more efficient, scalable and real-time solutions to manage water quality.

The use of machine learning (ML) and the internet of things (IoT) provides significant potential in addressing previous challenges. ML provides advanced algorithms for analysing complex, multidimensional datasets which enables the identification of patterns, prediction of water quality and detection of trends and anomalies that traditional methods overlook [8,14–20]. IoT systems complement this by facilitating continuous, real-time monitoring using sensors, communication networks and cloud-based analytics [21]. These technologies work together to provide stakeholders with valuable insights, helping them make informed decisions and adopt proactive, data-driven approaches to managing water resources.

The integration of ML and IoT aligns with global objectives such as the United Nations' sustainable development goals (SDGs). In particular, these technologies advance SDG 6 (Clean Water and Sanitation) by improving access to safe drinking water and SDG 14 (Life Below Water) by supporting the protection of aquatic ecosystems through effective pollution control. ML and IoT technologies play a critical role in ensuring water security for future generations by supporting more sustainable and informed water management practices.

1.1. Research gap for the review study

Many researchers have utilised ML and IoT-based systems to predict/monitor water quality. Unlike previous studies, this paper systematically examines the synergistic integration of ML and IoT for water quality assessment. Previous review studies have mostly focused on the applications of ML, the architectures of ML, and the applications of IoT in water quality research. Those studies did not explicitly describe the IoT methods, and the persisting challenges in the modern era. In addition, previous review studies did not report on the explainability of machine learning predictions, which has become popular nowadays. Explainability refers to revealing the decision-making process of machine learning models such that the end user knows the impact of input parameters on the water quality predictions. Therefore, the authors focused more on these areas which have not been thoroughly covered in previous review studies.

2. Machine learning in water quality evaluation

2.1. Overview of ML in water quality evaluation

Recently, machine learning has been widely used in water quality assessment which can identify patterns in data and make predictions that traditional statistical methods might overlook. The initial steps of water quality evaluation involve data collection and preprocessing through various techniques such as sensor networks, remote sensing, laboratory analysis, historical data, etc. [22]. Sensor networks gather real-time data from water quality sensors measuring parameters like pH, temperature, conductivity, turbidity, dissolved oxygen (DO), chlorine, nitrate, ammonia, total organic carbon (TOC), heavy metals and microbial presence [23,24]. Remote sensing has also been used in observing water bodies [25,26]. In the current context, it uses satellite imagery and aerial photography to monitor large water bodies, while laboratory analysis involves chemical, physical and biological examination of collected water samples [27,28]. Historical data collection utilises previously recorded water quality data to analyse patterns over time [29]. The next important step is data preprocessing which involves removing noise and missing values in the dataset to ensure accuracy [30]. Next, the data are scaled/normalised to ensure uniformity across different parameters. Afterwards, the correlations in the data can be assessed in order to select the correct algorithm which is paramount [31]. These preprocessing steps are essential for optimising the performance of machine learning models in water quality assessment.

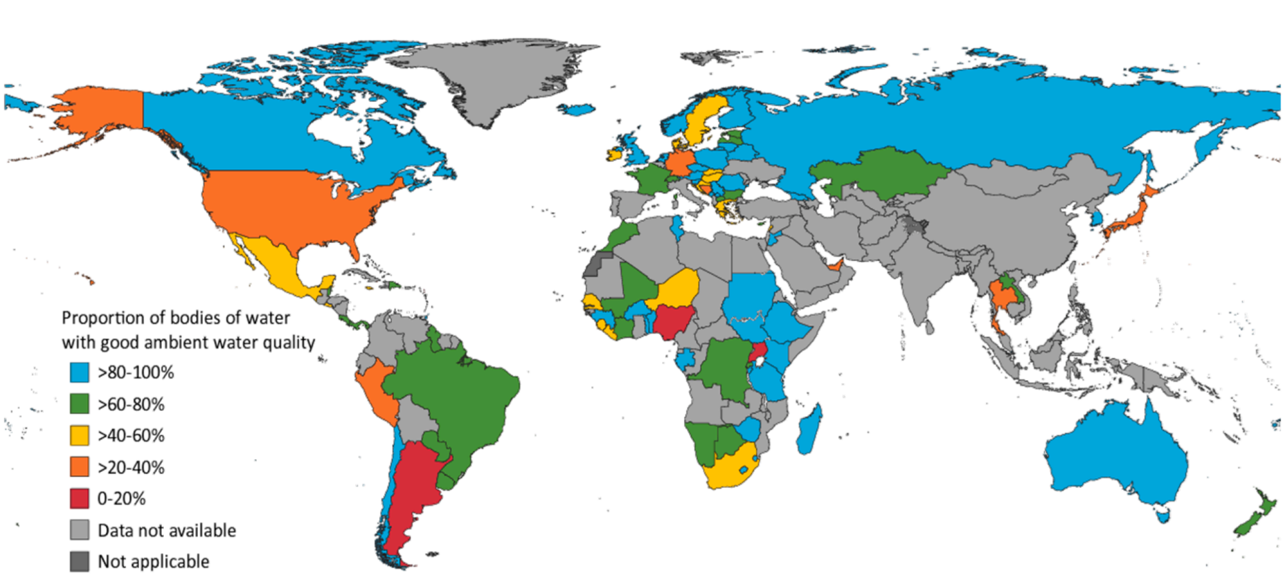

Fig. 1. Distribution of the ambient water quality among the water bodies across the world (2017–2020) Source: World Health Organisation [10].

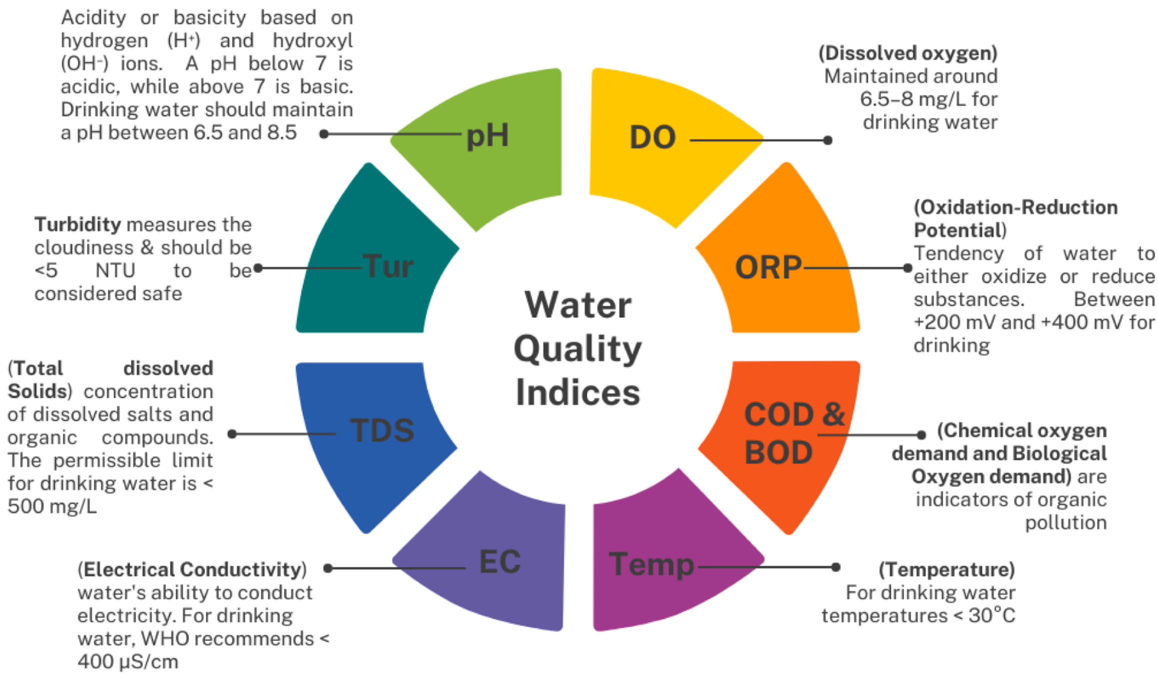

Fig. 2. Commonly used water quality indices.

As these are mostly labelled data (inputs and outputs are well defined), supervised learning like regression and classification have been often used in previous research. The generalised methodology of classification/regression is shown in Fig. 3. Classification refers to predicting a class (good/bad) and regression refers to predicting a continuous value (water quality index) [32–34]. On the other hand, clustering is one type of unsupervised learning technique that can group the water quality data based on certain characteristics (e.g. K-Means and hierarchical clustering) [35]. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) are two dimensionality reduction techniques that reduce dimensionality while retaining critical information [36]. Deep learning methods like convolutional neural networks (CNNs) for spatial data and Recurrent neural networks (RNNs) for time-series data have been used to capture non-linear relationships in water quality data [37,38]. Recent work in water quality research using ML includes classical machine learning, deep learning and hybrid models (machine learning models optimised with metaheuristic algorithms) [39,40]. Key applications of ML in water quality include forecasting, anomaly detection, classification, treatment optimisation, and resource management.

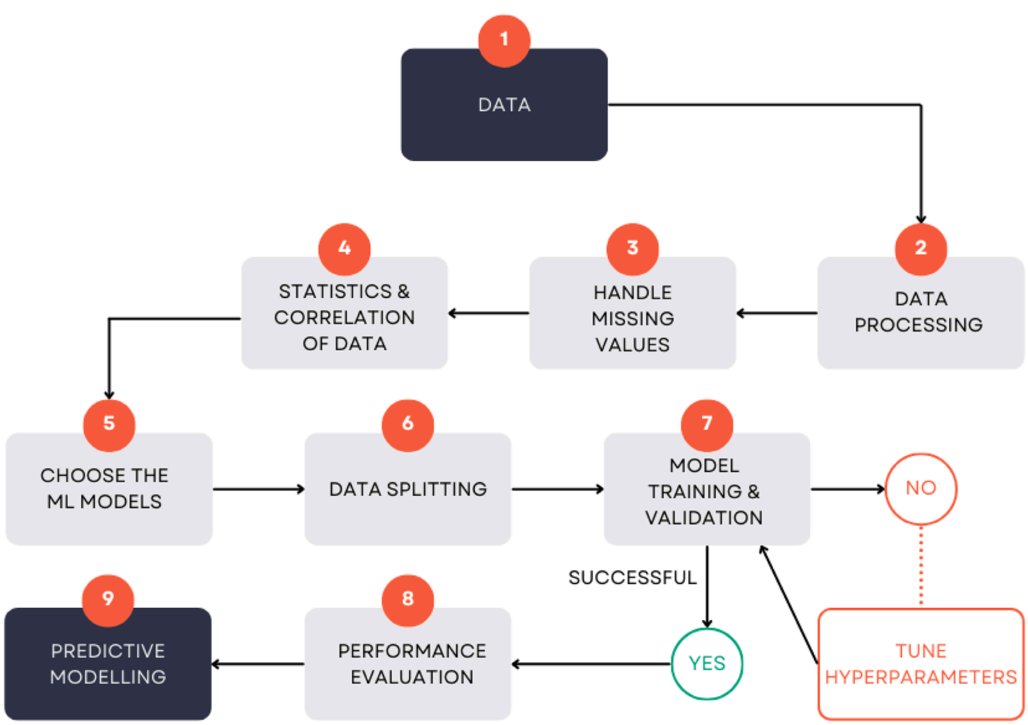

Fig. 3. Generalised framework for machine learning based regression/classification of water quality.

2.2. Prediction and forecasting of water quality

Prediction/forecasting mainly refers to regression models which yield continuous output. By using historical or real-time data, trained ML models can predict present/future water quality parameters. ML methods based on neural networks can analyse historical water quality data to predict future conditions [41,42]. These predictions help in understanding seasonal variations and the impact of external factors like industrial discharge or agricultural runoff on water quality [43]. For instance, predicting nitrate levels in a river can inform agricultural practices to minimise fertiliser runoff, thus protecting water quality [44].

Real-time forecasting uses data from sensor networks to provide immediate predictions on water quality [37]. Techniques like time-series analysis with recurrent neural networks (RNNs) are particularly effective in real-time forecasting [45,46]. These models can forecast parameters like pH, dissolved oxygen (DO) and turbidity which allows timely decision making. For example, if a significant drop in DO levels is predicted, immediate actions can be taken to aerate the water, preventing threats to aquatic life [47].

ML models can simulate various scenarios to predict their impact on water quality. This includes modelling the effects of climate change, population growth, or new industrial developments [48]. Scenario analysis helps policymakers and water managers develop long-term strategies for sustainable water management [49]. For instance, predicting the impact of increased urbanisation on a watershed can guide infrastructure development and pollution control measures [50].

Moreover, ML-driven early warning systems can alert authorities about potential water quality issues before they become critical. These systems integrate data from various sources, including sensors and weather forecasts, to predict events like harmful algal blooms or contamination incidents [51]. By providing early warnings, these systems enable timely preventive actions, protecting public health and the environment.

2.3. Water quality classification

Classification models are used to categorise water quality into different classes, such as safe, moderate and polluted [32]. Supervised learning models like decision trees, random forest and support vector machines have been commonly used for water quality classification [52]. These models are trained on labelled datasets, where water quality samples are categorised based on their parameters. For instance, a random forest model can classify water samples like potable, recreational, or agricultural based on parameters like pH, turbidity and microbial content [53]. Multiclass classification extends the binary (two-class) classification approach to categorise water quality into multiple classes [54]. Neural networks, particularly convolutional neural networks (CNNs), are effective in handling multiclass classification problems in water quality research [55]. These models can classify water quality into categories, such as excellent, good, fair and poor, providing a good understanding of water conditions. Threshold-based classification uses predefined thresholds for various water quality parameters to categorise water samples [56]. This approach can be combined with ML models to enhance accuracy. For example, a model might use thresholds for parameters like nitrate and phosphate levels to classify water samples into classes indicating different levels of pollution [57]. Also, the use of ensemble methods has recently become popular in this context. Ensemble learning techniques combine multiple ML models to improve classification performance. Methods like boosting and bagging enhance the robustness and accuracy of classification models. For instance, an ensemble model combining decision trees and support vector machines (SVMs) achieved higher classification accuracy for detecting water pollution compared to individual models [58,59].

Table 1 shows the use of machine learning methods in recent years in water quality research, including regression and classification. It is observed that many different algorithms have been used, and they have accurately predicted the water quality. In addition, recently, explainable artificial intelligence (XAI) methods have been used in water quality-related research to elucidate the decision-making process of machine learning models [60–62]. It is noteworthy that XAI methods neither affect models' accuracy nor increase the models' complexity, but rather underpin the decision-making process as shown in Fig. 4.

XAI methods can be mainly divided into *data-driven* XAI and *model-driven* XAI methods. Data-driven methods mainly focus on how a change in the inputs can affect the model predictions. On the contrary, model driven methods rely on models' internal mechanisms for the explanations. However, the authors noticed that XAI is very new to water quality research. For example, data driven methods like Shapley additive explanations (SHAP) have been widely used in water quality research. It gives a unique importance to each feature such that the users can identify which inputs mainly affect the quality of water [64,86].

Table 1 of the manuscript provides an evaluation of different ML models applied in water quality prediction, including classification and regression-based approaches. Based on the accuracy, and F1-score for classification, coefficient of determination (R^2), and root mean square error (RMSE) for regression reported across various datasets, their effectiveness can be compared in different water quality assessment scenarios.

Among classification models, ensemble methods such as extreme gradient boosting and random forest consistently outperform traditional statistical models, achieving accuracy levels closer to 99 %. These models are particularly effective in high-dimensional datasets and are widely used for categorising water quality into predefined classes such as "safe" or "polluted" [34,65–67]. Additionally, support vector machines display good performance in water quality classification [67]. In contrast, KNN shows lower and more inconsistent accuracy, making it less suitable for complex water quality assessments [66].

For continuous variable predictions (regression), such as estimating pH, dissolved oxygen, or heavy metal concentrations, Extreme gradient boosting, ANNs and random forest provide the most accurate results [68–71], with R^2 values reaching 99 % in certain studies. In addition, methods like long short-term memory (LSTM) and CNNs are highly dependent on large datasets and computational power, which may limit their application in regions with constrained resources. Table 2 provides a comparative analysis of commonly used machine learning models, highlighting their strengths and weaknesses in water quality prediction. Fig. 5 presents a comparative summary of the performance of various machine learning algorithms in terms of their frequency of use, minimum and maximum of accuracy % based on classification and R^2 % based on regression. The analysis highlights the variability in predictive capabilities among commonly used models in water quality research. Random Forest appeared most frequently ($n = 16$), demonstrating a broad range of accuracies, from a minimum of 12 % to a maximum of 99 %. Similarly, the artificial neural network was employed 15 times, yielding accuracies between 59 % and 99 %, indicating strong predictive accuracy when adequately trained and tuned.

Support vector machines and extreme gradient boosting were also commonly utilised, with frequencies of 12 and 11, respectively. Both models demonstrated high potential, with SVM showing accuracies ranging from 38 % to 96 %, and Extreme gradient boosting achieving up to 100 % accuracy in some cases, though with a lower bound of 59 %.

Simple models such as K-nearest neighbour, linear regression, and decision trees have often achieved reasonably good accuracies. Overall, more complex models—including neural networks, extreme gradient boosting, random forests, and support vector machines—have been widely adopted in water quality research and have demonstrated strong predictive performance. Notably, the recent use of ensemble models (i.e., combinations of multiple individual learners) has led to improved lower bounds in prediction accuracy for water quality assessments.

In terms of machine learning interpretability, at present, a few XAI methods have been used. Since water quality data are mostly structured,

Table 1
Performance comparison of machine learning models for water quality prediction.

Ref.	Machine learning models	Accuracy (%) Classification	F1 Score (%)	R^2 (%) Regression	RMSE	Data Source	Use of XAI
[67]	K-nearest neighbour	91	83	–	–	Water quality in Cork Harbour	Permutation Feature importance
	Naïve Bayes	94	87	–	–		
	Support vector machine	92	81	–	–		
	Extreme gradient boosting	100	99	–	–		
[65]	Support vector machine	88	78	–	–	Pearl River Estuary	Permutation Feature importance
	Random forest	98	94	–	–		
	Adaptive boosting	80	57	–	–		
	Gradient boosting	98	99	–	–		
	Ensemble model	99	99	–	–		
[66]	Decision tree	60	–	–	–	Kaggle water quality dataset	–
	K-nearest neighbour	61	–	–	–		
	Logistic regression	62	–	–	–		
	Random forest	63	–	–	–		
	Support vector regression	66	–	–	–		
	Adaptive boosting	64	–	–	–		
	Naïve Bayes	65	–	–	–		
	Extreme gradient boosting	70	–	–	–		
[72]	Multi-expression programming	–	–	96–97	11–16	Indus River	Permutation Feature importance
	Random forest	–	–	98	6–13		
[59]	Naïve Bayes	92	–	–	–	Kaggle water quality dataset	–
	Linear regression	79	–	–	–		
	Multilayer perceptron	92	–	–	–		
	K-nearest neighbour	88	–	–	–		
	Random forest	97	–	–	–		
	RotF	95	–	–	–		
	Adaptive boosting	97	–	–	–		
	Stacking	98	–	–	–		
	Bagging	96	–	–	–		
	Voting	93	–	–	–		
[34]	Random forest	99	98	–	–	Kaggle water quality dataset	–
	Extreme gradient boost	99	99	–	–		
	Adaptive boosting	99	99	–	–		
	Gradient boosting	99	99	–	–		
[73]	Support vector machine	–	–	36–76	0.4–12.4	El Kharga Oasis	–
	Adaptive neuro fuzzy interference system	–	–	68–100	0.2–5.9		
[74]	Random forest	–	–	84	11	An-Kim Hai irrigation system	Permutation Feature importance
	Decision tree	–	–	79	12		
	Multilayer perceptron	–	–	73	14		
	Support vector machine	–	–	53	18		
	Linear regression	–	–	42	20		
[68]	Extreme gradient boosting	–	–	73–90	0.08–0.25	Shenzhen	–
	Support vector regression	–	–	46–85	0.14–0.35		
	Random forest	–	–	12–77	0.22–0.34		
	Artificial neural network	–	–	15–79	0.17–0.44		
[75]	Ensemble trees	–	–	–	–	Savar	–
	Linear regression	–	–	–	–		
	Gaussian process regression	–	–	–	–		
	Artificial neural network	–	–	–	–		
	Support vector regression	–	–	–	–		
	Regression tree	–	–	–	–		
[76]	Multiple linear regression	–	–	9–63	–	Southeastern part of Obubra	Permutation Feature importance
	Multilayer perceptron	–	–	71–98	–		
[69]	Gradient boosting regression	–	–	55–95	0.05–4.1	Doucen plain, Algeria	–
	Artificial neural network	–	–	95–99	0–2.1		
[77]	Gene expression programming	–	–	94–99	9–37	Upper Indus Basin	Shapley additive explanations
	Deep neural network	–	–	92–99	0–55		
	Gaussian process regression	–	–	86–98	13–45		
[78]	Random forest	Near 75	Between 37–50	–	–	Namhan River Basin	Shapley additive explanations
	Extreme gradient boosting						
	Light gradient boosting			–	–		
	Artificial neural network			–	–		
	Convolutional neural network			–	–		
	TabNet			–	–		
[79]	Gradient boosting trees	–	–	–	–	Lake Loktak	Shapley additive explanations, Accumulated local effects
	Random forest	–	–	–	–		
	Deep neural network	–	–	–	–		
[80]	Multilayer perceptron	–	–	59–96	0.54–0.74	Wadi Dayqah Dam	Shapley additive explanations
	Support vector regression	–	–	59–96	0.58–0.74		
	Adaptive boosting	–	–	61–92	0.72–0.79		
	Least Absolute Shrinkage and Selection Operator	–	–	41–87	0.88–1.02		

(continued on next page)

Table 1 (*continued*)

Ref.	Machine learning models	Accuracy (%) Classification	F1 Score (%)	R^2 (%) Regression	RMSE	Data Source	Use of XAI
[81]	Decision tree	–	–	44–87	–	Suzhou	Shapley additive explanations
	Random forest	–	–	79–90	–		
	Gradient boosting tree	–	–	85–89	–		
	Support vector machine	–	–	38–69	–		
[70]	Artificial neural network	–	–	93	9.8	Shipra River, Ujjain	–
	Multiple linear regression	–	–	97	3.9		
	Support vector machine	–	–	94	8.2		
	Random forest	–	–	97	3.6		
	Extreme gradient boosting	–	–	99	3.2		
	Ensemble model	–	–	97	3.4		
[82]	Long-short term memory	–	–	85–95	0.01–0.09	Swan Canning Estuary	–
[61]	Extreme gradient boosting	–	–	–	1.872	GeumRiver, South Korea	Shapley additive explanations Partial Dependency plots
[83]	Random Forest	–	–	46–69	0.88–1.74	Texas Gulf Region	Shapley additive explanations
[71]	Multiple linear regression	–	–	60–82	1.1–2.4	Xiaoqing River	Shapley additive explanations
	Artificial neural network	–	–	64–82	1.3–2.3		
	Random forest	–	–	72–88	1.1–2.1		
	Extreme gradient boosting	–	–	77–90	1–1.8		
[84]	Random forest	–	–	82–99	0.09–0.29	Chesapeake Bay watershed	Shapley additive explanations
[85]	Extreme gradient boosting	–	–	59–85	0.019–0.226	Lake Okeechobee	Shapley additive explanations
	Light gradient boosting	–	–	57–84	0.019–0.230		
	Random forest	–	–	54–82	0.021–0.243		
	Support vector regression	–	–	45–80	0.023–0.266		

the following XAI methods can be used in future research. The authors observed that the majority of studies have used the SHAP method. Future research can investigate different XAI methods and how they rank the input parameters. Studies suggest that XAI can reveal the underlying reasoning behind ML predictions. However, no study has been conducted to verify the consistency of the explanations, which could be assessed using a questionnaire among domain experts. Table 3 shows a feature comparison of XAI approaches that can be considered in future work in water quality research.

2.4. Anomaly detection

Anomaly detection in this domain refers to identifying unusual patterns or sudden changes in water quality parameters that can indicate pollution events, system failures, or other issues requiring immediate attention. Unsupervised learning techniques like clustering and principal component analysis (PCA) are effective in detecting anomalies in water quality data [96,97]. These methods identify data points that deviate significantly from the norm, flagging them as potential anomalies. For example, K-means clustering can group similar water quality readings and highlight outliers that may indicate contamination [98].

Time-series analysis using models like long short-term memory (LSTM) networks can detect anomalies over time [96]. These models learn the normal patterns of water quality parameters and identify deviations from these patterns [99,100]. For instance, an LSTM model could detect an unexpected spike in heavy metal concentrations, prompting further investigation and corrective actions [101]. Also, rule-based systems can be integrated with ML models to enhance anomaly detection. These systems use predefined rules based on domain knowledge to flag anomalies [102]. For example, if the turbidity exceeds a certain threshold while the pH remains within normal ranges, the system might flag it as an anomaly, suggesting potential sediment contamination.

Also, integrating ML with physical hydrology models can significantly improve water quality predictions by combining data-driven insights with mechanistic environmental processes. Recent studies have explored hybrid approaches that merge ML algorithms with hydrodynamic and pollutant transport models. For instance, a study on the Great Lakes demonstrated that combining LSTM networks with a physics-based hydrodynamic model enhanced predictions of lake surface temperature, highlighting the synergy between ML and traditional modelling techniques [103]. Similarly, a three-dimensional (3D) hydrodynamic and water quality model was developed to simulate floating treatment wetlands, integrating physical flow structures with nutrient removal processes [104]. These examples suggest that fusing ML with hydrodynamic models could provide more reliable and interpretable water quality forecasts, particularly when dealing with complex aquatic systems.

Combining ML with real-time monitoring systems enables continuous surveillance of water quality. ML models process data from sensors in real-time, detecting anomalies instantly [105]. This real-time capability is crucial for early detection of pollution events or equipment failures, allowing for swift responses. For instance, a sudden drop in conductivity detected by real-time monitoring can signal a possible contamination event which can trigger immediate attention [106].

Fig. 6 shows the results from an anomaly detection method proposed by Tharayil et al. [107]. They used a hybrid deep learning which consists of long short-term memory networks, autoencoders with convolution and attention layers. The model was expected to calculate anomalies and predict using temporal features.

2.5. Optimisation of treatment processes

Machine learning can be used to optimise the operations of water treatment plants by predicting the necessary adjustments in treatment processes, improving efficiency and effectiveness. ML models can predict when equipment in water treatment plants is likely to underperform, allowing for timely maintenance and preventing breakdowns. Predictive maintenance models analyse historical data on equipment performance and failure rates to forecast maintenance needs [108]. For example, an ML model can predict the wear and tear of pumps and suggest maintenance schedules, reducing downtime and maintenance costs [109]. These include processes such as chemical dosing, filtration and disinfection. These models analyse real-time data to adjust treatment parameters dynamically, ensuring optimal performance [87]. For instance, an ML model can optimise the chlorine dosing process based on real-time measurements of microbial content, ensuring effective disinfection while minimising chemical usage [110]. Energy consumption is a significant operational cost in water treatment plants. ML models can optimise energy usage by predicting the most energy-efficient operating

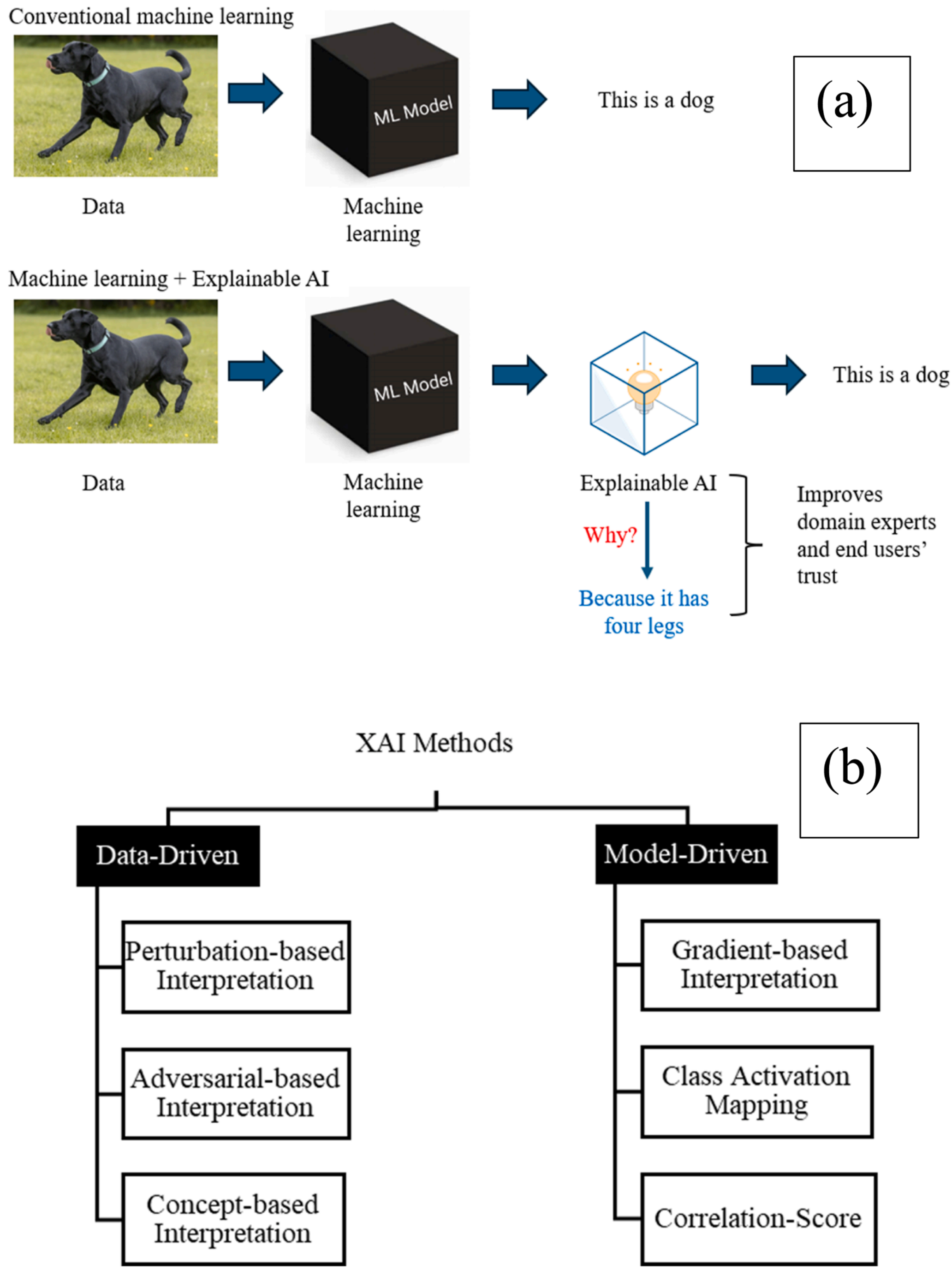

Fig. 4. XAI in machine learning research: (a) What is XAI and (b) Categories of XAI method, Source: Ukwaththa et al. [63].

conditions [111]. For example, an ML model can optimise the aeration process in wastewater treatment by predicting the minimum aeration needed to maintain DO levels, thus saving energy [112].

Use of ML models can ensure consistent water quality by predicting the outcomes of treatment processes [87]. These models analyse data from various stages of the treatment process to ensure that the final water quality meets regulatory standards.

2.6. Water resource management

ML supports decision-making in water resource management by providing accurate and timely information on water quality, helping to develop effective strategies for sustainable water management. They are integrated with data from various sources, including sensor networks, remote sensing and historical records, to provide a comprehensive view of water resources [110]. This integrated approach helps in understanding the interactions between different water bodies and the impact of various factors on water quality [113]. For instance, ML models can integrate data from rivers, lakes and groundwater to predict the overall water quality in a watershed [114].

Identifying pollution sources is crucial for effective water management. ML models can analyse spatial and temporal data to identify potential sources of pollution [115]. For example, clustering algorithms can identify hotspots of pollution in a river which can suggest nearby industrial or agricultural activities as potential sources. ML models can assist in the optimal allocation of resources for water management

Table 2
Comparison of general strengths and weaknesses of Machine Learning Models.

Model	Strengths	Weaknesses
Support Vector Machine	• Produces higher prediction accuracy and often shows higher generalisation. • Effective in high-dimensional spaces [87]	• Requires careful hyperparameter tuning (e.g. Kernel, C parameter) • Computationally intensive for large datasets [87]
Artificial Neural Network	• Can efficiently solve complex nonlinear problems [87]	• Model optimisation is difficult and time-consuming [88] • Requires large datasets for training. • Prone to overfitting without regularisation. • The black-box nature makes interpretation difficult.
Random Forest	• Achieves high accuracy • Model optimisation is relatively easy compared to a neural network • Bootstrap aggregating minimises overfitting [87]	• Tree-based predictions are not continuous. They provide the average score at the leaf node. This average can be slightly different from the actual value • Less interpretable than decision trees [87]
Decision Tree	• Achieves good accuracy • Model optimisation is relatively easy compared to a neural network • Model is interpretable at lower tree depths [89]	• Tree based predictions are not continuous. They provide the average score at the leaf node. This average can be slightly different from the actual value • Sensitive to small data variations.
Long Short-Term Memory	• Useful when predicting the future from the data in the past [87]	• Requires large amounts of data [87] • Computationally expensive [87]
Extreme gradient boosting	• High accuracy and fast execution [17] • Model optimisation is relatively easy compared to a neural network [88]	• Risk of overfitting to smaller datasets [90] • Tree based predictions are not continuous. They provide the average score at the leaf node. This average can be slightly different from the actual value [88]
k-Nearest Neighbors	• Easy implementation [87]	• Limited scalability for large datasets [91] • Computationally expensive for large datasets • For large datasets, accuracy can be lower [87]

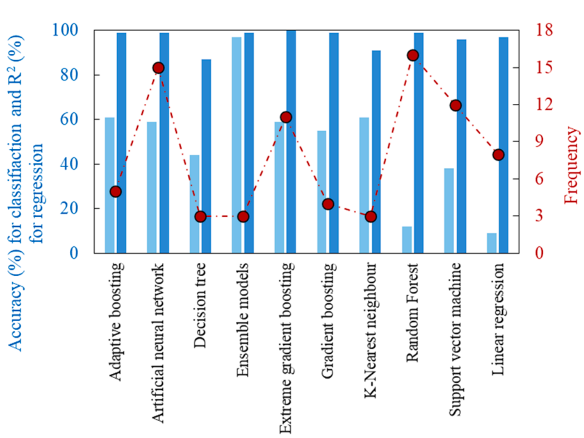

Fig. 5. Summary of the accuracies and frequency of usage of ML in recent water quality research.

[116]. These models analyse data on water demand, supply and quality to suggest the most efficient allocation of water resources. For instance, an ML model can predict water demand in different regions and suggest allocation strategies to ensure adequate supply while maintaining water quality [117]. ML provides valuable insights for developing water management policies [118]. By analysing data on water quality trends, pollution sources and the impact of various factors, ML models can inform policymakers about effective regulatory measures [119]. For example, ML models can suggest the most effective pollution control measures based on the analysis of historical pollution events and their causes [120]. Machine learning has revolutionised water quality evaluation by providing advanced tools for monitoring, predicting and managing water resources [121].

3. Internet of things in water quality monitoring

Similarly to machine learning, the scientific community has developed advanced techniques for real-time monitoring and assessment of surface and groundwater quality. This is driven by the internet of things (IoT), which integrates modern digital technologies such as wireless communication and cloud computing. IoT applications are now often used in many fields, including transportation, healthcare, smart homes and notably, water quality assessment [21]. IoT involves the interconnection of physical devices through sensors, networks and data analytics, allowing for real-time monitoring and control of various systems [122]. However, its practical implementation requires a detailed understanding of sensor architecture, communication protocols, data storage, and deployment challenges. IoT-based water monitoring systems typically rely on a network of low-power sensors that measure key parameters such as pH, dissolved oxygen (DO), turbidity, conductivity, and heavy metals. These sensors vary in power consumption, cost, and efficiency, making their selection critical for large-scale deployment. The generalised framework of IoT based system is shown in Fig. 7. It illustrates the simplified workflow of an IoT-based system: sensors embedded in the environment collect data, which is transmitted via IoT devices to a cloud server for storage and analysis. By automating data acquisition and reducing reliance on manual sampling, this framework enhances scalability, accuracy, and responsiveness to pollution events. Additionally, nowadays, the ML methods are combined with the data monitored in real time using IoT. That approach helps not only to observe real-time data but also to analsyse the data which helps real-time decisions.

In water quality monitoring, IoT devices can collect continuous data on critical parameters, transmit the data wirelessly and enable remote access for analysis and decision-making [123]. IoT's ability to provide 24/7 monitoring through an automated system is effective in water quality management. By detecting anomalies and pollution events in real time, IoT can facilitate rapid responses, reducing the time lag between contamination detection and corrective actions. This section aims to explore the various IoT applications, their technological components and their impact on the future of water quality assessment.

Moreover, IoT-enabled AI for water quality monitoring is crucial for sustainable development. Ensuring clean water is a key aspect of the sixth sustainable development goal (SDG) [124]. Without effective water quality monitoring, it is challenging to determine which water sources are safe for drinking. IoT-enabled AI can quickly detect and address water pollution from both specific and non-pointed sources [125]. This technology also supports the 14th SDG, which focuses on protecting aquatic life, by maintaining water quality within safe limits for the survival of aquatic plants and animals.

3.1. *The role of IoT in water quality assessment*

IoT systems for water quality assessment generally consist of three main components: sensors, connectivity modules and a cloud or edge computing infrastructure.

- Sensors: Specialised IoT sensors measure parameters such as pH, turbidity, temperature and chemical composition.

Table 3
Comparison of different aspects of XAI methods already used in water quality research.

Aspect	Shapley Additive Explanations [92]	Partial Dependence Plots [93]	Permutation Feature Importance [94]	Accumulated Local Effects [95]
Data/model-driven	Data-driven	Data-driven	Data-driven	Data-driven
Model-agnostic	Applicable to any model	Applicable to any model	Applicable to any model	Applicable to any model
Computational expense	High, especially for large datasets and complex models	Moderate; requires re-evaluations over grids	Efficient; only needs multiple evaluations per feature	Efficient; fewer evaluations compared to PDP
Sample dependency	Not strongly dependent; works well across samples	Strong dependency; assumes marginal feature distributions	Strongly dependent; random permutations impact results	Dependent; better than PDP in handling correlated features
Feature interaction handling	Captures feature interactions naturally	Limited; assumes independent features	Poor; evaluates features individually	Partially captures interactions locally
Use cases	Interpret complex models (both global and local explanations)	Visualise the marginal effect of features on prediction	Rank feature importance for the whole model	Interpret the local effect of features
Limitations	High computational cost; complex to interpret for high-dimensional data	Assumes feature independence; Struggles with correlated features	Sensitive to sampling variance and feature correlations	Difficult to interpret for categorical variables

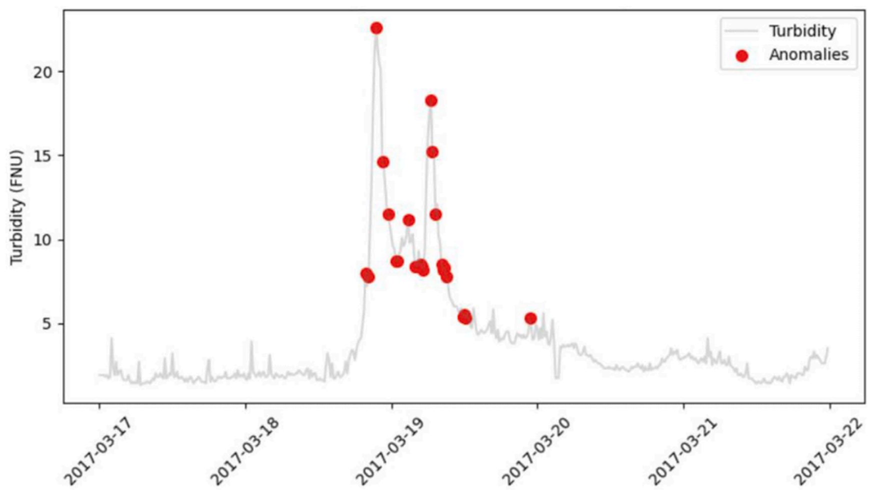

Fig. 6. Anomalies detection using hybrid deep learning, Source: Tharayil et al. [107].

- Connectivity: Wireless technologies such as Wi-Fi, LoRa, Zigbee, or cellular networks transmit the sensor data to a centralised server or cloud.
- Data Analytics: Cloud-based platforms or edge devices process the data using machine learning or artificial intelligence algorithms, generating insights and enabling predictive analytics.

Furthermore, IoT allows for broader geographic coverage, even in remote areas where manual sampling is impractical. Once collected, the data from IoT sensors can be stored in cloud-based systems for further analysis [122].

3.1.1. Hardware architectures and communication strategies

Currently, a variety of technologies are available for designing and developing IoT systems [126–129]. Many open-source platforms like Arduino, Raspberry Pi, ESP8266 and BeagleBone facilitate IoT development. These platforms support various communication technologies, including short-range options like Bluetooth and Wi-Fi, as well as long-range methods like GPRS, UMTS, 3G/4 G and LoRa for efficient data transmission. Additionally, IoT platforms are compatible with multiple identification technologies, such as NFC and RFID.

An IoT cyber-physical system consists of three main elements: the microcontroller, sensor and communication units [130,131]. The microcontroller interacts with the sensors and facilitates data transmission, either through an integrated communication unit or by connecting to an external communication module. The sensing unit collects physical data and connects to the microcontroller through various interfaces like analog input, digital input and I^2C (Inter-Integrated Circuit). The communication unit handles data transmission using technologies that can be wireless (such as Wi-Fi) or wired (such as Ethernet).

The sensor unit collects data, which is then processed and transmitted to the Internet by the microcontroller. Online services handle the analysis, visualisation and storage of this data, often using more powerful backend processing units. Nowadays, many low-cost sensors with various interface options are available, supporting numerous microcontrollers, making them suitable for applications in water management [127].

Nowadays, the integration of IoT in water quality has revolutionised real-time data collection, analysis and detection and decision making. However, its practical implementation requires a detailed understanding of sensor architecture, communication protocols, data storage, and deployment challenges. IoT-based water monitoring systems typically rely on a network of low-power sensors that measure key parameters such as pH, dissolved Oxygen (DO), turbidity, conductivity, and heavy metals [21]. These sensors vary in power consumption, cost, and efficiency, making the selection critical for large-scale deployment.

Sensor architecture is also important in IoT-enabled water quality assessment. Commonly used sensors include electrochemical sensors for pH and DO, optical sensors for turbidity, and biosensors for detecting contaminants such as heavy metals and nitrates [132–134]. These sensors often operate on low-power microcontrollers, such as ESP8266, ESP32, or Arduino-based platforms, which facilitate wireless connectivity and energy efficiency. Power consumption is a key consideration, as many water monitoring stations are deployed in remote areas with limited access to electricity [135,136]. Low-power communication strategies, such as duty cycling and energy harvesting techniques, help extend the operational life of these sensors.

Various communication protocols are used to transmit data in IoT-

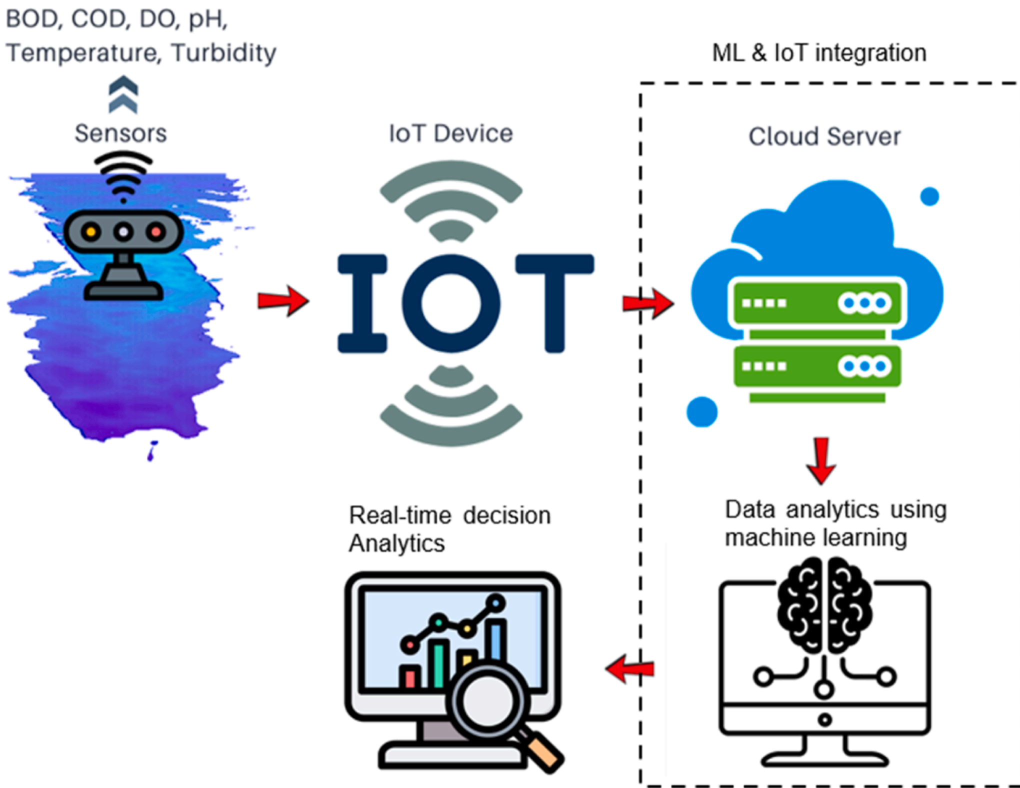

Fig. 7. Internet of things in water quality monitoring.

based water monitoring systems [128]. low-power wide-area networks (LPWANs), such as long range wide area networks (LoRaWAN) [137], are especially useful in remote or rural areas where energy efficiency is critical. In urban areas, more robust networks such as Wi-Fi, 4G, or 5G are often employed to support higher data transfer rates. Overall, these communication protocols have various transmission ranges and bandwidth as shown in Fig. 8. Short-range, high bandwidth technologies such as Wi-Fi and Bluetooth operate within tens of meters, providing high data throughput (Wi-Fi reaching up to 100 Mbps) but typically at the expense of higher power consumption. Zigbee provides a compromise with moderate data rates (approximately 100 Kbps) and slightly extended range compared to Bluetooth, while maintaining low power requirements. Radio-frequency identification (RFID) occupies the extreme low end of the spectrum, with extremely short range (~1 m) and minimal data rates (~1 Kbps), optimised for simple identification tasks rather than continuous monitoring.

At longer ranges, cellular technologies (including 3G, 4G/LTE, and 5G) provide high bandwidths across kilometers. However, their deployment may be limited by coverage availability, higher costs, and significant power demands, factors which are critical considerations for isolated environmental monitoring stations. Considering those, Low Power Wide Area Networks (LPWANs) offer promising alternatives. Licensed LPWAN technologies (e.g., Narrow band IoT (NB-IoT)) provide moderate bandwidths with operational ranges up to several kilometers. In contrast, unlicensed LPWAN technologies (e.g., LoRaWAN and Sigfox) achieve even longer ranges (exceeding 10 km) albeit at lower data rates (~1–100 Kbps). .

The choice of communication technology significantly impacts the efficiency and scalability of water quality monitoring systems. Various wireless communication protocols, such as LoRaWAN, ZigBee, Wi-Fi/

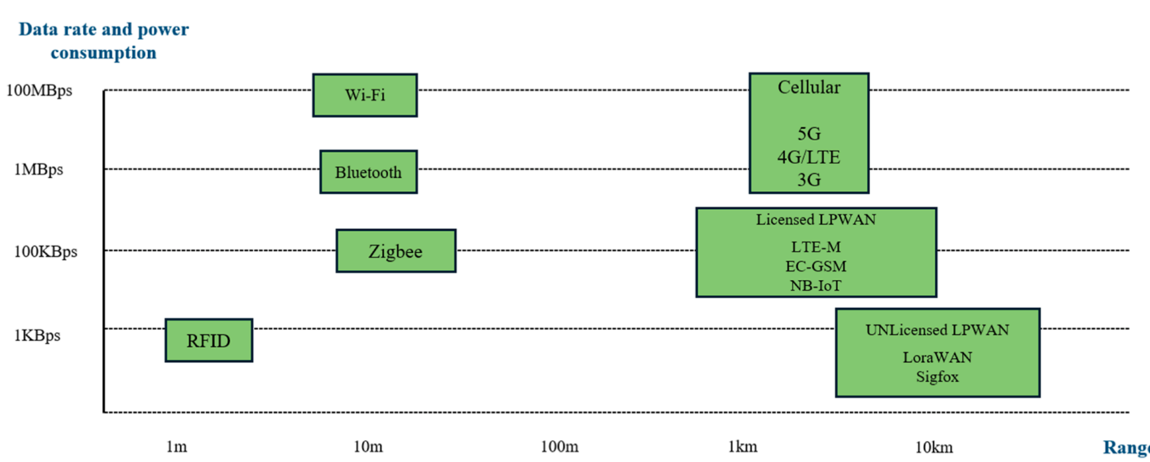

Fig. 8. IoT wireless protocols, Source: Essamlali et al., 2024 [138].

GSM, and NB-IoT, offer different trade-offs in terms of power consumption, range, data rate, cost, and suitability for different environments (Table 4).

In short-range deployments, bluetooth low energy (BLE) and Zigbee offer low-cost, low-power connectivity for localised monitoring [140, 141]. For large-scale networks, LoRaWAN and narrowband IoT (NB-IoT) enable long-range communication with minimal energy consumption, making them ideal for real-time water quality monitoring in remote regions. Cellular networks (e.g., 4G, 5G) and satellite communication are also employed in off-grid locations but come with higher costs and energy demands [142,143]. The selection of a communication protocol depends on factors such as data transmission frequency, power constraints, and environmental conditions.

The data storage and management aspect of IoT-driven water quality systems is equally important. Cloud-based platforms such as AWS IoT, Google Cloud IoT, and Microsoft Azure IoT offer real-time analytics, allowing environmental agencies and researchers to store, process, and visualise large datasets [144]. However, edge computing, which processes data locally at the sensor level before transmitting relevant insights, can reduce bandwidth usage, minimise latency, and enhance real-time decision-making [145]. This approach is particularly useful in applications requiring immediate responses to water quality fluctuations, such as early warning systems for contamination events.

3.1.2. Field deployment and challenges

Even while IoT technologies have a lot of potential to change water quality monitoring from reactive lab-based testing to proactive, real-time sensing, it is not always easy to convert this into practical, long-term implementations [146,147]. Sensor fouling is a significant issue that is commonly faced in the field, especially in natural water bodies. Sensors in lakes, rivers, and stormwater systems are frequently subjected to organic waste, biofilms, silt accumulation, and algae, all of which can quickly reduce sensor accuracy. For instance, in nutrient-rich waters, turbidity probes and pH sensors can coat in a matter of days, causing data drift that would not be detected until manual intervention takes place. This can counteract the noticeable benefit of "remote" monitoring by requiring periodic cleaning or recalibration, sometimes as often as weekly [148].

Data continuity is another challenge [148]. IoT devices provide round-the-clock (24/7) monitoring in theory, but in reality, data gaps are quite frequent. These could be caused by hardware failures, battery drain, or irregular power supplies (particularly in solar-powered remote systems). Data collection may also be interrupted by environmental variables such as intense rain, freezing temperatures, or physical damage from animals or debris [146,149]. Machine learning models, particularly those that depend on time-series consistency for anomaly detection or forecasting, can be negatively impacted by even brief disruptions.

Furthermore, network reliability becomes a major issue, especially in remote or wooded locations with spotty or poor cellular or satellite connectivity [150]. Even though protocols like NB-IoT and LoRaWAN are built for long-distance, low-power communication, they can malfunction because of geographical interference, access problems, or bad weather. One of the primary goals of IoT adoption is beaten when real-time monitoring switches back to delayed, batch processing in the absence of a steady network.

Human factors must also be considered. Environmental organisations or municipal agencies with little technical expertise run a large number of water quality monitoring activities [147]. It takes a certain amount of technical expertise to deploy and maintain IoT devices, especially when integrating them with cloud computing or machine learning operations. This expertise may not always be easily accessible. Even well-designed technologies run the danger of being underutilised in the absence of adequate training or intuitive user interfaces.

3.1.3. Ethical and governance issues of IoT based monitoring

Real-time data gathering, scalability, and predictive power are just a few of the impressive advantages that come with IoT technologies into water quality monitoring. However, there are also ethical and governance concerns that require more consideration. Concerns about data privacy, equity, openness, and accountability become more pressing as these technologies are used more and more in public, private, and even residential water systems.

In many applications, such as tracking water consumption in residential neighbourhoods, agricultural zones, or next to industrial facilities, data can be indirectly linked to identifiable behaviours, even though water quality data alone may appear innocuous [151]. Therefore, data privacy is one of the main issues. For example, usage patterns that represent daily habits, corporate processes, or regulatory infractions can be discovered through ongoing monitoring. Without stringent protections, there's a chance that third parties such as developers, insurers, or regulatory agencies will abuse data to target or penalise particular stakeholders or groups [152]. Strong privacy standards must be created as IoT systems grow in order to specify exactly who owns the data, who may access it, and under what circumstances it can be shared or examined [153].

Access equity is a major challenge as well. Many rural, isolated, or economically disadvantaged communities continue to rely on time-consuming, infrequent, and frequently less accurate manual sampling, even if well-funded cities and organisations may be able to afford the newest smart water infrastructure. There is a chance that the digital divide in environmental monitoring will exacerbate already-existing disparities in water safety and access. IoT-based solutions could unintentionally exacerbate environmental injustice in the absence of focused funding, capacity building, and inclusive design principles, particularly in areas where water insecurity is already a problem. Intentional policies that promote open-source tools, democratised access to monitoring technologies, and participatory monitoring frameworks that empower local people are necessary to address this [154,155].

3.2. Research progress on using IoT in water quality monitoring

Studies comparing IoT-based monitoring systems with traditional water quality assessment methods show that IoT systems offer superior temporal resolution, detecting issues like sudden pollution events more rapidly. Additionally, IoT systems can scale up more easily, providing continuous monitoring at a fraction of the operational cost of traditional methods.

Singh and Walingo [156] implemented a smart water quality monitoring system utilising IoT and wireless sensor networks (WSN).

Table 4
Comparison of Communication Technologies for Water Quality Monitoring Source [138,139].

Parameter	RFID	Bluetooth	LoRaWAN	ZigBee	Wi-Fi	5G	NB-IoT
Minimum Frequency (MHz)	0.125	2402	> 400	868	2500	< 800	> 0.003
Maximum Frequency (MHz)	433	2481	< 950	2400	5000	> 30,000	0.2
Data rate	Refer to Fig. 7						
Power consumption	Medium	High	Low	Low	Medium	High	Medium
Distance	Refer to Fig. 7						
Cost	High	Low	Low	Medium	Low	High	Low

The system included sensors to monitor parameters such as pH, conductivity and *Escherichia coli* concentration, combined with machine learning algorithms like Adaptive boosting to enhance prediction accuracy. Field data from water-treatment plants were used to assess performance. Despite challenges in real-time data accuracy, the study demonstrated improved monitoring efficiency, particularly in remote environments.

Wiryasaputra et al. [157] developed an IoT-based potable water quality monitoring and prediction model using cloud computing architecture. The system integrates sensors to measure water quality parameters such as pH, temperature, dissolved oxygen and electrical conductivity, transmitting the data via Narrow-Band-IoT technology to a cloud platform. Machine learning classifiers were applied to predict the potability of water, achieving higher accuracy with decision trees compared to other models. Alerts were sent to users via mobile notifications, enabling real-time monitoring and prediction of water quality.

Flores-Cortez [158] designed a low-cost IoT system to monitor water quality in developing countries, with a focus on aquaculture. The system comprised water sensors, a microcontroller and a Wi-Fi transceiver to monitor parameters such as temperature, pH and dissolved oxygen. Data were transmitted to a web-based dashboard for real-time analysis. Field testing in an aquaculture pond in El Salvador demonstrated the system's ability to provide stable, continuous monitoring, offering a promising solution for regions with limited water quality infrastructure.

Hemdan et al. [159] presented an IoT-based water quality monitoring system utilising machine learning algorithms for predictive analysis. The system measured parameters such as temperature, pH and conductivity and applied forecasting models, including long short-term memory networks and Facebook Prophet to predict water quality trends. The study highlighted the effectiveness of the Facebook Prophet model in terms of accuracy and resource utilisation, supporting decision-making in water management for smart cities. Table 5 provides a summary of empirical studies on IoT applications for water quality monitoring and assessment. It shows that IoT and IoT along with ML methods are increasingly being used in water quality monitoring research.

Due to the nature of current sensors, parameters like total dissolved solids, turbidity, electrical conductivity and pH are the most commonly studied indices. Developing portable and affordable sensors for detecting other parameters, such as heavy metals and other ions, would require a significant technological breakthrough [160]. **Future research** is likely to focus on alternative sensor technologies capable of measuring a broader range of parameters to accurately describe water quality. Achieving this would allow for the application of water quality index (WQI) correlations to obtain quick-WQI values, enabling the rapid determination of water source suitability for drinking.

Despite their many advantages, IoT-based systems come with their own set of challenges. One of the key issues is maintenance—sensors need regular calibration and upkeep, especially in harsh environmental conditions where exposure to contaminants or extreme weather can affect their accuracy. Security is another major concern, as IoT networks are susceptible to cyberattacks, making it crucial to implement strong encryption and security protocols. Additionally, the sheer volume of data generated by these sensors can be overwhelming. Without efficient data processing and storage solutions, managing and extracting meaningful insights from this flood of information becomes a significant hurdle. Overcoming these challenges is essential to fully employ the potential of IoT in water quality assessment.

Table 5
Recent IoT communication protocols and associated parameters in water quality monitoring systems.

Protocol	Parameters monitored	Machine learning models	Ref.
Wi-Fi	pH, Total Dissolved Solids, Turbidity	Random forest, Support vector machines, Linear regression, Naive Bayes, Decision trees	[161]
LoRaWAN/ GSM	pH, Temperature, Turbidity	–	[162]
LoRaWAN,5G	Temperature, humidity and pH	Linear regression, Stochastic gradient descent, Ridge regression	[163]
Sigfox	pH, Conductivity, Oxygenation and Temperature	Long Short-Term Memory Network	[164]
LoRaWAN	Temperature, Turbidity and relative humidity	–	[165]
Bluetooth, GSM, Zigbee, Wi-Fi	Turbidity, Total Dissolved Solids, Conductivity, BOD, nitrate and fecal coliform, pH	Decision tree	[166]
GSM	Temperature, pH and Turbidity	–	[167]
Wi-Fi	Temperature, pH and Dissolved oxygen	–	[168]
4G-LTE	pH, Dissolved oxygen, Turbidity and Conductivity.	–	[169]
Sigfox	pH and Turbidity	–	[170]
NB-IoT	Dissolved oxygen, pH, Temperature, Turbidity and Salinity	–	[171]
LoRaWAN	Temperature, Dissolved oxygen, Oxidation–reduction potential and pH	–	[172]
Wi-Fi/ Bluetooth	Dissolved oxygen, Oxidation–reduction potential, Dissolved Solids, Temperature, Conductivity and pH	Polynomial Regression	[173]

4. Conclusion

This review explored the integration of machine learning (ML) and the internet of things (IoT) in water quality monitoring and assessment. Traditional water quality evaluation methods, reliant on manual sampling and laboratory analysis, often fail to provide real-time insights which limits their effectiveness in addressing water pollution challenges. ML has emerged as a powerful tool for analysing complex water quality datasets, identifying patterns, predicting future trends and detecting anomalies. Meanwhile, IoT facilitates continuous, remote monitoring through sensor networks, ensuring real-time data collection and transmission. Together, these technologies enhance the accuracy, efficiency and scalability and provides a novel technological approach for water quality monitoring.

This study highlights that ensemble ML models, such as random forest, extreme gradient boosting, support vector machines and neural networks have been frequently used in water quality predictions which achieve over 95 % (Based on accuracy % in classification or R^2 % in regression) accuracy in water quality predictions. However, explainable AI (XAI) techniques remain underutilised, appearing in only a few of ML studies, which limits the transparency and trustworthiness of predictions.

IoT-based monitoring has significantly reduced contamination detection time, showing its potential for early warning systems in water resource management. IoT technologies hold great potential for revolutionising water quality monitoring, yet challenges such as sensor fouling, data continuity, network reliability, and human factors hinder their widespread adoption. Sensor fouling, particularly in natural water bodies, degrades sensor performance due to organic waste, biofilms, and algae, requiring frequent recalibration, which undermines the benefits of remote monitoring. Data gaps, caused by hardware failures, power issues, or environmental disruptions, further impact machine learning models reliant on time-series data. Network reliability is also a concern, especially in remote areas with poor connectivity, where IoT systems may revert to batch processing instead of real-time monitoring. Human factors, such as the lack of technical expertise in deploying and maintaining IoT devices, exacerbate the problem.

Additionally, ethical and governance issues, including data privacy, access equity, and environmental justice, need to be addressed as IoT technologies become more prevalent. Many disadvantaged communities still rely on manual monitoring methods, and the digital divide could worsen existing disparities in water safety.

The findings of this study have direct implications for improving real-world water quality monitoring systems. IoT-enabled ML models can enhance real-time pollution tracking, allowing authorities to detect contamination events significantly faster than traditional methods. The integration of AI-driven anomaly detection in smart water grids can optimise wastewater treatment processes, reducing chemical overuse and improving water reuse efficiency. Additionally, agriculture and industrial sectors can utilise ML-IoT to minimise runoff pollution, predict seasonal water quality changes, and ensure regulatory compliance.

5. Future research directions

To improve real-time performance, reliability and easy applicability, future research in the integration of machine learning and Internet of Things technologies for water quality monitoring should prioritise the following areas.

- Security and data integrity remain key concerns in distributed IoT networks. Blockchain can offer a promising solution for tamper-proof data logging and secure sensor-to-cloud communication [174]. The development of blockchain-enabled smart contracts could automate regulatory compliance tracking and water quality reporting.
- Shifting from centralised cloud computing to localised edge computing can significantly reduce latency and bandwidth usage. Future research should explore energy-efficient, AI-enhanced edge architectures that maintain predictive performance while enabling faster response in remote monitoring scenarios.
- There is a lack of comprehensive studies evaluating the cost and performance of ML-IoT systems against traditional water monitoring methods. Future research should benchmark these systems using real-world datasets across diverse environmental settings to quantify their economic and operational advantages, including long-term scalability and maintenance costs.
- Current IoT deployments often suffer from fragmented hardware and data standards. Future work should develop open, interpretable frameworks that ensure consistent data quality, support cross-regional monitoring, and enable smooth integration of ML algorithms across platforms.
- Adoption of XAI is essential to enhance trust, particularly in regulated sectors. Research should focus on integrating model interpretation tools such as SHAP into real-time water quality decision-making frameworks. Additionally, cross-comparison of XAI should be studied in order to investigate the differences/similarities in various XAI methods on how they explain the water quality data from a unique source.

CRediT authorship contribution statement

Gangani Dharmarathne: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **A.M.S.R. Abekoon:** Visualization, Investigation, Data curation. **Madhusha Bogahawaththa:** Writing – review & editing, Writing – original draft, Resources, Project administration, Conceptualization. **Janaka Alawatugoda:** Writing – review & editing, Project administration, Funding acquisition. **D.P.P. Meddage:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] A. Mohammadpour, et al., Assessment of drinking water quality and identifying pollution sources in a chromite mining region, J. Hazard. Mater. 480 (2024) 136050.
[2] C.K. Swain, Environmental pollution indices: a review on concentration of heavy metals in air, water, and soil near industrialization and urbanisation, Discover Environment 2 (1) (2024) 5.
[3] T.A. Kurniawan, et al., Implications of climate change on water quality and sanitation in climate hotspot locations: a case study in Indonesia, Water Supply 24 (2) (2024) 517–542.
[4] G.I. Edo, et al., Impact of environmental pollution from human activities on water, air quality and climate change, Ecological Frontiers (2024).
[5] S. Cook, S. Abolfathi, N.I. Gilbert, Goals and approaches in the use of citizen science for exploring plastic pollution in freshwater ecosystems: a review, Freshwater Science 40 (4) (2021) 567–579.
[6] D. Chen, et al., Global patterns of lake microplastic pollution: Insights from regional human development levels, Science of The Total Environment (2024) 176620.
[7] H. Tian, et al., Biodegradation of microplastics derived from controlled release fertilizer coating: Selective microbial colonization and metabolism in plastisphere, Science of The Total Environment 920 (2024) 170978.
[8] M. Guo, R. Noori, S. Abolfathi, Microplastics in freshwater systems: Dynamic behaviour and transport processes, Resources, Conservation and Recycling 205 (2024) 107578.
[9] M. Anwar, S.A. Shirazi, U. Mazhar, Spatial Distribution and Health Risk Assessment of Groundwater Pollution in Kotlakhpat Industrial Complex, Lahore, Water, Air, & Soil Pollution 235 (2) (2024) 157.
[10] U. Water, Progress on change in water-use efficiency: Global status and acceleration needs for SDG indicator 6.4. 1,2021, Food & Agriculture Org, 2021.
[11] M.G. Uddin, S. Nash, A.I. Olbert, A review of water quality index models and their use for assessing surface water quality, Ecol. Indic. 122 (2021) 107218.
[12] R. Singh, et al., Water quality monitoring and management of building water tank using industrial internet of things, Sustainability. 13 (15) (2021) 8452.
[13] F. Lezzar, D. Benmerzoug, and I. Kitouni, "IoT for monitoring and control of water quality parameters," 2020.
[14] J. Sarafaraz, F.A. Kaleybar, J.M. Karamjavan, N. Habibzadeh, Predicting river water quality: an imposing engagement between machine learning and the QUAL2Kw models (case study: aji-Chai, river, Iran), Results. Eng. 21 (2024) 101921.
[15] M. Yousefi, V. Oskoei, H.R. Esmaeli, M. Baziar, An innovative combination of extra trees within adaboost for accurate prediction of agricultural water quality indices, Results. Eng. 24 (2024) 103534.
[16] L. Kandasamy, A. Mahendran, S.H.V. Sangaraju, P. Mathur, S.V. Faldu, M. Mazzara, Enhanced remote sensing and deep learning aided water quality detection in the Ganges River, India supporting monitoring of aquatic environments, Results. Eng. 25 (2025) 103604.
[17] M. Zhu, et al., A review of the application of machine learning in water quality evaluation, Eco-Environment & Health 1 (2) (2022) 107–116.
[18] K. Khosravi, et al., Enhanced water quality prediction model using advanced hybridized resampling alternating tree-based and deep learning algorithms, Environmental Science and Pollution Research (2025) 1–20.
[19] B. Ghiasi, et al., Uncertainty quantification of granular computing-neural network model for prediction of pollutant longitudinal dispersion coefficient in aquatic streams, Sci. Rep. 12 (1) (2022) 4610.
[20] R. Noori, et al., A non-threshold model to estimate carcinogenic risk of nitrate-nitrite in drinking water, J. Clean. Prod. 363 (2022) 132432.
[21] F. Jan, N. Min-Allah, D. Düştegör, Iot based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications, Water. (Basel) 13 (13) (2021) 1729.
[22] T. Paepae, P.N. Bokoro, K. Kyamakya, From fully physical to virtual sensing for water quality assessment: a comprehensive review of the relevant state-of-the-art, Sensors 21 (21) (2021) 6971.
[23] M.C. Onojake, C. Obi, A.E.D. Mahmoud, Water Quality Monitoring Using Sensors and Models. Artificial Intelligence and Modeling for Water Sustainability, CRC Press, 2023, pp. 97–127.
[24] S.N. Zainurin, et al., Advancements in monitoring water quality based on various sensing methods: a systematic review, Int. J. Environ. Res. Public Health 19 (21) (2022) 14080.
[25] M.M. Hason, I.S. Abbood, S. aldeen Odaa, Land cover reflectance of Iraqi marshlands based on visible spectral multiband of satellite imagery, Results. Eng. 8 (2020) 100167.

[26] H. Mahir Mahmod, I.S. Abbood, A.N. Hanoon, SURFACE AREA EVALUATION OF MOSUL DOM LAKE USING SATELLITE IMAGERY TECHNIQUE, Technology. 6 (1) (2020) 85–100.
[27] M.H. Gholizadeh, A.M. Melesse, L. Reddi, A comprehensive review on water quality parameters estimation using remote sensing techniques, Sensors 16 (8) (2016) 1298.
[28] E.U. Alum, et al., Physico-chemical and bacteriological analysis of water used for drinking and other domestic purposes in Amaozara Ozizza, Afikpo North, Ebonyi State, Nigeria, Nigerian Journal of Biochemistry and Molecular Biology 38 (1) (2023) 1–8.
[29] S. Heydari, M.R. Nikoo, A. Mohammadi, R. Barzegar, Two-stage meta-ensembling machine learning model for enhanced water quality forecasting, J. Hydrol. (Amst) 641 (2024) 131767.
[30] C. Shen, A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water. Resour. Res. 54 (11) (2018) 8558–8593.
[31] S.J. Mohammed, S.L. Zubaidi, S. Ortega-Martorell, N. Al-Ansari, S. Ethaib, K. Hashim, Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective, Cogent. Eng. 9 (1) (2022) 2143051.
[32] N. Nasir, et al., Water quality classification using machine learning algorithms, J. Water. Process. Eng. 48 (2022) 102920.
[33] A. Fernández del Castillo, C. Yebra-Montes, M. Verduzco Garibay, J. de Anda, A. Garcia-Gonzalez, M.S. Gradilla-Hernández, Simple prediction of an ecosystem-specific water quality index and the water quality classification of a highly polluted river through supervised machine learning, Water. (Basel) 14 (8) (2022) 1235.
[34] M.Y. Shams, A.M. Elshewey, E.S.M. El-Kenawy, A. Ibrahim, F.M. Talaat, Z. Tarek, Water quality prediction using machine learning models based on grid search method, Multimed. Tools. Appl. 83 (12) (2024) 35307–35334.
[35] G. Shenbagalakshmi, A. Shenbagarajan, S. Thavasi, M.G. Nayagam, R. Venkatesh, Determination of water quality indicator using deep hierarchical cluster analysis, Urban. Clim. 49 (2023) 101468.
[36] A. Mazher, Visualization framework for high-dimensional spatio-temporal hydrological gridded datasets using machine-learning techniques, Water. (Basel) 12 (2) (2020) 590.
[37] Á.F. Gambín, E. Angelats, J.S. González, M. Miozzo, P. Dini, Sustainable marine ecosystems: Deep learning for water quality assessment and forecasting, IEEe Access. 9 (2021) 121344–121365.
[38] H. Wan, R. Xu, M. Zhang, Y. Cai, J. Li, X. Shen, A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods, J. Hydrol. (Amst) 612 (2022) 128081.
[39] U.M. Ismail, K. Bani-Melhem, M.F. Khan, H. Elnakar, Optimizing membrane bioreactor performance in wastewater treatment using machine learning and meta-heuristic techniques, Results. Eng. 25 (2025) 104626.
[40] S. Alnaimat, O. Mohsen, H. Elnakar, Perfluorooctanoic Acids (PFOA) removal using electrochemical oxidation: a machine learning approach, J. Environ. Manage 370 (2024) 122857.
[41] U. Ewuzie, O.P. Bolade, A.O. Egbedina, Application of deep learning and machine learning methods in water quality modeling and prediction: a review, Current trends and advances in computer-aided intelligent environmental data engineering (2022) 185–218.
[42] A.R. Jadhav, P.D. Pathak, R.Y. Raut, Water and wastewater quality prediction: current trends and challenges in the implementation of artificial neural network, Environ. Monit. Assess. 195 (2) (2023) 321.
[43] J. Wu, Z. Wang, A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory, Water. (Basel) 14 (4) (2022) 610.
[44] Y. Deng, X. Ye, X. Du, Predictive modeling and analysis of key drivers of groundwater nitrate pollution based on machine learning, J. Hydrol. (Amst) 624 (2023) 129934.
[45] Y. Fu, Z. Hu, Y. Zhao, M. Huang, A long-term water quality prediction method based on the temporal convolutional network in smart mariculture, Water. (Basel) 13 (20) (2021) 2907.
[46] K.P. Wai, M.Y. Chia, C.H. Koo, Y.F. Huang, W.C. Chong, Applications of deep learning in water quality management: a state-of-the-art review, J. Hydrol. (Amst) 613 (2022) 128332.
[47] G.A. López-Ramírez, A. Aragón-Zavala, Wireless sensor networks for water quality monitoring: a comprehensive review, IEEe Access. 11 (2023) 95120–95142.
[48] T.M. Tung, Z.M. Yaseen, A survey on river water quality modelling using artificial intelligence models: 2000–2020, J. Hydrol. (Amst) 585 (2020) 124670.
[49] N.J. Messina, R.M. Couture, S.A. Norton, S.D. Birkel, A. Amirbahman, Modeling response of water quality parameters to land-use and climate change in a temperate, mesotrophic lake, Science of the Total Environment 713 (2020) 136549.
[50] C. Hesse, V. Krysanova, Modeling climate and management change impacts on water quality and in-stream processes in the Elbe River Basin, Water. (Basel) 8 (2) (2016) 40.
[51] W. Zhang, Hybrid Meteorological Forecasting: ML-Driven Predictions of Lake Michigan's Lake-Effect precipitation for Urban Preparedness, Northeastern University, 2023.
[52] H. Lu, X. Ma, Hybrid decision tree-based machine learning models for short-term water quality prediction, Chemosphere 249 (2020) 126169.
[53] C. Gakii and J. Jepkoech, "A classification model for water quality analysis using decision tree," 2019.
[54] M. Bourel, A. Segura, Multiclass classification methods in ecology, Ecol. Indic. 85 (2018) 1012–1021.
[55] S. Liu, L. Xu, Q. Li, X. Zhao, D. Li, Fault diagnosis of water quality monitoring devices based on multiclass support vector machines and rule-based decision trees, IEEe Access. 6 (2018) 22184–22195.
[56] K.A. Thompson, et al., Detecting industrial discharges at an advanced water reuse facility using online instrumentation and supervised machine learning binary classification, Front. Water. 4 (2022) 1014556.
[57] K.A. Thompson, E.R. Dickenson, Using machine learning classification to detect simulated increases of de facto reuse and urban stormwater surges in surface water, Water. Res. 204 (2021) 117556.
[58] N.H.A. Malek, W.F. Wan Yaacob, S.A. Md Nasir, N. Shaadan, Prediction of water quality classification of the Kelantan River Basin, Malaysia, using machine learning techniques, Water. (Basel) 14 (7) (2022) 1067.
[59] E. Dritsas, M. Trigka, Efficient data-driven machine learning models for water quality prediction, Computation 11 (2) (2023) 16.
[60] M. Al Saleem, F. Harrou, Y. Sun, Explainable machine learning methods for predicting water treatment plant features under varying weather conditions, Results. Eng. 21 (2024) 101930.
[61] J. Park, W.H. Lee, K.T. Kim, C.Y. Park, S. Lee, T.Y. Heo, Interpretation of ensemble learning to predict water quality using explainable artificial intelligence, Science of the Total Environment 832 (2022) 155070.
[62] M. Nallakaruppan, E. Gangadevi, M.L. Shri, B. Balusamy, S. Bhattacharya, S. Selvarajan, Reliable water quality prediction and parametric analysis using explainable AI models, Sci. Rep. 14 (1) (2024) 7520.
[63] J. Ukwaththa, S. Herath, D. Meddage, A review of machine learning (ML) and explainable artificial intelligence (XAI) methods in additive manufacturing (3D Printing), Mater. Today Commun. (2024) 110294.
[64] R.K. Makumbura, et al., Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature, Results. Eng. 23 (2024) 102831.
[65] T. Yan, A. Zhou, S.L. Shen, Prediction of long-term water quality using machine learning enhanced by Bayesian optimisation, Environmental Pollution 318 (2023) 120870.
[66] P. Rawat, M. Bajaj, V. Sharma, S. Vats, 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), IEEE, 2023, pp. 1108–1114.
[67] M.G. Uddin, S. Nash, A. Rahman, A.I. Olbert, Performance analysis of the water quality index model for predicting water state using machine learning techniques, Process Safety and Environmental Protection 169 (2023) 808–828.
[68] S. Tian, et al., Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms, Environmental Science and Pollution Research 30 (7) (2023) 18617–18630.
[69] A. Gaagai, et al., Application of water quality indices, machine learning approaches, and GIS to identify groundwater quality for irrigation purposes: a case study of Sahara Aquifer, Doucen Plain, Algeria, Water. (Basel) 15 (2) (2023) 289.
[70] U. Mohseni, C.B. Pande, S.C. Pal, F. Alshehri, Prediction of weighted arithmetic water quality index for urban water quality using ensemble machine learning model, Chemosphere 352 (2024) 141393.
[71] S. Wang, H. Peng, S. Liang, Prediction of estuarine water quality using interpretable machine learning approach, J. Hydrol. (Amst) 605 (2022) 127320.
[72] A. Aldrees, M.F. Javed, A.T.B. Taha, A.M. Mohamed, M. Jasiński, M. Gono, Evolutionary and ensemble machine learning predictive models for evaluation of water quality, Journal of Hydrology: Regional Studies 46 (2023) 101331.
[73] H. Ibrahim, et al., Evaluation and prediction of groundwater quality for irrigation using an integrated water quality indices, machine learning models and GIS approaches: a representative case study, Water. (Basel) 15 (4) (2023) 694.
[74] B.Q. Lap, et al., Predicting Water Quality Index (WQI) by feature selection and machine learning: a case study of An Kim Hai irrigation system, Ecol. Inform. 74 (2023) 101991.
[75] A.M. Sajib, M.T.M. Diganta, A. Rahman, T. Dabrowski, A.I. Olbert, M.G. Uddin, Developing a novel tool for assessing the groundwater incorporating water quality index and machine learning approach, Groundw. Sustain. Dev. 23 (2023) 101049.
[76] M.E. Omeka, Evaluation and prediction of irrigation water quality of an agricultural district, SE Nigeria: an integrated heuristic GIS-based and machine learning approach, Environmental Science and Pollution Research 31 (41) (2024) 54178–54203.
[77] A. Aldrees, M. Khan, A.T.B. Taha, M. Ali, Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches, J. Water. Process. Eng. 58 (2024) 104789.
[78] H. Jeong, et al., Machine learning-based water quality prediction using octennial in-situ Daphnia magna biological early warning system data, J. Hazard. Mater. 465 (2024) 133196.
[79] S. Talukdar, et al., Optimisation and interpretation of machine and deep learning models for improved water quality management in Lake Loktak, J. Environ. Manage 351 (2024) 119866.
[80] S. Majnooni, et al., Smarter water quality monitoring in reservoirs using interpretable deep learning models and feature importance analysis, J. Water. Process. Eng. 60 (2024) 105187.
[81] H. Li, et al., The insightful water quality analysis and predictive model establishment via machine learning in dual-source drinking water distribution system, Environ. Res. 250 (2024) 118474.

[82] A. Saeed, A. Alsini, D. Amin, Water quality multivariate forecasting using deep learning in a West Australian estuary, Environmental Modelling & Software 171 (2024) 105884.
[83] R. Wang, J.H. Kim, M.H. Li, Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach, Science of the Total Environment 761 (2021) 144057.
[84] Z. Zhang, J. Huang, S. Duan, Y. Huang, J. Cai, J. Bian, Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed, Ecol. Indic. 140 (2022) 108977.
[85] S.S. Hasani, M.E. Arias, H.Q. Nguyen, O.M. Tarabih, Z. Welch, Q. Zhang, Leveraging explainable machine learning for enhanced management of lake water quality, J. Environ. Manage 370 (2024) 122890.
[86] K. Merabet, et al., Predicting water quality variables using gradient boosting machine: global versus local explainability using SHapley Additive Explanations (SHAP), Earth. Sci. Inform. 18 (3) (2025) 1–34.
[87] M. Lowe, R. Qin, X. Mao, A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring, Water. (Basel) 14 (9) (2022) 1384.
[88] D. Meddage, D. Mohotti, K. Wijesooriya, C. Lee, K. Kwok, Interpolating wind pressure time-histories around a tall building-A deep learning-based approach, Journal of Wind Engineering and Industrial Aerodynamics 256 (2025) 105968.
[89] Y. Luo, H.H. Tseng, S. Cui, L. Wei, R.K. Ten Haken, I. El Naqa, Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling, BJR| Open 1 (1) (2019) 20190021.
[90] N. Malik, A. Kalonia, S. Dalal, D.N. Le, Optimized XGBoost Hyper-Parameter Tuned Model with Krill Herd Algorithm (KHA) for Accurate Drinking Water Quality Prediction, SN. Comput. Sci. 6 (3) (2025) 263.
[91] Y. Wang, Research on Spam Filters using: SVM, Naïve Bayes, and KNN, in: 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI2023), Atlantis Press, 2023, pp. 574–580.
[92] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[93] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[94] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[95] D.W. Apley, J. Zhu, Visualizing the effects of predictor variables in black box supervised learning models, Journal of the Royal Statistical Society Series B: Statistical Methodology 82 (4) (2020) 1059–1086.
[96] K. Choi, J. Yi, C. Park, S. Yoon, Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines, IEEe Access. 9 (2021) 120043–120065.
[97] G. Apostolakos, "Operational Anomaly Detection Using Clustering Methods and Machine Learning Models," 2024.
[98] D. Ramotsoela, A. Abu-Mahfouz, G. Hancke, A survey of anomaly detection in industrial wireless sensor networks with critical water system infrastructure as a case study, Sensors 18 (8) (2018) 2491.
[99] J. Pyo, et al., Long short-term memory models of water quality in inland water environments, Water. Res. X. (2023) 100207.
[100] E.M. Dogo, N.I. Nwulu, B. Twala, C. Aigbavboa, A survey of machine learning methods applied to anomaly detection on drinking-water quality data, Urban. Water. J. 16 (3) (2019) 235–248.
[101] E. El-Shafeiy, M. Alsabaan, M.I. Ibrahem, H. Elwahsh, Real-time anomaly detection for water quality sensor monitoring based on multivariate deep learning technique, Sensors 23 (20) (2023) 8613.
[102] B. Dhamodharan, Beyond Traditional Methods: a Novel Approach to Anomaly Detection and Classification Using AI Techniques, Transactions on Latest Trends in Artificial Intelligence 3 (3) (2022).
[103] P. Xue, et al., Integrating Deep Learning and Hydrodynamic Modeling to Improve the Great Lakes Forecast, Remote Sens. (Basel) 14 (11) (2022) 2640.
[104] Y. Wang, X. Gao, B. Sun, Y. Liu, Developing a 3D hydrodynamic and water quality model for floating treatment wetlands to study the flow structure and nutrient removal performance of different configurations, Sustainability. 14 (12) (2022) 7495.
[105] U. Ahmed, R. Mumtaz, H. Anwar, S. Mumtaz, A.M. Qamar, Water quality monitoring: from conventional to emerging technologies, Water Supply 20 (1) (2020) 28–45.
[106] D. Jalal, T. Ezzedine, Toward a smart real time monitoring system for drinking water based on machine learning, in: 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), IEEE, 2019, pp. 1–5.
[107] S.M. Tharayil, N.K. Alomari, D.K. Bubshait, Detecting Anomalies in Water Quality Monitoring Using Deep Learning, in: SPE Water Lifecycle Management Conference and Exhibition, OnePetro, 2024.
[108] S.Mohamed Almazrouei, F. Dweiri, R. Aydin, A. Alnaqbi, A review on the advancements and challenges of artificial intelligence based models for predictive maintenance of water injection pumps in the oil and gas industry, SN. Appl. Sci. 5 (12) (2023) 391.
[109] I. Rojek, M. Jasiulewicz-Kaczmarek, M. Piechowski, D. Mikołajewski, An artificial intelligence approach for improving maintenance to supervise machine failures and support their repair, Applied Sciences 13 (8) (2023) 4971.
[110] J. Jawad, A.H. Hawari, S.J. Zaidi, Artificial neural network modeling of wastewater treatment and desalination using membrane processes: a review, Chemical Engineering Journal 419 (2021) 129540.
[111] C. Varadharajan, et al., Can machine learning accelerate process understanding and decision-relevant predictions of river water quality? Hydrol. Process. 36 (4) (2022) e14565.
[112] D.W. Dunnington, B.F. Trueman, W.J. Raseman, L.E. Anderson, G.A. Gagnon, Comparing the Predictive performance, interpretability, and accessibility of machine learning and physically based models for water treatment, ACS. ES. T. Eng. 1 (3) (2020) 348–356.
[113] A.Y. Sun, B.R. Scanlon, How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions, Environmental Research Letters 14 (7) (2019) 073001.
[114] M. Sit, B.Z. Demiray, Z. Xiang, G.J. Ewing, Y. Sermet, I. Demir, A comprehensive review of deep learning applications in hydrology and water resources, Water Science and Technology 82 (12) (2020) 2635–2670.
[115] X. Xiang, Q. Li, S. Khan, O.I. Khalaf, Urban water resource management for sustainable environment planning using artificial intelligence techniques, Environ. Impact. Assess. Rev. 86 (2021) 106515.
[116] S. Behmel, M. Damour, R. Ludwig, M. Rodriguez, Water quality monitoring strategies—A review and future perspectives, Science of the Total Environment 571 (2016) 1312–1329.
[117] C. Valerio, L. De Stefano, G. Martínez-Muñoz, A. Garrido, A machine learning model to assess the ecosystem response to water policy measures in the Tagus River Basin (Spain), Science of the Total Environment 750 (2021) 141252.
[118] L. Ho, P. Goethals, Machine learning applications in river research: Trends, opportunities and challenges, Methods Ecol. Evol. 13 (11) (2022) 2603–2621.
[119] G. Chhipi-Shrestha, H.R. Mian, S. Mohammadiun, M. Rodriguez, K. Hewage, R. Sadiq, Digital water: artificial intelligence and soft computing applications for drinking water quality assessment, Clean. Technol. Environ. Policy. 25 (5) (2023) 1409–1438.
[120] A.N. Ahmed, et al., Machine learning methods for better water quality prediction, J. Hydrol. (Amst) 578 (2019) 124084.
[121] F. Ghobadi, D. Kang, Application of machine learning in water resources management: a systematic literature review, Water. (Basel) 15 (4) (2023) 620.
[122] J.B. Ajith, R. Manimegalai, V. Ilayaraja, An IoT based smart water quality monitoring system using cloud, in: 2020 International conference on emerging trends in information technology and engineering (ic-ETITE), IEEE, 2020, pp. 1–7.
[123] Z. Hassan, G. Hossain, M.M. Islam, Internet of Things (IoT) based water quality monitoring system, Educ. Res. 2 (4) (2020) 168–180.
[124] M. Wang, A.B. Janssen, J. Bazin, M. Strokal, L. Ma, C. Kroeze, Accounting for interactions between Sustainable Development Goals is essential for water pollution control in China, Nat. Commun. 13 (1) (2022) 730.
[125] N. Thai-Nghe, N. Thanh-Hai, N. Chi Ngon, Deep learning approach for forecasting water quality in IoT systems, International Journal of Advanced Computer Science and Applications 11 (8) (2020) 686–693.
[126] S.C. Olisa, C.N. Asiegbu, J.E. Olisa, B.O. Ekengwu, A.A. Shittu, M.C. Eze, Smart two-tank water quality and level detection system via IoT, Heliyon. 7 (8) (2021).
[127] S. Pasika, S.T. Gandla, Smart water quality monitoring system with cost-effective using IoT, Heliyon. 6 (7) (2020).
[128] K.L. Tsai, L.W. Chen, L.J. Yang, H.J. Shiu, H.W. Chen, IoT based smart aquaculture system with automatic aerating and water quality monitoring, Journal of Internet Technology 23 (1) (2022) 177–184.
[129] S.K. Vasudevan, B. Baskaran, An improved real-time water quality monitoring embedded system with IoT on unmanned surface vehicle, Ecol. Inform. 65 (2021) 101421.
[130] Y. Wang, I.W.H. Ho, Y. Chen, Y. Wang, Y. Lin, Real-time water quality monitoring and estimation in AIoT for freshwater biodiversity conservation, IEEe Internet. Things. J. 9 (16) (2021) 14366–14374.
[131] H.M. Yasin, et al., IoT and ICT based smart water management, monitoring and controlling system: a review, Asian Journal of Research in Computer Science 8 (2) (2021) 42–56.
[132] H. Xiang, et al., Sensors applied for the detection of pesticides and heavy metals in freshwaters, J. Sens. 2020 (1) (2020) 8503491.
[133] P. Yeh, N. Yeh, C.H. Lee, T.J. Ding, Applications of LEDs in optical sensors and chemical sensing device for detection of biochemicals, heavy metals, and environmental nutrients, Renewable and Sustainable Energy Reviews 75 (2017) 461–468.
[134] K. Lal, S.A. Jaywant, K.M. Arif, Electrochemical and optical sensors for real-time detection of nitrate in water, Sensors 23 (16) (2023) 7099.
[135] A.H. Gatea, A. Al-Ibadi, and M. Alaziz, "Optimizing workplace performance by using ESP32 to monitor and enhance environmental quality in smart office solutions".
[136] A. Martikkala, A. Lobov, and I. Flores Ituarte, "Affordable Modular Iot Gateway for Iot-Sensor Data Collection," Available at SSRN 4559684.
[137] L.K. Tolentino, et al., IoT-based automated water monitoring and correcting modular device via LoRaWAN for aquaculture, International Journal of Computing and Digital Systems 10 (2021) 533–544.
[138] I. Essamlali, H. Nhaila, M. El Khaili, Advances in machine learning and IoT for water quality monitoring: a comprehensive review, Heliyon. (2024).
[139] H.N.S. Aldin, M.R. Ghods, F. Nayebipour, M.N. Torshiz, A comprehensive review of energy harvesting and routing strategies for IoT sensors sustainability and communication technology, Sens. Int. 5 (2024) 100258.
[140] M. Ouaissa, M. Ouaissa, I.U. Khan, Z. Boulouard, J. Rashid, Low-power wide Area network for large Scale internet of things: architectures, communication protocols and recent Trends, CRC Press, 2024.
[141] T. Polonelli, "Ultra-low power iot applications: from transducers to wireless protocols," 2021.
[142] S.T. Ahmed, "Implementation And Optimization Of A Secure, Scalable, And Robust Long-Range Low-Power Mesh Network For Enhanced Geo-Location Accuracy And Efficient Data Aggregation," 2024.

[143] M. Jouhari, N. Saeed, M.S. Alouini, E.M. Amhoud, A survey on scalable LoRaWAN for massive IoT: Recent advances, potentials, and challenges, IEEE Communications Surveys & Tutorials 25 (3) (2023) 1841–1876.
[144] S. Daniel, S. Brightwood, and J. Oluwaseyi, "Cloud-based big data analytics (aws, azure, google cloud)," 2024.
[145] K. Gatlin, "Real-Time Analytics on Amazon Web Services and Google Cloud: Unlocking Data-Driven Insights," 2024.
[146] A. Gaddam, T. Wilkin, M. Angelova, J. Gaddam, Detecting sensor faults, anomalies and outliers in the internet of things: a survey on the challenges and solutions, Electronics. (Basel) 9 (3) (2020) 511.
[147] D. Adhikari, et al., A comprehensive survey on imputation of missing data in internet of things, ACM. Comput. Surv. 55 (7) (2022) 1–38.
[148] A. Marengo, Navigating the nexus of AI and IoT: a comprehensive review of data analytics and privacy paradigms, Internet of Things (2024) 101318.
[149] S. Dutta, AI and IoT-Based Smart Healthcare Solutions in Urban Area, Trends in Health Informatics 2 (1) (2025) 18–26.
[150] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, B. Qureshi, An overview of IoT sensor data processing, fusion, and analysis techniques, Sensors 20 (21) (2020) 6076.
[151] C. Li, B. Palanisamy, Privacy in Internet of Things: From principles to technologies, IEEe Internet. Things. J. 6 (1) (2018) 488–505.
[152] K.R. Sollins, IoT big data security and privacy versus innovation, IEEe Internet. Things. J. 6 (2) (2019) 1628–1635.
[153] C. Maple, Security and privacy in the internet of things, Journal of cyber policy 2 (2) (2017) 155–184.
[154] S. Suklabaidya, "Towards inclusive societies: Leveraging IoT for community development and education," 2024.
[155] J. Shahid, R. Ahmad, A.K. Kiani, T. Ahmad, S. Saeed, A.M. Almuhaideb, Data protection and privacy of the internet of healthcare things (IoHTs), Applied Sciences 12 (4) (2022) 1927.
[156] Y. Singh, T. Walingo, Smart Water Quality Monitoring with IoT Wireless Sensor Networks, Sensors 24 (9) (2024) 2871 [Online]. Available: https://www.mdpi.com/1424-8220/24/9/2871.
[157] R. Wiryasaputra, C.Y. Huang, Y.J. Lin, and C.T. Yang, "An IoT Real-Time Potable Water Quality Monitoring and Prediction Model Based on Cloud Computing Architecture," Sensors, vol. 24, no. 4, doi: 10.3390/s24041180.
[158] O.O. Flores-Cortez, A Low-Cost IoT System for Water Quality Monitoring in Developing Countries, in: 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), 2024, pp. 596–597, https://doi.org/10.1109/CCNC51664.2024.10454847, 6-9 Jan. 2024.
[159] E.E.D. Hemdan, Y.M. Essa, M. Shouman, A. El-Sayed, A.N. Moustafa, An efficient IoT based smart water quality monitoring system, Multimed. Tools. Appl. 82 (19) (2023) 28827–28851, https://doi.org/10.1007/s11042-023-14504-z, 2023/08/01.
[160] M. Miller, A. Kisiel, D. Cembrowska-Lech, I. Durlik, T. Miller, IoT in water quality monitoring—Are we really here? Sensors 23 (2) (2023) 960.
[161] M.A. Rahu, et al., IoT and machine learning solutions for monitoring agricultural water quality: a robust framework, Mehran University Research Journal Of Engineering & Technology 43 (1) (2024) 192–205.
[162] Z. Mansor, N.N.S. Abdul Latiff, Water pollution detection system for illegal toxic waste dumps. Materials Innovations and Solutions in Science and Technology: With a Focus on Tropical Plant Biomaterials, Springer, 2023, pp. 73–81.
[163] Y. Li, X. Wang, Z. Zhao, S. Han, Z. Liu, Lagoon water quality monitoring based on digital image analysis and machine learning estimators, Water. Res. 172 (2020) 115471.
[164] P. Boccadoro, V. Daniele, P. Di Gennaro, D. Lofù, P. Tedeschi, Water quality prediction on a sigfox-compliant iot device: The road ahead of waters, Ad. Hoc. Netw. 126 (2022) 102749.
[165] S. Sendra, L. Parra, J.M. Jimenez, L. Garcia, J. Lloret, LoRa-based network for water quality monitoring in coastal areas, Mobile Networks and Applications 28 (1) (2023) 65–81.
[166] B.K. Jha, G. Sivasankari, K. Venugopal, Cloud-based smart water quality monitoring system using IoT sensors and machine learning, International Journal of Advanced Trends in Computer Science and Engineering 9 (3) (2020).
[167] A.P. Rao, K.G. Sarman, G.V.P. Kumar, S.D. Yerra, Water quality monitoring using remote control boat, in: International Conference on Cognitive Computing and Cyber Physical Systems, Springer, 2022, pp. 201–212.
[168] T.M. Murugan, R.K. Shankar, P. Shivkumar, S.R. Kumar, K. Gayathri, A. Jeyam, Monitoring and controlling the desalination plant using IoT, Measurement: Sensors 27 (2023) 100720.
[169] B. Esakki, et al., Design of amphibious vehicle for unmanned mission in water quality monitoring using internet of things, Sensors 18 (10) (2018) 3318.
[170] P. Di Gennaro, D. Lofú, D. Vitanio, P. Tedeschi, P. Boccadoro, WaterS: a Sigfox-compliant prototype for water monitoring, Internet Technology Letters 2 (1) (2019) e74.
[171] C. Jamroen, N. Yonsiri, T. Odthon, N. Wisitthiwong, S. Janreung, A standalone photovoltaic/battery energy-powered water quality monitoring system based on narrowband internet of things for aquaculture: Design and implementation, Smart Agricultural Technology 3 (2023) 100072.
[172] L. González, A. Gonzales, S. González, A. Cartuche, A low-cost IoT architecture based on LPWAN and MQTT for monitoring water resources in andean wetlands, SN. Comput. Sci. 5 (1) (2024) 144.
[173] A.D. Gupta, et al., Devising an iot-based water quality monitoring and ph controlling system for textile etp, in: 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, 2023, pp. 1–6.
[174] V. Gugueoth, S. Safavat, S. Shetty, D. Rawat, A review of IoT security and privacy using decentralized blockchain techniques, Comput. Sci. Rev. 50 (2023) 100585 ed.