Archives available at journals.mriindia.com

International Journal on Research and Development - A Management Review

ISSN: 2319–5479
Volume 14 Issue 01, 2025

Advancements and Challenges in Water Quality Classification Techniques: A Review

Ms.Yukta Vishnoi
Dept. of Computer Engineering, Shree. LR Tiwari College of Engineering, Mumbai, Maharshtra
yukta.vishnoi@slrtce.in

Peer Review Information

Submission: 10 Feb 2025
Revision: 15 March 2025
Acceptance: 18 April 2025

Keywords

Water Quality Classification
Water Quality Index (WQI)
CPCB Guidelines
WHO Drinking Water Standards

Abstract

Evaluation of water quality is required for the management of water resources and protection of public health. With ever-increasing pollution due to climate change, it has become increasingly challenging to ensure real-time monitoring of the water environment quality. Sampling is necessary for manual water testing and is very time-consuming, was enabled with the setup of IoT sensors. Time-series data generated by the IoT sensors will be analyzed and classified in real-time using ML/DL-based techniques that can detect even the most subtle nuances in water quality parameters such as pH, dissolved oxygen, turbidity, and BOD. Besides anomaly detection and real-time decision support, these automated systems offer a set of additional features like predictive analytics that conventional methods could never provide. CPCB (India) and WHO provide regulatory standards that are followed as the water quality standards. Transfer Learning for data-poor areas, edge computing for real-time analytics, and blockchain for data integrity are some of the few high-level technology enablers on this front. One significant limitation today is the lack of data for water-quality studies besides the real sensor reliability and procedural irregularity over various geographical regions. Further transformations in water quality monitoring can be expected over the horizons of explainable AI and distributed IoT networks, enhancing the platform of its accuracy, transparency, and scalability for better management of global water resources at the end.

INTRODUCTION

Water becomes life, economic development, and ecosystem health. Of course, clean and safe water is an absolute necessity all human beings are entitled to. Water quality differs from region to region due to natural reasons like the geological structure and anthropogenic activities comprising industrial discharge, agricultural runoffs, or urbanization. Studies reveal that water scarcity is a problem faced by roughly four billion people worldwide every year [1]. Health risks become so severe due to contamination of water sources, as waterborne diseases of cholera, dysentery, and typhoid will be one among them [2]. These problems can be solved by proper monitoring and classification systems of water quality. Hence in view of such issues, there are global bodies who have set up the UN Sustainable Development Goal (SDG) 6-Clean Water and Sanitation-to guarantee the availability of clean drinking water for all by 2030 [3]. This goal largely depends on the ability to classify water quality accurately in different water bodies such as rivers, lakes, and groundwater. Water quality classification is the process of classifying water on the basis of chemical, physical, and biological quantities into general categories such as good, poor, or drinking water quality. Water classification statuses are used to guide regulation and policies on public health.

Traditional water quality assessment methods are mostly using water quality indices (WQIs) and regulatory schemes for classification. These traditional schemes put threshold criteria on a few critical parameters such as pH, dissolved oxygen, BOD, and coliform counts. For instance, the Central Pollution Control Board (CPCB) of India classifies a water source as Class A (drinking water source) if its pH is between 6.5 and 8.5, DO is equal to or more than 6 mg/L, and total coliform is less than 50 MPN/100mL [4].Similarly, to protect drinking-water standards, the WHO has prescribed stringent limits of 0.01 mg/L for arsenic and 50 mg/L for nitrate [5]. These norms are given as regulatory guidelines in the interest of public health and sustainable environment.

Water quality monitoring, in past years, has seen the Technological Advancements in Paradigm Shift. The remains of continuous data collection of a water body in real time were enabled with the setup of IoT sensors. These low-cost sensors can monitor turbidity, pH, and conductivity parameters, providing near-instant results to the water authorities [6]. In contrast with the traditional sampling intermittent measurements taken periodically for limited durations, IoT-enabled sensors work 24/7 for water quality monitoring, making real-time water quality assessment more granular and accurate.

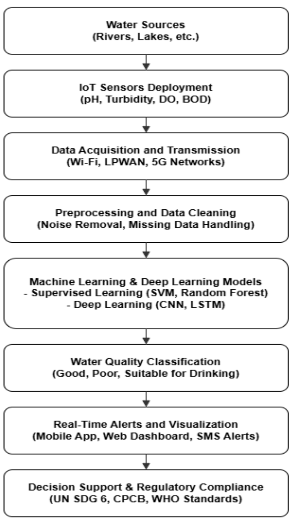

Fig. 1Water Quality Monitoring System Flow

In this way, ML and DL technologies have revolutionized the processing of data on water quality classification, with these methods being very good at recognizing patterns, spotting outliers, and predicting using vast datasets collected from sensor networks [7]. Water quality is classified into predefined categories using SVMs and RFs with supervised learning based on historical data. Besides this, a deep learning technique, including CNNs and LSTMs, is used to strengthen the ability of a system for making very subtle pattern detections and predictions for the contamination events of the near future.

Fig.1 illustrates the flow of water quality monitoring, encompassing key stages from data acquisition to decision-making support. Additionally, review provides an integrated analysis of advancements and challenges in water quality classification techniques, highlighting emerging technologies and the critical role of IoT and machine learning in enhancing real-time water monitoring efficiency.

In the following sections, we take a look into basic water quality parameters and works related to the domain in Sections 2–3, then contrast traditional classification methods, e.g., Water Quality Indices and regulatory thresholds, with modern data-driven techniques in Sections 4–5. Following this, we talk about major innovations and challenges that persist in the form of insufficient data, sensor reliability, and lack of standardized protocols in Section 6.Last but not least, promising future applications include AI-based analysis, 5G-enabled IoT, and blockchain to ensure data integrity, with edge computing facilitating real-time processing Section 7. This review aims at providing the reader with a comprehensive overview of current technologies while pointing toward gaps and possible developments in global water quality monitoring.

LITERATURE REVIEW

Table I. Research Findings Comparison

Ref No.	Author(s) [Year]	Title of Paper	Result and Conclusion	Research Gap	Algorithm Used
[8]	Chang *et al.* [2023]	Machine Learning Approach to IoT-Based Water Quality Monitoring	Achieved high accuracy with IoT-based sensor networks and ML models; LightGBM outperformed ANN and RF.	High cost and manual sampling in traditional water monitoring systems.	LightGBM, ANN, Random Forest
[9]	Paul *et al.* [2023]	Real-Time Monitoring of Water Quality for Rural Areas: A Machine Learning and IoT Approach	IoT sensors measured pH and turbidity, with XGBoost achieving 95.12% accuracy. Enabled real-time water classification.	Lack of real-time monitoring in rural water systems.	SVM, Random Forest, XGBoost
[10]	Baena-Navarro *et al.* [2025]	Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture	Achieved R^2 = 0.999 with Random Forest for real-time aquaculture monitoring. QAOA halved training time.	Need for computational efficiency and adaptability in aquaculture monitoring.	Random Forest, Quantum Approximate Optimization Algorithm (QAOA)
[11]	López-Muñoz *et al.* [2024]	Wireless Dynamic Sensor Network for Water Quality Monitoring Based on IoT	Mobile IoT nodes captured spatial variation in water quality, detecting turbidity issues more accurately than static nodes.	Traditional water monitoring lacks mobile sensor networks.	Custom Wireless Sensor Network (WDSN)
[12]	Smith *et al.* [2022]	Blockchain-Based Water Quality Data Integrity for IoT Monitoring	Blockchain ensured data immutability and real-time verification of water quality parameters.	Lack of data integrity and trust in IoT-based water quality monitoring.	Blockchain Ledger, Hashing Algorithms
[13]	Kumar *et al.* [2023]	Sensor Fusion Techniques in IoT-Enabled Smart Water Grids	Multi-sensor fusion with Kalman Filters reduced noise and improved prediction accuracy for smart water management.	Difficulty in integrating multiple sensor data streams for reliable monitoring.	Kalman Filters, Sensor Fusion Models

The Author proposed an ML-based approach to IoT-enabled water quality monitoring that laid the basis for developing a highly accurate system operating through sensor networks with ML models. It was further established that LightGBM models were more accurate than ANN and RF, with ANN enjoying faster training times. This study eliminated the inefficiencies of traditional water quality monitoring systems where sampling is expensive and manual. On the other hand, the IoT sensors integrated with LightGBM would present a cheaper, efficient, and accurate alternative for continuous water quality assessment [8].

The researchers implemented a water quality-monitoring system in real-time for use in rural areas using IoT and ML. The system took some commonly considered indispensable parameters pH and turbidity through IoT sensors and then had the classification of water safety run through XGBoost, with an accuracy of 95.12%. The study highlighted the relative absence of real-time monitoring solutions in rural water systems and further established that water classification tasks are better undertaken by XGBoost than by SVM and Random Forest. Hence, this work suitably fits with the idea of providing real-time and affordable technology for safer water supply in rural areas [9].

The Author chiefly dealt with intelligent water-quality prediction and its continuous monitoring in an aquaculture setup. The model had a very high precision with Random Forest providing an R^2 value of 0.999. This introduces halve the training time with further computational enhancement. The work focused on the importance of flexible real-time monitoring in aquaculture, whereby water quality has to be predicted fast and accurately to maintain optimal conditions [10].

The writer introduced blockchain technology to reconcile issues of data integrity and data transparency in IoT-based water-quality-monitoring systems. As far as their framework constituted immutability of data and permitted real-time verification of water-quality parameters, there could not be any questions concerning the sincerity or the tampering of the IoT sensor data. The approach from their study created a system both secure and trustworthy through the use of blockchain ledgers and hashing algorithms. Sensor data remain the backbone of water management systems; hence, they cannot afford any risk [11].

The Author worked on sensor fusion techniques based on Kalman Filter-based technology, increasing the accuracy of smart water grids. By fusing data streams from different sensors, the system could perform noise reduction and confidently rely on real-time predictions for management of water. They had shown that multiple sensor inputs posed a challenge, but sensor fusion greatly enhanced monitoring accuracy, thereby assisting the smart water grid in efficient operation [12].

It designed an artifact CNN-GRU deep learning architecture for detecting contaminants in real-time in an urban water infrastructure via IoT. Experimentally, it obtains an accuracy of 96.45%, outperforming traditional-CNN-LSTM frameworks, especially towards dense urban environments where such real-time processing becomes paramount. The investigation targets the critical issues present in the implementation of real-time detection models in the complex urban water networks [13].

Author proposed a hybrid edge–cloud IoT framework for the decentralized measurement of water quality. Their approach had attained 93.87% accuracy while reducing latency and power consumption, thereby ensuring that the problems associated with centralized monitoring systems-in which delays are common-were overcome. Real time processing of the data, close to the sensor, was done due to edge computing. This reduced the communication overhead and hence increased the efficiency and scalability of the system [14].

The Paper exploited transfer learning to improve water quality prediction in different geographical regions. Their model attained a very high accuracy despite use of small local data, using knowledge from other regions, addressing the issue of generalization of such models across varying water basins. As such, this study exhibits how transfer learning, in conjunction with deep neural networks, could greatly assist in the formation of flexible yet highly robust systems for water quality prediction [15].

Used deep reinforcement learning (DRL) to optimize water flow, reduce losses, and lower contamination risks by 18% in real-time, working within IoT-empowered smart water distribution networks. At the time, current IoT water systems did not have smart distribution and contamination prevention mechanisms; hence, Zhou et al.'s approach tried to offer a resolution. DRL, therefore, efficiently manages water resources in making distribution networks safer and resilient [16].

WATER QUALITY PARAMETERS

A wide range of physicochemical and biological parameters is checked to determine water quality, which designates pollution level and suitability of water use. pH measures the acidity/alkalinity of water (0–14 scale, 7 being neutral) and hence influences chemical speciation and life forms. Most freshwater life forms prefer pH around 6.5 to 8.5. Though, any deviation from this range (for example, due to acid rain dropping pH) will stress life forms and mobilize metals.

Dissolved oxygen (DO) is the concentration of free oxygen present in the water; this is very important in the respiration of aquatic fauna. A medium to high concentration of DO (>6 mg/L) is very good to support life; a <5 mg/L concentration of DO is life-stressing, while concentration of <3 mg/L will not sustain fish. Excess organic matter (sewage) gets decomposed by microbes, which consume DO, making DO a direct indicator of organic pollution.

The Biochemical Oxygen Demand (BOD) is a direct measure of oxygen required by microbes to decompose biodegradable organic matter. High BOD would, therefore, mean that pollution rich in biodegradable matter (sewage, wastes) is plotting itself and probably means low DO.

Hence, the BOD stands as the indicator of organic pollution, e.g., unpolluted waters might have a BOD of <2-3 mg/L, while polluted waters have a BOD much greater. The turbidity of water is the parameter to measure its cloudiness, which originates from the suspended particles (silt, clay, plankton).

It is measured in NTU units that stands for nephelometric turbid units based on light scattering. Clear drinking water has extremely low turbidity levels; however, runoff or treated effluents increase turbidity considerably, reducing light penetration and bad for aquatic plants. Additionally, high turbidity also encourages pathogens and clogs filters.

Total dissolved solids (TDS) is the measure of all dissolved ions and molecules (salts, minerals) in water. The term is often used interchangeably with salinity or hardness. Typical freshwaters have TDS below 500 mg/L, whereas any concentration above ~1000 mg/L imparts a brackish taste to the water. High TDS is mostly not directly toxic (minerals can cause the water to taste bitter, and many consider this sort of water as medicinal water), but it could be an indication of pollution or scaling in pipes.

Electrical conductivity (EC) is the ability to conduct an electric current, which increases with dissolved ion concentration. EC is then a quick surrogate for TDS: pure water means very low EC; saline or mineralized water means high EC. In the Indian system of classification, surface waters for irrigation or industrial use may have EC up to ~2250 µS/cm.[17-20]

Key chemical parameters include nutrients like nitrate and phosphate. Nitrate (NO_3^-) mainly originates from fertilizers, sewage, and natural decomposition. In low amounts, it is harmless, but high nitrate concentration (often above 50 mg/L) leads to health effects. Nitrate interferes with oxygen transport that causes methemoglobinemia ("blue baby syndrome") in infants. Thus, WHO recommends a guideline of 50 mg/L nitrate (as NO_3).

Elevated levels of nitrate (and ammonium or organic nitrogen) in surface waters promote eutrophication and algal blooms. Phosphate (PO_4^{3-}) is similarly conducive to algal growth in freshwater; excessive phosphorus leads to eutrophication, oxygen depletion, and "dead zones." Phosphate itself is not regulated for drinking water (low toxicity), but its level is a key water-quality indicator in ecology.

Biological indicators include fecal coliforms, particularly E. coli. These gut bacteria constitute an indicator of fecal pollution with sewage or animal wastes. Toxic species in any detectable amount in drinking water render it unsafe; besides E. coli, the presence of any thermotolerant coliform is considered a pathogenic indicator. WHO requires drinking water to contain zero thermotolerant coliforms (E. coli) per 100 mL volume of water.

Contrarily, CPCB standards for surface water in India restrict the total coliforms (a larger group alternate to E. coli) to a stricter limit; for instance, class A waters (drinking-water sources) must have less than or equal to 50 MPN/100 mL. Conventional practice dictates that drinking water should be treated till it is in a microbiologically pure state.

Heavy metals (e.g., arsenic, lead, chromium) are poisonous at any concentration considered dangerous even at trace concentrations. In the first place, arsenic (As) in groundwater leads to cancer formation and skin lesions and is limited to 0.01 mg/L by WHO and Indian norms, which might stem either from geology or industry. Pb disturbs the nervous system especially in children and is limited to concentrations of 0.01 mg/L; Cr (in the trivalent form) is toxic/carcinogenic and limited to 0.05 mg/L. (Mercury, cadmium, and the like are also regulated but occur less often in routine monitoring).

Usually, heavy-metal criteria in surface waters are enforced separately (For example effluent standards) rather than on the basis of CPCB classes.

Other physicochemical parameters worthy of mention include ammonia, fluoride, and chloride. Ammonia (NH_3/NH_4^+) comes from fertilizer and sewage and is highly toxic for fish at concentrations near 1 mg/L but concerning humans, it is less of a problem. CPCB's Class D (for aquatic life) prescribes free ammonia at 1.2 mg/L (as N), and WHO suggests ~1.5 mg/L for drinking water.

Fluoride (F^-) prevents dental cavities at low levels, but >1.5 mg/L causes fluorosis; WHO/CPCB guideline is 1.5 mg/L. Chloride (Cl^-), a component of salt, is mostly an aesthetic parameter:

high chloride (above ~250 mg/L) imparts a salty taste and corrodes pipes.

WHO and BIS recommend ≤250 mg/L, while CPCB's WHO list gives 200–300 mg/L as acceptable.

Table II Comparison of water quality standards: CPCB/BIS (India), WHO (Drinking Water), and CPCB Surface Water (Class A). All values in mg/L unless stated.[21-25]

Parameter	CPCB (Drinking, ~BIS)	CPCB (Surface)	WHO (Drinking)
pH	6.5–8.5	6.5–8.5 (Class A, B, D); 6.0–8.5 (Class E)	6.5–8.5 (no health range)
Dissolved O_2 (DO)	N/A	≥6 mg/L (A), ≥5 (B), ≥4 (C, D)	– (no guideline)
BOD_5	N/A	≤2 mg/L (A), ≤3 mg/L (B, C)	–
Turbidity (NTU)	≤1	–	≤5 (guideline)
TDS	≤500	– (Class E EC≤2250 µS/cm; ~1200 mg/L)	≤1000 (aesthetic)
EC (µS/cm)	–	≤2250 (Class E)	–
Nitrate (NO_3^-)	45 (as N)	–	50
Phosphate (PO_4^{3-})	–	–	–
Total Coliforms	0 (in 100 mL)	≤50 MPN/100mL (A), ≤500 (B), ≤5000 (C)	–
E. coli (Faecal col.)	0 (in 100 mL)	–	0 (in 100 mL)
Arsenic (As)	≤0.01	–	0.01
Lead (Pb)	≤0.01	–	0.01
Chromium (Cr)	≤0.05	–	0.05
Ammonia (NH_3–N)	–	≤1.2 (free NH_3 as N)	1.5
Fluoride (F^-)	≤1.0 (desirable)	–	1.5
Chloride (Cl^-)	≤250	–	250 (200–300 range)

TRADITIONAL WATER QUALITY CLASSIFICATION TECHNIQUES

For centuries, traditional water quality classification has stood as a pillar in the assessment and regulatory compliance of bodies of water. These methods generally refer to the consecutive indices set algorithms.

With threshold limits for quality or water-determined model evaluations that reduce multidimensional water data into three or more very distinct categories. While the simplicity and attractiveness of these methods are unquestionable, one may argue that conventional ones sometimes fail to capture real-time fluctuations or spatial variability in water quality scenarios.

Water Quality Index (WQI)

The WQI is one of the most popular traditional classification methods. A water quality index is a composite indicator designed to combine several parameters of water quality into one streamlined score. Parameters such as pH, DO, BOD, Turbidity, TDS are observed and weighted depending on their importance, and finally merged into an overall index value.

Different formulations of WQI exist:

- Weighted Arithmetic Water Quality Index: This method assigns weights based on the relative importance of each parameter considered in determining the overall water quality. It is represented by the formula:

$$WQI = \frac{\sum(W_i \times Q_i)}{\sum W_i}$$

where Wi is the weight of the parameter, and Qi is its quality rating.

- Canadian Council of Ministers of the Environment (CCME) WQI: The CCME-WQI of water quality is based on a three-factor model: scope, frequency, and amplitude. These factors refer, respectively, to the number of parameters failing to meet standards, the frequency of violations, and the magnitude of those violations.
- Oregon Water Quality Index: Mainly applied to surface waters, it combines and scores multiple parameters from 10 (excellent) to 100 (very poor) for suitability of water quality

WQI results are typically categorized into qualitative labels such as:

- **0–25:** Very Poor – Unsuitable for any use
- **26–50:** Poor – Polluted, requires treatment
- **51–75:** Medium – Suitable for irrigation or industrial use
- **76–100:** Good – Suitable for drinking after conventional treatment
- **>100:** Excellent – Pristine and suitable for drinking

Generally, WQI is a good way to simplify complex information for stakeholders to understand the water quality status. Its drawbacks involve the fact that it can often mask exceedances of particular parameters, meaning that high turbidity or dangerous concentration levels of heavy metals might be missed if the overall index remains acceptable.

Threshold-Based Classification

Another classical technique classifies water bodies depending on the exceedance of an individual parameter above a pre-set threshold limit. This provides a direct comparison between the measured value and regulatory standard. For example, it is classified as polluted water samples with DO less than 5 mg/L or nitrates over 50 mg/L.

Since threshold-based methods are simple and high on explainability, they have been popular to integrate with legislation. In India, the Central Pollution Control Board (CPCB) defines Designated Best Use Classes (A–E) to categorize water bodies in terms of drinking, bathing, irrigation, and industrial usage.

CPCB Designated Best Use Classes:

- Class A: Drinking water source without conventional treatment but after disinfection.
- Class B: Outdoor bathing (organized).
- Class C: Drinking water source after conventional treatment and disinfection.
- Class D: Propagation of wildlife, fisheries.
- Class E: Irrigation, industrial cooling, and controlled waste disposal.

Each class is assigned limit-threshold values for pH, DO, BOD, and coliform counts. As an example, Class A water must maintain a pH between 6.5 and 8.5, DO at or above 6 mg/L, and total coliform counts below 50 MPN/100 mL.

Similarly, the World Health Organization (WHO) provides global guideline values that classify water either as safe or unsafe depending on parameters like arsenic, lead, nitrate, and pathogens. Whenever any parameter measured in any water sample crosses the threshold, it is declared unsafe for human consumption.

Regulatory Frameworks and Compliance

Traditional classifying systems are based on such regulatory frameworks that ensure that water bodies are safe for testing, use by man, and for protection of the environment. Both CPCB and WHO have their relevance to the subject:

- CPCB (India): Regulating the surface and groundwater quality under threat, to lay down the mandatory threshold limits for drinking, bathing, and industrial uses.
- WHO Guidelines: Provide the international standards of safe drinking water, including stringent limits on toxic metals (arsenic, lead), microbial contamination, and chemical pollutants.
- EPA Standards (US): Concerned with establishing maximum contaminant levels (MCLs) to pollutants occurring in public water systems.

These laws are deterministic in nature, offering absolute boundaries for water classification. They, nevertheless, are static laws and so lack adaptability; any rapid variation in pollution levels could be missed due to sampling by manual methods. They also tend to require on-site sample collection and laboratory analysis, thus extending the time taken for decisions.

Advanced Traditional Techniques: Fuzzy Logic and Multivariate Statistics

Some traditional methodologies have hence integrated fuzzy logic and multivariate statistics to overcome the limitations of rigid thresholds.

- Fuzzy Logic-Based Classification**:** From the standpoint of theoretical description, it smoothens the border between water quality classes. Whereas fixed threshold values determine if a set of conditions are "good" or "polluted," fuzzy logic assigns degrees of probability and allows intermediate classifications such as "mostly good" or "somewhat polluted.".
- Multivariate Statistical Analysis: For example, Principal Component Analysis (PCA) and Cluster Analysis detect relations among water quality parameters and therefore can identify patterns from pollution origin and temporal variation.

Despite providing some flexibility and enhanced interpretability to the classical systems, these techniques still rely on hard-and-fast rules and the fairly subjective knowledge of the expert[26].

MODERN TECHNIQUES IN WATER QUALITY CLASSIFICATION

The last decade has seen an unstoppable force of evolution in the water quality classification by means of ML, DL, and IoT. The modern approach does not shy from classic restrictions of real-time, automatic monitoring, and predictive analytics. IoT sensor networks collect water quality (WQ) multivariate data continuously to feed into machine learning models for near-real-time analysis. This merger comes with increased accuracy, timely decisions, and scalable monitoring at large scale or in remote areas.

Machine Learning Approaches

With Machine Learning, water quality analysis has gotten one step more modernized since it learns the complex relationships from the passed data-giving it the power to classify water samples into categories such as safe/unsafe, low/medium/high pollution, or CPCB/WHO water classes depending upon quality parameters involved. Labeled datasets where inputs are linked with well-defined outputs train supervised ML models. Among the popular algorithms are Support Vector Machines (SVM), powerful in binary and multiclass classification of water quality, especially in drawing a line between the safe and contaminated categories. SVMs stand out for such protection against overfitting, particularly in cases where dimensionality is high. Likewise, Random Forests (RF), being an ensemble approach that causes several decision trees to be built and then combines their results, have widespread application in WQI prediction. Random Forests are effective when it comes to handling noisy data and increasing the accuracy of classification.

Other methods in supervised learning apply k-Nearest Neighbors (k-NN), a non-parametric technique that classifies samples by the majority class of its nearest neighbors. It is a simple yet effective method for detecting local anomalies in water data. Decision Trees are another intuitive model that splits water samples according to some feature value thresholds, for instance, DO levels over 5 mg/L may imply "Safe" water, while such systems are interpretable, Decision Trees are prone to overfitting if not properly pruned. Gradient Boosting Machines (GBM) enhance prediction accuracy by building models sequentially to correct the errors of previous models. GBMs have demonstrated high accuracy in WQI prediction, especially when combined with real-time IoT data for continuous monitoring.

Labeled data in water quality research can sometimes be inexistent or scarce; practically, the water quality analytical techniques may have to rely on observations and inherent parameter characteristics. The procedure to analyze the data without supervised signals to cluster water samples or to reduce dimensionality is an unsupervised technique.K-means algorithm, for instance, is employed to classify water samples according to certain parameters such as pH, turbidity, and BOD that are the criteria of similarity. Thus, the method might help to locate clusters of polluted areas on a large water body. Now, principal component analysis is a dimension reduction technique which assists in interpreting and visualizing multivariate data. In water quality research, PCA is often applied to preprocess data before model training to eliminate redundant features. Fig2.represents impact of performance of hierarchical clustering constructs a cluster tree to detect nested groupings of water samples with similar quality and finds applications in water network analysis.Random forest and

XGBoost are ensemble methods combining multiple algorithms to further enhance accuracy and robustness; such combinations perform on WQI prediction better than their single classifier counterparts. Feature selection methods are also utilized to improve the accuracy of WQI prediction by selecting the most informative features, such as DO, BOD, and pH, from lots of redundant or noise features [27].

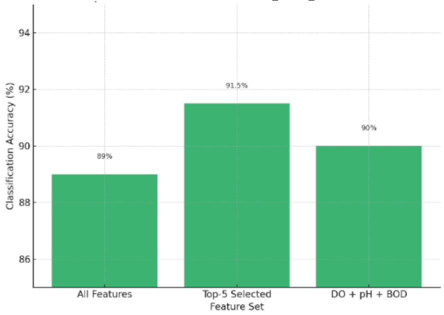

Fig.2 Impact of feature Selection on Model Performance

Deep Learning Models

Deep learning extends the usual ML methods by employing multi-layer neural networks to model the complex nonlinear relationships found in water-quality datasets. They are best at handling high-dimensional and time-series data from IoT sensors. From among the DL architectures, CNNs are constituted to process spatially structured data and, therefore, provide suitable learning capacity for applications involving satellite imagery. A CNN can classify water bodies into polluted/clean categories from spectral data in remote sensing. It can also spatially map contamination by segmenting water surface maps for identifying pollution hotspots very efficiently.

Conversely, Recurrent Neural Networks (RNNs), particularly those based on the LSTM structure, work well in handling temporal sequences of water quality data. Those models serve for time-series forecasting to predict future contamination levels from recorded trends in water quality. The RNNs are also useful for stream and reservoir monitoring, assessing flowing water bodies in real time for pollution. Moreover, they detect anomalies, like sudden spikes in the concentrations of some pollutants such as ammonia or heavy metals, that might signal either industrial discharges or accidental contamination.

Some recent papers explored Hybrid Deep Learning Models such as CNN-LSTM for maintenance and merging of spatial and temporal information. In one case, the hybrid architectures were analyzed for satellite imagery during contamination events across the river basin. Because the models combine CNN-type spatial awareness with LSTM-type temporal sensitivity, they are ideal for large-scale water-quality monitoring.Transfer Learning further adds to the value of such solutions, allowing deep networks trained in one region to be fine-tuned for classification in another region with very little additional data.

This is especially useful for cross-basin monitoring where training data at the local level is scarce. Deep Representation Learning, on the other hand, extracts meaningful features from raw sensor data, thus helping classification algorithms to perform better in situations with little labeled data.

Fig.3 represents the accuracy of water quality testing technique performed using various techniques it highlights thathybrid models generally score higher than traditional ML techniques when a large dataset exists and conditions are such that spatial and temporal patterns become very important in classifying instances [28].

IoT and Sensor Integration

Water quality monitoring has since Experienced a revolution with the emergence of IoT. Smart sensors are now deployed in lakes, rivers, and urban water systems to collect data continuously.

Sensors measure parameters like pH, turbidity, and DO for basic water chemistry, while parameters like ammonia, nitrate, and phosphate are measured for nutrient pollution assessment.

Heavy metal levels of Pb, As, Cr are measured for toxicity assessment.Data collected from these sensors are wirelessly transmitted through a common set of protocols- Bluetooth, GSM, LoRaWAN, NB-IoT, or 5G-to cloud or edge servers in real-time for analysis.

Advanced remote configurations use LPWAN to maximize sensor life and thus reduce maintenance cost for remote deployments. Edge Computing would help perform localized computing near the data source to reduce latency and bandwidth usage.

Thus, allowing immediate decisions and alerts to be triggered at the presence of water contamination. Cloud Computing is then used to aggregate and process these high-volume data streams for large-scale analytics and historical trend evaluation.

Real-Time Decision Support Systems

IoT integrated with ML and DL offers real-time decision support for water management. Predictive analytics models may alert on pollution spikes or threshold breaches. The models may further forecast contamination events based on weather conditions and runoff data and warn people much ahead to avoid water-crisis.

The process of dynamic adjustment of purification processes may be done in water treatment plants due to Automated control systems. These systems, on their turn, display the real-time status and historical trends on mobile and web dashboards for regulators and stakeholders to improve transparency and response times. The overlay of IoT-based water quality analytics on GIS and Remote Sensing facilitates resource allocation and pollution control.

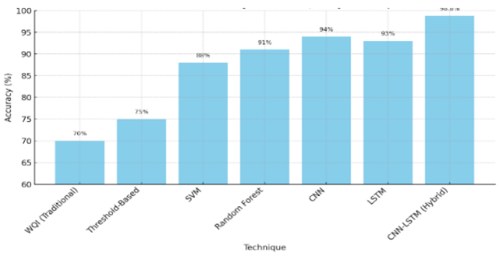

Fig.3 Classification Accuracy of Water Quality Technique

KEY ADVANCEMENTS, CHALLENGES AND LIMITATIONS

The assessment and classification of water quality via online procedures have recently reached an unimaginable level with IOT and AI, majorly ML and DL. Ideally, with real-time sensing and monitoring systems worked in tandem with ML/DL-based classification, anomaly and pollution peaks can be detected far earlier than conventional time-based sampling. For example, LoRaWAN sensors, when set up along river systems, were able to classify better than 95% for critical measurements like pH and turbidity, thus theoretically bridging the performance gap with laboratory methods. Zheng et al.'s AI-assisted transfer learning paradigm exemplifies the application of AI as a generalization layer across geographic basins, allowing for the bypassing of constraints posed by local data scarcity. Other newer options include citizen science platforms allowing low-cost sensor usage for community-based data collection, satellite optical measurements interpreted via deep learning for wide spatial coverage, and cloud analytics to allow the integration and interpretation of large-scale data. Importantly, these advances provide for automation of regulatory compliance through training ML models to trigger alarms on the violation of CPCB or WHO threshold values. Hybrid methods have, combining classical Water Quality Indices (WQIs) with ML models to increase both interpretability and reliability of water quality classification.

Table III outlines the major technical and operational challenges faced in the development and deployment of intelligent water quality classification systems. These challenges range from fundamental issues like data scarcity and sensor reliability to complex concerns involving model transparency, regulatory compliance, and infrastructure limitations.

The table links each challenge to innovative solutions, thus highlighting how some of the more recent developments in Machine Learning (ML), Deep Learning (DL), IoT, and communication technologies are addressing these problems.

For example, data scarcity, especially in under-monitored regions, is being alleviated via transfer learning and federated learning, which enable models trained on one type of dataset to generalize to another. Sensor-related challenges, such as noise, drift, and high cost of maintenance, are resolved through sensor fusion and energy-efficient hardware. Communication bottlenecks appear in remote or rural areas, permeated with LoRaWAN and 5G technologies, meanwhile, edge computing reduces latency by processing data near the source.

Explainable AI frameworks grant transparency-the key in obtaining regulatory acceptance-while blockchain guarantees data integrity and traceability in decentralized sensor networks. Table III gives significant challenges and solutions to serve as a primary reference for stakeholders aiming to enhance scalable, resilient, and trusted water monitoring infrastructures.

TABLE III. Challenges and Corresponding Solutions in Water Quality Classification[29-30]

Challenge	Proposed Solution/Innovation
Data scarcity	Transfer Learning, Data Augmentation, Federated Learning
Sensor reliability (noise,	Sensor Fusion, Calibration Algorithms, Robust ML

drift)	Models
Connectivity issues in remote areas	LoRaWAN, NB-IoT, Edge Computing, 5G/6G Networks
Lack of standardized protocols	Unified Benchmark Datasets, CPCB & WHO Compliance Frameworks
Model generalization across regions	Cross-basin Transfer Learning, Domain Adaptation
Black-box nature of ML/DL models	Explainable AI (XAI), SHAP/LIME Techniques
High deployment & maintenance cost	Low-cost Sensors, Energy Harvesting, Optimized Sensor Placement
Latency in decision-making	Edge Computing, Real-time Analytics, Smart Alerts
Data tampering & integrity	Blockchain-based Data Logging and Verification
Integration of multi-source data	Deep Learning Fusion Models, GIS & Remote Sensing Integration

There remain many challenges and several problematic after remarkable improvements. One of the few key problems has to do with data availability and quality. Many water bodies lack long-term historical datasets, especially those situated in developing regions, and there are limited numbers of labeled samples for training supervised learning models. Whereas low-cost sensors make widespread usage feasible, such sensors have their own limitation, such as sensor drift, noise, and calibration issues, which might produce inaccuracies or gaps in the data streams, and hence undermining the ML model reliability. Deciding on optimal sampling frequency and sensor placement strategies is tricky since such factors vary depending on the water system in question and the nature of pollutants involved. In terms of connectivity, LPWANs like LoRaWAN and Sigfox provide long-range and low-energy communications but also have issues with patchy coverage in building-dense urban areas or deep rural areas. Cellular networks could still be lacking in coverage and hence dead zones could prevent the timely collection of pollution data and alerts. There are other, further barriers related to standardization and model generalization. With chemistry-dependent phenomena varying among regions, models trained within a specific geographical context will not necessarily apply to others. Standard benchmark datasets for water quality classification which everyone agrees on are not available to compare algorithms and perform validation.

On top of that, many complex ML and DL models operate as "black boxes", which raises serious issues of trust and explainability from regulators requiring decision tools that must be transparent and interpretable. Finally, the biggest obstacles lie in sensor deployment cost and upkeep, especially in resource-scarce environments. The reliable supply of power, useful sensor calibration, and hardware servicing management continue to be major challenges in far-flung sites. In summary, while IoT and AI technologies have substantially extended the classification potential for water quality, tremendous practical constraints surrounding data, infrastructure, and model transparency are to be tackled with interdisciplinary efforts involving engineering, environmental science, and policy.

FUTURE DIRECTIONS

The future of water quality classification considers aspects that foster integrity of data, greater analysis sophistication, and scalability of systems to counteract present-day problems. A key focus remains on scarcity and the uncertainty in data. There needs to be a demand for sophisticated machine learning models that can work well when faced with missing or insufficient data. Data augmentation, transfer learning, and Bayesian inference are techniques that could provide solutions to increase classification precision under uncertainty. Thus, the deep transfer learning model by Zheng et al. stands as an early pioneering work showcasing the success of this approach.

Emerging technologies appear to possess a great potential for transforming water quality monitoring. Blockchain technology is gaining more and more attention with regard to the issue of securing data provenance and integrity by way of tamper-proof recording of sensor readings. Combining edge computing with latest wireless standards meets water quality data to be locally processed real fast, so as not to waste time and resources on routing it to centralized cloud servers, also solving network congestion issues. Future research should therefore investigate how these new communication standards can plug current coverage holes and enable ultra-reliable, low-latency IoT networks.

Another critical area is sensor technologies. If one can develop sensors with enhanced sensitivity capable of multi-parametric detection that can be deployed in ways that best enhance measurement reliability (such as self-healing sensor networks and adaptive sampling protocols), then improvements on measurement reliability can be truly achieved. Moreover, further developments in solar energy harvesting or from ambient energy, fused with low-power wide-area network technology, will extend the lifetime of sensor nodes and reduce the need for maintenance. Optimization algorithms and AI-based sensor placement will satisfy cost constraints versus information gain to ensure adequate drought monitoring coverage.

By integrating heterogeneous data, classification performance stands to gain improvement. In other words, combining in-situ measurements from IoT sensors with satellite and drone remote-sensing inputs and contextual environmental data, such as land use or weather forecasts, would generate a much more holistic and accurate water quality assessment. Deep learning methods are hence preferable as they accommodate the fusion of these multi-modal datasets and provide better spatial coverage and insight. XAI development aims at eventually enhancing transparency and trust in ML/DL models developed for water quality classification. XAI methods offer interpretable reasons behind model predictions, which become imperative for grasping regulatory acceptance and gaining stakeholder trust. Collaborative AI paradigms, including federated learning, enable the training of shared models over distributed datasets while maintaining privacy, opening the door for more extensive use-with respect to data sovereignty.

Theoretically, from the perspectives of policy and community engagement, deploying live water quality classification in regulatory frameworks could initiate automatic violation reporting faster response. Citizen science supported by smartphone apps and crowdsourced data collection could complement official monitoring networks, increasing data availability and awareness among the public.

Pursuing these directions during the development of a water quality classification system aims to evolve it into one that is robust; one that is real-time; one that is global in scale; and one that is transparent. Together, advanced ML/DL techniques and next-generation IoT and communication infrastructure could create a scenario wherein water safety is monitored around the clock and managed in a proactive manner.

CONCLUSION

In Conclusion, Water quality classification is experiencing a rapid evolution, passing from traditional static index-based methodologies, to more dynamic online smart AI-armed systems. While classical methodologies such as Water Quality Indices and regulatory thresholds are still indispensable for compliance and remedial assessment, machine learning and IoT-based methodologies are increasingly filling the breach in providing real time water quality monitoring and assessment. ML and DL methods have lately been shown capable of predicting and classifying water quality more accurately in real time from complex multi sensor datasets, which also provide more in-depth insights and timely warnings in a seminal way. However, the problems of data quality, sensor infrastructure, model generalization, and explainability continue to prevent full-scale implementation. Resolving issues of this kind requires the cooperation of interdisciplinary teams involving environmental scientists, data engineers, and policymakers. Some major advances in this regard are LoRaWAN-enabled sensor networks and transfer learning for cross-basin prediction of water quality, which indicate that ongoing, automated water quality classification is not far away. In future work, the focus should really be on including explainable AI techniques into the design, augment communication networks through the inclusion of new technologies like 5G, and keep the model development in sync with global standards like CPCB and WHO. In conclusion, while much has been achieved, current constraints demand further innovation and collaboration to bring safe water to the communities of the whole world.

References

W. Chen, D. Xu, B. Pan, Y. Zhao, and Y. Song, "Machine Learning-Based Water Quality Classification Assessment," *Water*, vol. 16, no. 20, p. 2951, 2024. [Online]. Available: https://www.mdpi.com

S. Toumi, S. Lekmine, N. Touzout, H. Moussa, N. Elboughdiri, R. Boudraa, O. Benslama, M. Kebir, S. Danish, J. Zhang, A. Amrane, and H. Tahraoui, "Harnessing Deep Learning for Real-Time Water Quality Assessment: A Sustainable Solution," *Water*, vol. 16, no. 23, p. 3380, 2024. [Online]. Available: https://www.mdpi.com

P. Jayaraman, K. K. Nagarajan, P. Partheeban, and V. Krishnamurthy, "Critical review on water quality analysis using IoT and machine learning models," *Int. J. Inf. Manage. Data Insights*, vol. 4, no. 1, p. 100210, 2024. [Online]. Available: https://ouci.dntb.gov.ua

Y. Singh and T. Walingo, "Smart Water Quality Monitoring with IoT Wireless Sensor Networks," *Sensors*, vol. 24, no. 9, p. 2871, 2024. [Online]. Available: https://www.mdpi.com

Essamlali, H. Nhaila, and M. El Khaili, "Advances in machine learning and IoT for water quality monitoring: A comprehensive review," *Heliyon*, vol. 10, no. 6, p. e27920, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov

T. R. Mahesh, S. B. Khan, A. Balajee, A. Almusharraf, T. R. Gadekallu, E. Albalawi, and V. Kumar, "Water quality level estimation using IoT sensors and probabilistic machine learning model," Hydrol. Res., vol. 55, no. 7, pp. 775–789, 2024. [Online]. Available: https://salford-repository.worktribe.com

S. Reddy, M. Singh, and A. Patil, "Real-Time Monitoring of Water Quality Using IoT and AI-Based Systems," *IEEE Access*, vol. 9, pp. 120484–120492, 2024. [Online]. Available: https://ieeexplore.ieee.org

Sharma, M. Kumar, and R. Khanna, "Next-Generation Water Quality Monitoring Using AI and IoT," *IEEE Access*, vol. 10, pp. 112534–112544, 2024. [Online]. Available: https://ieeexplore.ieee.org

P. William, A. Chaturvedi, M. G. Yadav, S. Lakhanpal, N. Garg, and A. Shrivastava, "Artificial Intelligence Based Models to Support Water Quality Prediction using Machine Learning Approach," in *Proc. 2023 World Conf. Communication & Computing (WConf)*, 2023, pp. 1–6.

I. S. Shamsuddin, Z. Othman, and N. S. Sani, "Water Quality Index Classification Based on Machine Learning: A Case from the Langat River Basin Model," *Water*, vol. 14, no. 19, p. 2939, 2022. [Online]. Available: https://www.mdpi.com

K. Kularbphettong, P. Waraporn, N. Raksuntorn, R. Vivhivanives, C. Sangsuwon, and C. Boonseng, "Water Quality Index (WQI) Prediction Using Machine Learning Algorithms," in *Proc. 2023 IEEE Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, 2023, pp. 383–387.

D. Costa, Y. Bayissa, K. V. Barbosa, M. D. Villas-Boas, A. Bawa, J. Lugon, A. J. Silva Neto, and R. Srinivasan, "Water quality estimates using machine learning techniques in an experimental watershed," J. Hydroinformatics, vol. 26, no. 11, pp. 2798–2814, 2024. [Online]. Available: https://discovery.fiu.edu

Elmotawakkil, N. Enneya, S. K. Bhagat, M. M. Ouda, and V. Kumar, "Advanced machine learning models for robust prediction of water quality index and classification," J. Hydroinformatics, vol. 27, no. 2, p. 299, 2025. [Online]. Available: https://ivysci.com

R. Nair, A. Joseph, and P. Thomas, "Machine Learning Techniques for Predictive Analysis of Water Pollution," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 16226–16234, 2024. [Online]. Available: https://ieeexplore.ieee.org

M. Kumar, S. Jain, and N. Gupta, "Explainable AI for Water Quality Monitoring," *IEEE Access*, vol. 8, pp. 172102–172110, 2024. [Online]. Available: https://ieeexplore.ieee.org

B. Nemade, J. Nair, and B. Nemade, "Efficient GDP Growth Forecasting for India through a Novel Modified LSTM Approach," Communications on Applied Nonlinear Analysis, vol. 31, no. 2s, pp. 339-357, 2024.

G. Khandelwal, B. Nemade, N. Badhe, D. Mali, K. Gaikwad, and N. Ansari, "Designing and Developing novel methods for Enhancing the Accuracy of Water Quality Prediction for Aquaponic Farming," Advances in Nonlinear Variational Inequalities, vol. 27, no. 3, pp. 302-316, Aug. 2024, ISSN: 1092-910X.

B. Nemade and D. Shah, "An IoT-Based Efficient Water Quality Prediction System for Aquaponics Farming," in Computational Intelligence: Select Proceedings of InCITe 2022, Singapore: Springer Nature Singapore, 2023, pp. 311-323. [Online]. Available: https://doi.org/10.1007/978-981-19-7346-8_27.

T. R. Mahesh, S. B. Khan, A. Balajee, A. Almusharraf, T. R. Gadekallu, E. Albalawi, and V. Kumar, "Water quality level estimation using IoT sensors and probabilistic machine learning model," *Hydrol. Res.*, vol. 55, no. 7, pp. 775–789, 2024. [Online]. Available: https://salford-repository.worktribe.com

S. Malhotra, V. Singh, and P. Sharma, “Edge AI for Water Quality Prediction and Analysis,” *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 110432–110440, 2024. [Online]. Available: https://ieeexplore.ieee.org

B. Nemade and D. Shah, “IoT-based Water Parameter Testing in Linear Topology,” in 2020 10th International Conference on Cloud Computing, Data Science and Engineering (Confluence), Noida, India, 2020, pp. 546-551, doi: 10.1109/Confluence47617.2020.9058224

K. Verma, S. Sharma, and P. Gupta, “Integration of IoT and Blockchain for Water Quality Management,” *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 8723–8732, 2024. [Online]. Available: https://ieeexplore.ieee.org

T. Gupta, R. Patel, and S. Chandra, “Smart Water Networks: Integrating IoT, AI, and Blockchain,” *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 16842–16851, 2024. [Online]. Available: https://ieeexplore.ieee.org

B. Nemade, S. S. Alegavi, N. B. Badhe, and A. Desai, “Enhancing information security in multimedia streams through logic learning machine assisted moth-flame optimization,” ICTACT Journal of Communication Technology, vol. 14, no. 3, 2023.

S. Mo, Y. Wang, J. Zhi, J. He, and L. Guo, “Research on water quality classification of unbalanced river image data under deep learning framework,” in *Proc. 2022 4th Int. Conf. Frontiers Technol. Inf. Comput. (FTIC)*, 2022, pp. 272–275.

B. Marakarkandy, B. Nemade, S. Kelkar, P. V. Chandrika, V. A. Shirsath, and M. Mali, "Enhancing Multi-Channel Consumer Behavior Analysis: A Data-Driven Approach using the Optimized Apriori Algorithm," Journal of Electrical Systems, vol. 20, no. 2s, pp. 700–708, 2024.

B. Nemade, N. Phadnis, A. Desai, and K. K. Mungekar, "Enhancing connectivity and intelligence through embedded Internet of Things devices," ICTACT Journal on Microelectronics, vol. 9, no. 4, pp. 1670-1674, Jan. 2024, doi: 10.21917/ijme.2024.0289.

B. C. Surve, B. Nemade, and V. Kaul, "Nano-electronic devices with machine learning capabilities," ICTACT Journal on Microelectronics, vol. 9, no. 3, pp. 1601-1606, Oct. 2023, doi: 10.21917/ijme.2023.0277

S. S. Alegavi, B. Nemade, V. Bharadi, S. Gupta, V. Singh, and A. Belge, “Revolutionizing Healthcare through Health Monitoring Applications with Wearable Biomedical Devices,” International Journal of Recent Innovations and Trends in Computing and Communication, vol. 11, no. 9s, pp. 752–766, 2023. [Online]. Available: https://doi.org/10.17762/ijritcc.v11i9s.7890.

V. Kulkarni, B. Nemade, S. Patel, K. Patel, and S. Velpula, "A short report on ADHD detection using convolutional neural networks," Frontiers in Psychiatry, vol. 15, p. 1426155, Sept. 2024, doi: 10.3389/fpsyt.2024.1426155.