ISSN: 2096-3246
Volume 57, Issue 02, October, 2025

A COMPREHENSIVE SURVEY OF AI-DRIVEN AND BIG DATA ANALYTICS-BASED APPROACHES FOR INTELLIGENT IOT-ENABLED WATER QUALITY MONITORING

C. V. Radhakrishnan[1], Dr. S. Ashok Kumar[2]
[1]Research Scholar, School of Computer Studies,
A.V.P. College of Arts and Science (Co-education), Tirupur, Tamil Nadu, India
[2]Research Supervisor, Assistant Professor and Dean A&R, School of Computer Studies,
A.V.P. College of Arts and Science (Co-education), Tirupur, Tamil Nadu, India

Abstract
Ensuring access to safe water is vital for public health and environmental sustainability. However, conventional water quality monitoring methods are limited by high costs, lack of scalability, and delayed analysis. This survey explores recent innovations in IoT-based Water Quality Monitoring Systems (IoT-WQMS) integrated with Machine Learning (ML) and Deep Learning (DL) to enable real-time, automated water quality assessment. The review covers architectures utilizing multi-parameter sensors (e.g., pH, turbidity, TDS, DO, temperature), along with essential data processing techniques such as imputation and normalization. Advanced feature selection methods (RF-MOA, ensemble voting) and hyperparameter tuning techniques (QPSO, Grid Search) are discussed for model optimization. ML models like XGBoost, Random Forest, and ANN, as well as DL models such as CNN-LSTM and MS-CAGRU, demonstrate predictive accuracies up to 99.9%, supporting early contamination detection and regulatory compliance. Applications span urban rivers, aquaculture, and groundwater systems, offering actionable insights for efficient and sustainable water management. The paper also addresses key challenges including sensor calibration, data heterogeneity, and model adaptability, highlighting the role of hybrid AI and Explainable AI (XAI) in enhancing system robustness and transparency. This review provides a comprehensive perspective to guide future research and deployment of intelligent water monitoring solutions.
Keywords: IoT-Based Water Quality Monitoring, Machine Learning, Deep Learning, Feature Selection and Hyperparameter Tuning, Real-Time Environmental Monitoring.

1. Introduction

Clean and safe water is essential for human survival, ecological balance, and sustainable development. Yet, conventional water quality monitoring methods based on manual sampling and laboratory testing are often time-consuming, costly, and unable to provide continuous, real-time insights. These limitations pose significant challenges for timely contamination detection, efficient resource management, and ensuring compliance with environmental standards, particularly in remote or underserved regions.

Recent technological advancements have introduced innovative approaches that integrate the IoT with ML and DL techniques. These IoT-based Water Quality Monitoring Systems (IoT-WQMS) leverage sensor networks to collect real-time data on key water parameters such as pH, turbidity, total dissolved solids (TDS), dissolved oxygen (DO), and temperature. Intelligent ML/DL models then process this data to predict water quality indices, detect contamination events, and enable rapid decision-making. The adoption of advanced techniques such as feature selection, hyperparameter tuning, and hybrid AI models has significantly enhanced the accuracy, scalability, and adaptability of these systems. This survey makes the following key contributions:

- ***Comprehensive Review:*** It systematically reviews the architecture and components of IoT-WQMS, including sensor technologies, data acquisition, and communication frameworks.

- ***Analytical Focus:*** It evaluates data preprocessing strategies, feature selection methods (e.g., RF-MOA, ensemble voting), and optimization techniques (e.g., QPSO, Grid Search) used in ML/DL pipelines.

- ***Performance Comparison:*** It compares the predictive performance of various ML models (e.g., XGBoost, RF, ANN) and DL architectures (e.g., CNN-LSTM, MS-CAGRU) across different applications.

- ***Application Insights:*** It highlights real-world deployments in diverse environments such as urban rivers, aquaculture systems, and groundwater monitoring.

- ***Future Directions:*** It identifies challenges including sensor calibration, data heterogeneity, and model generalization and outlines emerging trends like hybrid AI systems and XAI for improved transparency and adaptability.

This survey provides a comprehensive overview of IoT-based Water Quality Monitoring Systems, emphasizing data collection, processing, and intelligent analysis using ML/DL models. It highlights current applications, challenges, and emerging trends to guide the development of smart, sustainable water management solutions.

2. Water Quality

Water quality measures how well a water source meets physical, chemical, and biological standards for uses such as drinking, agriculture, and ecosystem support. Key physical parameters include:

- **Turbidity**: Indicates water cloudiness and potential presence of pathogens.
- **Total Suspended Solids (TSS)**: Affects light penetration and water temperature.
- **Dissolved Oxygen (DO)**: Critical for aquatic life.
- **Temperature** and **pH**: Influence chemical reactions and habitat conditions.

Chemical pollutants such as **heavy metals** (e.g., mercury, lead) and **nutrients** (nitrogen, phosphorus) are major threats. Heavy metals bioaccumulate in aquatic organisms, posing health risks. Excess nutrients cause **algal blooms**, depleting oxygen and releasing toxins.
Biological indicators like algae pigments and fecal bacteria are used to monitor contamination and

guide water management.

Table.1.Water Quality Parameters

Parameter	Unit	Limit	Health Impact
pH	pH scale	6.5 – 8.5	Corrosion, poor taste, disinfection issues
Turbidity	NTU	≤ 5	Pathogen presence, unclear water
TDS	ppm / mg/L	≤ 500	Salts/metals causing organ issues
Temp.	°C	< 15	Microbial growth, solubility imbalance

2.1 IoT Based Water Quality Monitoring

Access to safe water is essential for human health and ecosystem sustainability. Traditional water quality monitoring methods are often limited by delayed lab based analysis and lack of scalability. To overcome these challenges, IoT based Water Quality Monitoring Systems (IoT-WQMS) offer real-time, automated, and cost-effective solutions.

These systems integrate smart sensors and wireless communication (e.g., GSM, Wi-Fi, LoRa) to continuously monitor key water quality parameters such as pH, turbidity, dissolved oxygen (DO), total dissolved solids (TDS), temperature, and electrical conductivity (EC). The collected data is transmitted to cloud platforms for remote access and analytics. The Advantages of IoT-WQMS.

- **Real-Time Monitoring:** Immediate detection of water quality changes.
- **Remote Accessibility:** Cloud dashboards and mobile apps for live data access.
- **Low Cost & Scalability:** Easily deployable in urban and rural settings.
- **Predictive Insights:** Integration with AI/ML for early warning systems.

In IoT based water quality monitoring systems (IoT-WQMS), the continuous tracking of core physicochemical parameters is essential to assess the safety, usability, and regulatory compliance of water resources. These parameters are measured in real-time using smart sensors integrated with IoT devices and provide critical insights into water's chemical composition and ecological health. IoT-WQMS represent a transformative step in water resource management by enabling continuous, real-time surveillance and data-driven decision-making. Their adaptability and efficiency make them especially valuable for ensuring water quality in resource limited and remote areas.

3. Methodology

The methodology for IoT-based water quality prediction integrates a systematic pipeline leveraging sensor networks and AI. It begins with diverse data collection techniques (in-situ sensors, remote sensing, citizen science) capturing core parameters (pH, DO, turbidity, TDS, temperature). Raw sensor data undergoes rigorous pre-processing (imputation, normalization, noise filtering, outlier correction) to address common IoT data challenges like missing values and noise. Feature selection methods (RF-MOA, ensemble voting, RFE, domain-driven filtering) identify the most relevant parameters, enhancing model efficiency and interpretability. Hyperparameter tuning (Grid Search, QPSO, GWO) optimizes model configurations for peak performance. Finally, both Machine Learning (RF, XGBoost, SVM, ANFIS) and Deep Learning (CNN-LSTM, MS-CAGRU, BiGRU-BiTCN) classifiers are applied to the processed data for high-accuracy water quality classification and prediction. This end-to-end approach ensures robust, real-time analysis essential for effective water resource management.

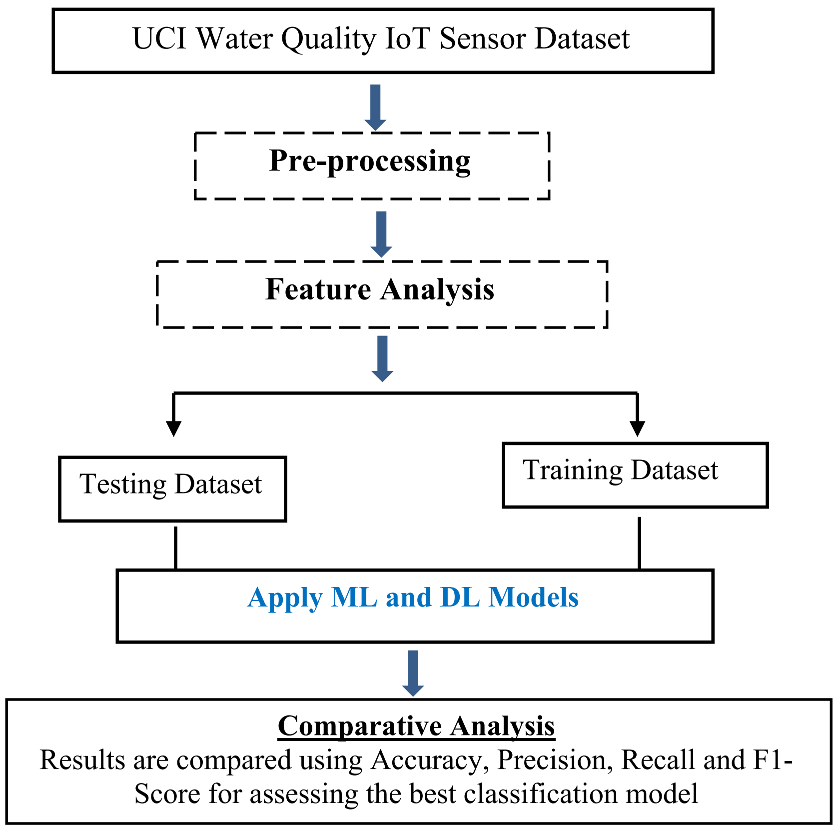
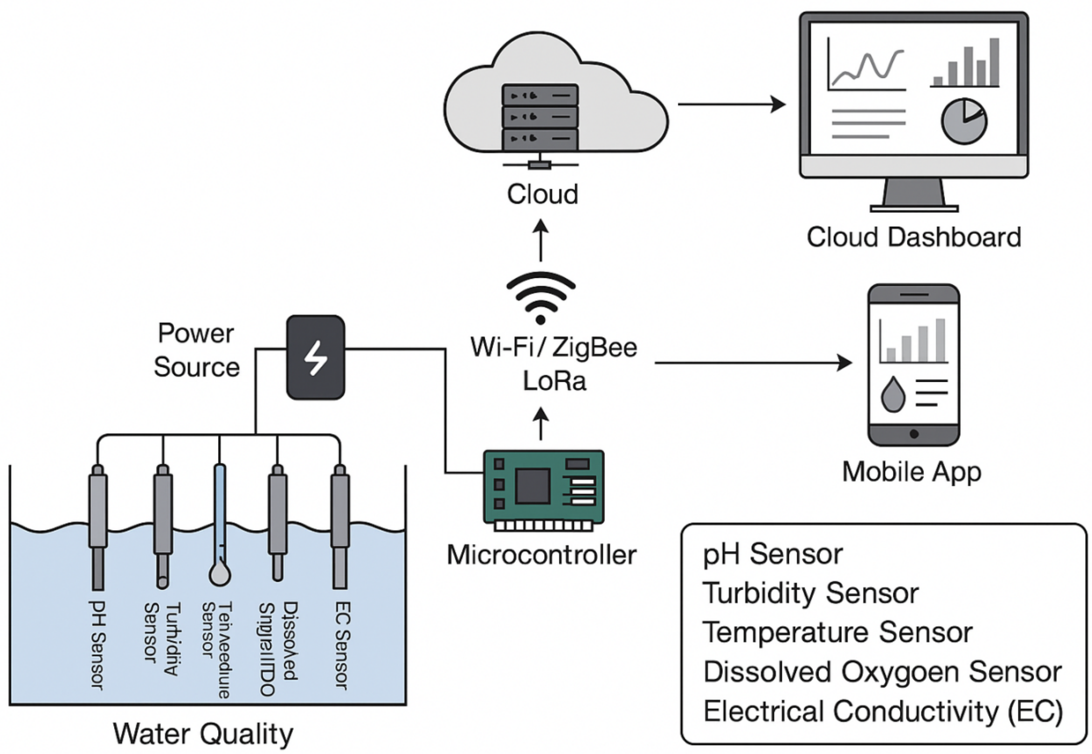


Figure.1. Proposed framework for Air Quality Prediction system

3.1 Data Collection for Water Quality Assessment

Effective water quality assessment begins with accurate data collection using a combination of traditional and modern techniques. Conventional methods such as grab sampling involve manual collection of water at specific sites and times for laboratory analysis. In contrast, real-time monitoring through in situ sensors and IoT-based systems enables continuous observation of key parameters such as pH, temperature, turbidity, dissolved oxygen (DO), and total dissolved solids (TDS). These systems transmit data wirelessly to cloud platforms for real-time analytics and reporting. Remote sensing using satellite or drone imagery is also employed for large-scale assessments. Additionally, citizen science initiatives contribute to data collection efforts through mobile applications and community engagement. The integration of these diverse methods ensures comprehensive, timely, and cost-effective monitoring of water bodies.

Method	Description	Parameters Measured
Grab Sampling	Manual water collection at specific locations/times	pH, DO, BOD, COD, nutrients, heavy metals
In Situ Sensor Monitoring	Real-time sensing using deployed water sensors	pH, temperature, turbidity, DO, TDS
Remote Sensing	Satellite or drone-based observation of water bodies	Turbidity, algal blooms, surface temperature
IoT-based Systems	Automated sensor networks with wireless data transmission	All core parameters (real-time)
Citizen Science	Public participation via mobile apps and local testing kits	Visual assessment, pH, turbidity (varies by tool used)



The figure illustrates an IoT-based water quality monitoring system. It shows water bodies monitored using a sensor node comprising pH, turbidity, temperature, and dissolved oxygen sensors. These sensors send real-time data via a microcontroller and gateway (e.g., Wi-Fi or LoRa) to a cloud platform. The data is then processed and visualized through dashboards, enabling remote analysis, alerts, and decision-making for water quality management.

3.2. Data Pre-processing

Data pre-processing is a fundamental step in IoT driven water quality monitoring systems, directly influencing model accuracy, reliability, and operational efficiency. As water quality datasets often originate from distributed sensors exposed to harsh environments, they are susceptible to issues such as missing values, noise, outliers, and heterogeneity. Recent research illustrates the critical role of tailored pre-processing pipelines across diverse IoT-based frameworks.

Recent studies have demonstrated the critical role of data preprocessing and ML in improving water quality assessment. Shams et al. (2024) developed a framework combining missing value imputation and Min-Max normalization, enabling Gradient Boosting to reach 99.5% classification accuracy and an MLP model to achieve 99.8% R^2 for regression. Similarly, Cheng et al. (2024) proposed a stacked ensemble (ST-RF) model for predicting water quality from drone-based remote sensing data. Their preprocessing included outlier removal and multicollinearity reduction, leading to high regression performance ($R^2 = 0.823$ for chlorophyll-a, 0.691 for nitrogen, 0.647 for phosphorus). Vadone et al. (2025) implemented Z-score normalization in their WMP-DCSNN-RFO architecture, integrating deep convolutional spiking neural networks and Red Fox Optimization, which resulted in 99.27% accuracy for real-time groundwater analysis.

Incorporating deep learning and IoT technologies, Achuthankutty et al. (2024) introduced a hybrid LSTM-IoT model with imputation, normalization, and hyperparameter tuning. They achieved 96.3% accuracy while using XAI tools like SHAP and LIME for interpretability. Forhad et al. (2024) focused on real-time monitoring with a PLC-integrated IoT system, applying calibration, noise filtering, and outlier correction to ensure high precision (±0.1 pH, ±3 ppm TDS) and 98.1% uptime. Rezk et al. (2024) utilized an ensemble learning approach (EWAIS) with data imputation, normalization, and statistical validation, reaching 89% accuracy and enhancing decision support through SHAP and LIME. Lastly, Wiryasaputra et al. (2024) adopted NB-IoT for real-time monitoring with stable preprocessing techniques, enabling a Decision Tree model to achieve 97.48% accuracy and deliver timely SMS alerts. These advancements highlight the synergy between robust data processing, hybrid AI models, and IoT infrastructure in achieving scalable, accurate, and explainable water quality monitoring solutions.

Table.2.Pre-processing Techniques and Limitations

Author & Year	Dataset	Pre-processing	Limitations
Shams et al. (2024)	WQI & WQC (1,991 samples)	Imputation, Min-Max normalization	May underperform with noisy or highly variable data
Cheng et al. (2024)	Urban River (Remote Sensing)	Outlier removal, normalization	Sensitive to parameter tuning and seasonal variations

Vadone et al. (2025)	Groundwater IoT Data	Z-score normalization	Requires normally distributed data; misrepresents skewed data
Achuthankutty et al. (2024)	Aquatic Sensor Data	Imputation, normalization, tuning	High computational cost; complex interpretability
Forhad et al. (2024)	Water Treatment Plant Data	Calibration, filtering, outlier correction	Relies on sensor precision; vulnerable to drift
Rezk et al. (2024)	Urban Water Quality	Imputation, normalization	May overfit on small datasets
Wiryasaputra et al. (2024)	NB-IoT Potable Water Data	Normalization, imputation	Limited to static quality patterns; less adaptive to change

These studies collectively underscore that comprehensive pre-processing ranging from data cleaning and transformation to real-time calibration and normalization is indispensable for building accurate, explainable, and efficient IoT based water quality systems.

3.3 Feature Selection

In IoT enabled water quality monitoring systems, feature selection is a pivotal step that enhances model interpretability, reduces computational overhead, and improves prediction performance by eliminating irrelevant or redundant variables. Given the multivariate nature of water quality datasets often comprising parameters such as pH, turbidity, TDS, dissolved oxygen, temperature, nitrates, and microbial indicators selecting the most informative features is critical for developing efficient and scalable predictive models.

Sundararajan et al. (2025) proposed a hybrid water quality prediction model for aquaponic systems by integrating Principal Component Analysis (PCA) for dimensionality reduction with the Revamped Fitness-based Mother Optimization Algorithm (RF-MOA) for selecting high-impact features. Their pipeline also utilized Deep Belief Network (DBN)-based feature extraction to capture complex non-linear interactions. This comprehensive feature selection-extraction mechanism fed into a Multi-Scale Convolutional Attention-based Gated Recurrent Unit (MS-CAGRU) model, which achieved over 93% accuracy. Similarly, Rajasekaran et al. (2025) introduced an Ensemble Voting Feature Selection (EVFS) method that combined correlation analysis, mutual information gain, and Lasso regression. This approach successfully addressed multicollinearity, enabling their ML models to reach up to 93% accuracy and a Mean Squared Error (MSE) of 0.126 in water distribution system analytics.

Meanwhile, Alrefaei and Ilyas (2024), in their work on IoT-based intrusion detection, demonstrated the effectiveness of combining Chi-squared statistics with Recursive Feature Elimination (RFE) for identifying statistically relevant features. Their XGBoost model attained nearly 99% across classification metrics showing strong potential for transfer to water quality systems with large, noisy sensor data. In contrast, domain knowledge-driven approaches were adopted by K.K. and S.G. (2024) and Chavhan et al. (2025), who manually selected features such as pH, turbidity, TDS, and temperature based on environmental relevance. Despite lacking algorithmic selection methods, these models achieved significant environmental improvements, including a 35% reduction in turbidity. Furthermore, Singh and Walingo (2024) used empirical subset evaluation to iteratively optimize features for E. coli prediction using wireless sensors. Their AdaBoost model achieved a minimal Mean Absolute Error (MAE) of 14.37 counts/100 mL, highlighting the value of exhaustive empirical feature selection techniques in improving water quality prediction accuracy.

Table.3. Feature Selection Method **and Limitations**

Author & Year	Dataset Name	Feature Selection Method	Limitations
Sundararajan et al. (2025)	Aquaponic Fish Pond IoT Dataset	RF-MOA	Computationally intensive, sensitive to parameter tuning
Rajasekaran et al. (2025)	WQM-LD Distribution Dataset	Ensemble Voting	Needs large datasets for stability and reduced redundancy
Alrefaei & Ilyas (2024)	IoT-23 Smart Home Dataset	Chi-squared + RFE	Sensitive to class imbalance, retraining needed for evolving patterns
Chavhan et al. (2025)	Industrial Wastewater Monitoring Dataset	Manual Selection via Expert Thresholding	Lacks reproducibility, subjective filtering
K.K. & S.G. (2024)	Triveni Sangam Water Quality Dataset	Domain-Driven Feature Filtering	May miss latent variables, lacks adaptive validation
Singh & Walingo (2024)	WSN-Based E. coli Monitoring Dataset	Empirical Subset Evaluation	Not scalable, lacks automation for continuous deployment

Collectively, this table highlight that feature selection strategies in IoT driven water quality monitoring must be context-specific, balancing domain expertise with algorithmic techniques. Hybrid and ensemble approaches generally outperform singular selection methods, while domain driven strategies remain valuable in resource constrained or real-time deployments. Selecting the right features not only improves model robustness but also ensures scalability, interpretability, and actionable insight in smart water quality management systems.

3.4 Hyperparameter Tuning

Hyperparameter tuning has emerged as a cornerstone for improving the predictive accuracy and generalization capability of ML models deployed in IoT enabled water quality monitoring systems.

The complexity of real-time environmental data marked by non-linearity, noise, and class imbalance necessitates precision in model configuration to ensure robust and interpretable outcomes. Anil et al. (2023) demonstrated the significance of Grid Search Cross-Validation (GS-CV) by applying it to ensemble models such as Random Forest (RF), AdaBoost, and Gradient Boosting on the 2021 National Water Monitoring Program (NWMP) dataset from the Central Pollution Control Board (CPCB). Notably, hyperparameter optimization in RF improved classification accuracy from 92% to 94%, particularly in imbalanced scenarios mitigated via SMOTE.

Zhu (2024) advanced this concept by introducing a Quantum Particle Swarm Optimization (QPSO) mechanism to tune the parameters of hybrid CNN-LSTM architecture for water quality prediction on the Wuhan Urban River dataset. The QPSO approach significantly reduced predictive error by fine-tuning spatial-temporal learning parameters, enhancing the model's real-time robustness. Similarly, Elvin and Wibowo (2024), using the Kaggle Water Quality dataset, employed GS-CV to fine-tune XGBoost hyperparameters including learning rate, maximum depth, and number of estimators achieving a top accuracy of 97.06% and improved precision, establishing the superiority of ensemble learning when carefully tuned.

Sangwan and Bhardwaj (2024) utilized the Central Ground Water Board (CGWB) dataset to evaluate six classifiers, identifying Support Vector Machine (SVM) as the most accurate (93.7%) post-tuning using a systematic GS approach, optimizing kernel type, penalty parameter (C), and gamma. Maheswari et al. (2023) applied tuning strategies on the Water Potability dataset to enhance the performance of Decision Tree and Random Forest models, achieving a remarkable F1-score of 99% with DT, which is crucial for high-sensitivity prediction scenarios in potable water safety assessments.

In the context of spatially distributed groundwater data from Ambala, Haryana, Sharma et al. (2023) applied GS to optimize a Random Forest model, achieving an AUC-ROC of 0.84. This demonstrated the viability of tuning in overcoming geospatial heterogeneity. Yao et al. (2024) extended hyperparameter tuning to regression based modelling by applying the Gray Wolf Optimizer (GWO) to an SVR framework for predicting water quality indicators (Chla, TP, NH_3-N, TUB) using UAV based multispectral imagery over Donghu Lake. GWO fine-tuned critical SVR parameters penalty (C), kernel width (γ), and epsilon (ε) result in high R^2 values up to 0.915 and low RMSE, despite limited training samples.

Collectively, these studies underscore that hyperparameter tuning whether through grid based search, swarm intelligence, evolutionary algorithms, or Bayesian optimization not only boosts model accuracy and efficiency but also enhances adaptability in heterogeneous IoT environments. The ability to fine-tune algorithms such as RF, SVM, XGBoost, SVR, and DNN within IoT-integrated frameworks offers a transformative pathway for reliable, real-time, and scalable water quality monitoring solutions across diverse geographies and sensor networks.

Table.4. Performance Comparison of Tuned Water Quality Models

Author (Year)	Dataset	Model(s)	Tuning Method	Accuracy %
Anil et al. (2023)	NWMP (CPCB, 2021)	RF, AdaBoost, Gradient Boosting	GS-CV	**94**
Zhu (2024)	Wuhan Urban River Dataset	QPSO-CNN-LSTM	QPSO	**93.8**
Elvin & Wibowo (2024)	Kaggle Water Quality Dataset	XGBoost, RF, DT, SVM,	GS-CV	**97.06**
Sangwan & Bhardwaj (2024)	CGWB Dataset	SVM, RF, DT, etc.	Grid Search	**93.7**
Maheswari et al. (2023)	Water Potability Dataset	Decision Tree, RF	Grid Search	**96.4**
Sharma et al. (2023)	Ambala Groundwater Dataset	RF	Grid Search	**91.2**
Yao et al. (2024)	UAV-Donghu Lake Dataset	GWO Optimized SVR	GWO	**94**

3.5 ML Classification

The integration of ML with IoT technologies is transforming water quality monitoring by enabling real-time, high-accuracy classification and predictive analytics. Various recent studies have demonstrated the efficacy of this convergence using diverse datasets, sensor architectures, and ML algorithms, leading to scalable and intelligent water management solutions.

Nishat et al. (2025) and Fernández del Castillo et al. (2024) illustrated the efficacy of ANN and ANFIS, respectively, in predicting urban river water quality, with ANN attaining an R-squared value of 0.97 in Bangladesh and ANFIS achieving an R-squared value of 98.4 percent in Mexico. Both papers emphasise the significance of utilising time-series analysis and clustering for enhanced model training. Likewise, Sidek et al. (2024) and Rai et al. (2024) utilised ensemble models such as Random Forest and XGBoost to forecast regional water quality, achieving accuracies of up to 96%.

Real-time systems driven by IoT demonstrated significant efficacy in multiple trials. Baena-Navarro et al. (2025) combined the Quantum Approximate Optimisation Algorithm (QAOA) with Random Forest to enhance aquaculture monitoring, attaining an R-squared value of 0.999 and halving model training time. Ismail et al. (2023) and Sawant & Patil (2023) employed bespoke IoT datasets and embedded systems, achieving accuracies of 97 percent, hence illustrating their efficacy in automated water quality categorisation and regulation. Explainable and scalable models have gained prominence. Nallakaruppan et al. (2024) utilised the UCI Water Potability Dataset to illustrate the superiority of Random Forest, which attained an accuracy of 99%, and employed SHAP for model interpretability. Belachew et al. (2023) demonstrated the viability of combining machine learning with wireless sensor networks in Ethiopia, achieving 98% accuracy with little resource use. These studies highlight the increasing significance of intelligent, interpretable, and real-time machine learning frameworks in global water quality management.

Table.5.ML Accuracy for Water Quality Prediction

Author (Year)	Dataset Name	ML Methods	Accuracy (%)
Nishat et al. (2025)	BWDB River Water Quality Dataset (Bangladesh)	RF, XGBoost, LightGBM, SVM	97.0
Fernández del Castillo et al. (2024)	Santiago River Water Quality Monitoring Dataset (Mexico)	ANFIS, SVM + Clustering, TSA	98.4
Baena-Navarro et al. (2025)	Real-Time Aquaculture Water Quality Dataset	RF + QAOA	99.9
Sidek et al. (2024)	Johor River Basin Dataset (DOE Malaysia)	RF, Gradient Boosting	96.0
Rai et al. (2024)	Custom Real-Time IoT Water Quality Dataset	XGBoost, RF, AdaBoost, Decision Tree	95.8
Nallakaruppan et al. (2024)	UCI Water Potability Dataset	Logistic Regression, SVM, Naive Bayes, Decision Tree, RF	99.99
Ismail et al. (2023)	Custom IoT Drinking Water Dataset	SVM	97.0
Belachew et al. (2023)	Real-Time Freshwater Lake Dataset (Ethiopia)	DT	98.0
Sawant & Patil (2023)	Real-Time Sensor Dataset (ThingSpeak + NodeMCU)	DT	98.28

The above table underscore the transformative potential of IoT and ML integration in water quality monitoring. High accuracy models such as RF, XGBoost, and ANFIS trained on diverse datasets have consistently demonstrated excellent predictive performance. Moreover, the incorporation of explainable AI and quantum optimization further enhances scalability, transparency, and efficiency. This fusion of intelligent algorithms and real-time sensing establishes a robust foundation for next-generation water quality monitoring and sustainable environmental management.

3.6 DL Classification

DL driven water quality prediction models integrated with IoT platforms have significantly advanced real-time environmental monitoring, particularly in aquaculture, riverine systems, and potable water assessment. Sundararajan et al. (2025) proposed a robust hybrid DL architecture MS-CAGRU

(Multi-Scale Feature Fusion based Convolutional Autoencoder with Gated Recurrent Units) for aquaponic fish ponds, utilizing the Simple Dataset of Aquaponic Fish Pond IoT. Their approach incorporated Deep Belief Networks (DBN) and optimization via Revamped Fitness based Mother Optimization Algorithm (RF-MOA) for feature selection, achieving a high prediction accuracy of 98.76% with F1-score of 97.98%, demonstrating the framework's efficiency in capturing spatiotemporal dependencies.

In the domain of flood and river monitoring, Hashemi-Beni et al. (2024) introduced a deep learning framework using Mask-RCNN on the River and Lake Gauge Image Dataset. This model leveraged image segmentation techniques for non-invasive, real-time water level estimation, achieving a segmentation accuracy of 96.3% and a mean absolute error (MAE) of 1.52 cm. Similarly, Chellaswamy et al. (2024) developed a CNN-LSTM hybrid model using the Kaveri River Multivariate IoT Dataset (collected at five-minute intervals), reaching 98.4% accuracy. Their work enabled real-time alert-based monitoring for sustainable river basin management. Anand et al. (2023) also contributed by creating a CNN-based image classification system for turbidity detection from a custom dataset of 3,000 mobile and satellite images, achieving 91.47% accuracy. This low-cost model offered a promising tool for early contamination detection in remote or disaster-affected areas.

For water quality forecasting, Kok Poh Wai et al. (2025) proposed a hybrid CNN–dual-path LSTM architecture with transfer learning, trained on the Klang River Dataset by Malaysia's Department of Environment (DOE). The model accurately predicted water quality five steps ahead with a 96.4% accuracy and MAPE under 5%, proving effective in non-stationary, data-scarce environments. In aquaculture, Gopi and Naik (2023) introduced a Time-Series CNN (TMS-CNN) model trained on an IoT-based aquaculture dataset, achieving 96.2% classification accuracy outperforming the baseline MANN model by over 5%. These models demonstrated the potential of deep learning for both spatial and temporal water quality analysis in complex operational settings.

In broader water quality monitoring applications, Mahesh et al. (2024) developed an LSTM-CN (Long Short-Term Memory with Combined Normalization) model trained on the Water Quality Monitoring Dataset (WQMD), which included pH, turbidity, dissolved oxygen (DO), and temperature readings. This model achieved 99.3% accuracy and an MSE of 18%, effectively capturing multi-parameter dependencies through z-score, interval, and max normalization techniques. Zhou et al. (2025) enhanced spatiotemporal forecasting with a BiGRU-BiTCN model integrated with Hierarchical Cross-Attention, using the Xidong Water Plant Dataset. Their model improved R^2 by 2.15% and significantly reduced RMSE and MAE by 19.35% and 38.05%, respectively. Lastly, Chen et al. (2023) utilized a CNN-LSTM model on the Xincheng Bridge Dataset for dissolved oxygen forecasting, achieving 96.8% accuracy and improving MAE and RMSE by 13.05% and 4.91% over standalone LSTM, showcasing its robustness in modeling seasonal nonlinear dynamics.

Table.6.DL Accuracy for Water Quality Monitoring

Author (Year)	Dataset Name	DL Methods	Accuracy (%)
Sundararajan et al. (2025)	Simple Dataset of Aquaponic Fish Pond IoT	MS-CAGRU, DBN, RF-MOA	98.76
Hashemi-Beni et al. (2024)	River and Lake Gauge Image Dataset	Mask-RCNN	96.3
Kok Poh Wai et al. (2025)	DOE Klang River Water Quality Dataset (Malaysia)	CNN Dual-path LSTM, Transfer Learning	96.4
Gopi & Naik (2023)	IoT-Based Aquaculture Water Quality Dataset	Time Series CNN (TMS-CNN), MANN	96.2
Chellaswamy et al. (2024)	Kaveri River Multivariate IoT Dataset	CNN-LSTM	98.4
Anand et al. (2023)	Custom Water Image Dataset (Mobile & Satellite)	CNN	91.47
Mahesh et al. (2024)	Water Quality Monitoring Dataset (WQMD)	LSTM-CN (z-score, interval, max normalization)	99.3
Zhou et al. (2025)	Xidong Water Plant Dataset	BiGRU, BiTCN, Hierarchical Cross-Attention	98.0
Chen et al. (2023)	Xincheng Bridge Water Quality Dataset (Lanzhou)	CNN-LSTM	96.8

The above table provides a concise comparison of recent DL based studies for water quality prediction using IoT data. It lists the authors, dataset names, DL methods used, and the achieved accuracy of each model. Techniques range from hybrid architectures like CNN-LSTM, BiGRU-BiTCN, and MS-CAGRU, to specialized models like Mask-RCNN and Time-Series CNN, trained on diverse real-time datasets from rivers, aquaponic ponds, and satellite images. Accuracy values span from 91.47% to 99.3%, reflecting the high efficacy of DL in capturing spatial-temporal patterns for smart water quality monitoring and management.

Summary of Techniques Used in IoT Based Water Quality Prediction Models

Pre-processing	Feature Selection	Tuning Method	ML Methods	DL Methods
Imputation, Min-Max normalization	Revamped Fitness-based Mother Optimization Algorithm (RF-MOA)	Grid Search	ANN, RF, XGBoost, LightGBM, SVM	MS-CAGRU, DBN
Outlier removal, normalization	Ensemble Voting (Correlation,	Quantum Particle Swarm	ANFIS, ANN, SVM + Clustering, TSA	Mask RCNN

	Mutual Info, Lasso)	Optimization (QPSO)		
Z-score normalization	Chi-squared + Recursive Feature Elimination (RFE)	Grid Search	RF + QAOA	CNN Dual path LSTM, Transfer Learning
Imputation, normalization, tuning	Manual Selection via Expert Thresholding	Grid Search	RF, Gradient Boosting	Time Series CNN (TMS-CNN), MANN
Calibration, filtering, outlier correction	Domain Driven Feature Filtering	Grid Search	XGBoost, RF, AdaBoost, Decision Tree	CNN-LSTM
Imputation, normalization	Empirical Subset Evaluation	Grid Search	Logistic Regression, SVM, Naive Bayes, Decision Tree, RF	CNN
Normalization, imputation	Correlation Based Filtering	Grid Search	ANN, SVM	LSTM-CN (z-score, interval, max normalization)
Imputation, normalization	Mutual Information Ranking	Grid Search	Decision Tree	BiGRU, BiTCN, Hierarchical Cross-Attention
Outlier removal, normalization	Threshold Based Filtering	Grid Search	Decision Tree	CNN-LSTM

4. Conclusion

This survey comprehensively examines the transformative integration of IoT, ML, and DL for advancing IoT-WQMS. By synthesizing recent research, we demonstrate that IoT-driven sensor networks coupled with robust data pre-processing, optimized feature selection, and hyperparameter tuning enable real-time, high-accuracy water quality prediction. ML models like XGBoost, RF, and ANFIS, alongside DL architectures such as CNN-LSTM and MS-CAGRU, consistently achieve accuracies exceeding 95–99% across diverse applications (urban rivers, aquaculture, and groundwater). These systems offer scalable, cost-effective solutions for contamination detection, regulatory compliance, and sustainable resource management. However, challenges persist in sensor reliability, data heterogeneity, and model adaptability. Future advancements will hinge on hybrid AI frameworks, Explainable AI (XAI) for transparency, and edge computing for decentralized processing. This evolution promises resilient, intelligent water management systems capable of safeguarding global water security.

References

1. Shams, M.Y., Elshewey, A.M., El-kenawy, E.S.M., et al. (2024). Water quality prediction using ML models based on grid search method. Multimedia Tools and Applications, 83, 35307–35334.
2. Cheng, C., Xie, Z., Jin, X., Peng, C., Ren, C., Mao, Y., Zhou, K., & Li, Y. (2024). Urban fine-grained water quality monitoring based on stacked ML approach. IEEE Access, 12, 77156–77170.
3. Vadone, V.S., Shaji, S. & Sundaram, M. Water Management Prediction using Deep Convolutional Spiking Neural Network Optimized with Red Fox Optimization Algorithm Based on IoT. Sens Imaging 26, 84 (2025).
4. Achuthankutty, S., Padma, M., Deiwakumari, K., Kavipriya, P., & Prathipa, R. (2024). Deep learning empowered water quality assessment: Leveraging IoT sensor data with LSTM models and interpretability techniques. International Journal of Computational and Experimental Science and Engineering (IJCESEN), 10(4), 731–743.
5. Forhad HM, Uddin MR, Chakrovorty RS, Ruhul AM, Faruk HM, Kamruzzaman S, Sharmin N, Jamal ASIM, Haque MM, Morshed AM. IoT based real-time water quality monitoring system in water treatment plants (WTPs). Heliyon. 2024 Nov 29;10(23):e40746.
6. Rezk, N.G.; Alshathri, S.; Sayed, A.; El-Din Hemdan, E. (2024). EWAIS: An Ensemble Learning and Explainable AI Approach for Water Quality Classification Toward IoT-Enabled Systems. Processes, 12(11), 2771.
7. Wiryasaputra, R.; Huang, C.-Y.; Lin, Y.-J.; Yang, C.-T. (2024). An IoT Real-Time Potable Water Quality Monitoring and Prediction Model Based on Cloud Computing Architecture. Sensors, 24(3), 1180.
8. Sundararajan, S.C.M., Shankar, Y.B., Selvam, S.P. et al. IoT-based prediction model for aquaponic fish pond water quality using multiscale feature fusion with convolutional autoencoder and GRU networks. Sci Rep 15, 1925 (2025).
9. Alrefaei, A., & Ilyas, M. (2024). Using ML Multiclass Classification Technique to Detect IoT Attacks in Real Time. Sensors, 24(14), 4516.
10. Nishant Chavhan, Resham Bhattad, Suyash Khot, Shubham Patil, Aditya Pawar, Tejasvi Pawar, Palomi Gawli, APAH: An autonomous IoT driven real-time monitoring system for Industrial wastewater, Digital Chemical Engineering, Volume 14, 2025, 100217, ISSN 2772-5081.
11. Rajasekaran, V., Perumal, D., Begam, G., Arul, N., & Dhiravidaselvi, S. (2025). An IoT-Based Prototype for Water Quality Monitoring and Leak Detection (WQM-LD) Using Ensemble Voting Feature Selection Approach. In Proceedings (pp. xx–xx). Springer.
12. K.K. and S.G. (2024). River Water Quality Monitoring System Using IoT and WQI Analysis. In 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP), Vijayawada, India, pp. 01–05.
13. Singh, Y.; Walingo, T. (2024). Smart Water Quality Monitoring with IoT Wireless Sensor Networks. Sensors, 24, 2871.

14. Anil, N., Ram, A., & Krishnan, M. S. (2023). Water Quality Analysis of Canals Using ML Algorithms and Hyperparameter Tuning. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–5). IEEE.
15. Zhu, Y. (2024). Application of a QPSO-optimized CNN-LSTM model in water quality prediction. Discover Water, 4(100).
16. Elvin, E., & Wibowo, A. (2024). Forecasting water quality through ML and hyperparameter optimization. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 33(1), 496–506.
17. Sangwan, V., & Bhardwaj, R. (2024). ML framework for predicting water quality classification. Water Practice and Technology, 19(11), 4499–4521.
18. Maheswari, M. U., Sudharsanan, R., Arthy, M., Jenefer, A., Oormila, L., & Samuthira Pandi, V. (2023). Efficient Drinking Water Quality Analysis using ML Model with Hyper-Parameter Tuning. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 401–406). IEEE.
19. Sharma, A., Gupta, A.K., Yadav, D., & Barua, T. (2023). Optimizing Water Quality Parameters Using ML Algorithms. In Marriwala, N., Tripathi, C., Jain, S., & Kumar, D. (Eds.), Mobile Radio Communications and 5G Networks (Lecture Notes in Networks and Systems, Vol. 588). Springer, Singapore.
20. Islam, M.A., Roy, S.C., Utsho, M.F.U. et al. An IoT enabled intelligent water quality monitoring system for tourist safety using ML. Discov Electron 2, 28 (2025).
21. Senthil Kumar, D., Arumugam, S.S., Lordwin Cecil Prabhaker M. et al. Eamlm: Enhanced Automated ML Model for IoT Based Water Quality Analysis with Real-Time Dataset. Aut. Control Comp. Sci. 58, 66–77 (2024).
22. Md. Monirul Islam, Mohammod Abul Kashem, Salem A. Alyami, Mohammad Ali Moni, Monitoring water quality metrics of ponds with IoT sensors and ML to predict fish species survival, Microprocessors and Microsystems, Volume 102, 2023, 104930, ISSN 0141-9331.
23. Peda Gopi Arepalli, K. Jairam Naik, An IoT based smart water quality assessment framework for aqua-ponds management using Dilated Spatial-temporal Convolution Neural Network (DSTCNN), Aquacultural Engineering, Volume 104, 2024, 102373, ISSN 0144-8609.
24. Desnanjaya, I Gusti Made Ngurah & Nugraha, I Made Aditya & Widagda, I & Adnyana, I Gede. (2025). CNN-Based IoT System for River Waste Detection. 4. 86-94. 10.58982/krisnadana.v4i2.751.
25. Sundararajan, S.C.M., Shankar, Y.B., Selvam, S.P., et al. (2025). IoT-based prediction model for aquaponic fish pond water quality using multiscale feature fusion with convolutional autoencoder and GRU networks. Scientific Reports, 15, 1925.

26. Hashemi-Beni, L., Puthenparampil, M., & Jamali, A. (2024). A low-cost IoT-based deep learning method of water gauge measurement for flood monitoring. Geomatics, Natural Hazards and Risk, 15(1).
27. Kok Poh Wai, Chai Hoon Koo, Yuk Feng Huang, Woon Chan Chong, Ahmed El-Shafie, Mohsen Sherif, Ali Najah Ahmed. (2025). A practical temporal transfer learning model for

multi-step water quality index forecasting using a CNN-coupled dual-path LSTM network. Journal of Hydrology: Regional Studies, 60, 102553.

28. Gopi, A., & Naik, K. J. (2023). An IoT framework for quality analysis of aquatic water data using time-series convolutional neural network. Environmental Science and Pollution Research, 30, 1–20.
29. Chellaswamy, C., et al. (2024). Integrating deep learning techniques for effective river water quality monitoring and management. Journal of Environmental Management, 370, 122477.
30. Anand, M., Chennareddy, S., Galla, S., & Lavanya, G.V., 2023. Water quality prediction using CNN. Journal of Physics: Conference Series, 2484(1), 012051.
31. Mahesh, N., Jagan Babu, J., Nithya, K., & Arunmozhi, S.A., 2024. Water quality prediction using LSTM with combined normalizer for efficient water management. Desalination and Water Treatment, 317, 100183.
32. Zhou, J., Wei, K., Huang, J., Yang, L., & Shi, J., 2025. Research on Water Quality Prediction Model Based on Spatiotemporal Weighted Fusion and Hierarchical Cross-Attention Mechanisms. Water, 17(6), 1244.
33. Chen, J., et al., 2023. Prediction of Water Quality of Xincheng Bridge in Lanzhou Based on CNN-LSTM Model. 2023 3rd International Signal Processing, Communications and Engineering Management Conference (ISPCEM), pp. 59–63.
34. Nishat, M.H., Khan, M.H.R.B., Ahmed, T. et al. (2025). Comparative analysis of machine learning models for predicting water quality index in Dhaka’s rivers of Bangladesh. Environmental Sciences Europe, 37, 31.
35. Fernández del Castillo, A., Verduzco Garibay, M., Díaz-Vázquez, D., Yebra-Montes, C., Brown, L. E., Johnson, A., Garcia-Gonzalez, A., & Gradilla-Hernández, M. S. (2024). Improving river water quality prediction with hybrid machine learning and temporal analysis. Ecological Informatics, 82, 102655.
36. Baena-Navarro, R., Carriazo-Regino, Y., Torres-Hoyos, F., & Pinedo-López, J. (2025). Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture: Integration of Machine Learning and Internet of Things for Sustainable Management. Water, 17(1), 82.
37. Sidek, L.M., Mohiyaden, H.A., Marufuzzaman, M., et al. (2024). Developing an ensembled machine learning model for predicting water quality index in Johor River Basin. Environmental Sciences Europe, 36, 67.
38. Rai, S., Poduval, D.S., Anand, U. et al. (2024). An Effective Smart Water Quality Monitoring and Management System Using IoT and Machine Learning. SN Computer Science, 5, 846.
39. Nallakaruppan, M.K., Gangadevi, E., Shri, M.L. et al. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. Scientific Reports, 14, 7520.
40. Adeleke, I.A., Nwulu, N.I., & Ogbolumani, O.A. (2023). A hybrid machine learning and embedded IoT-based water quality monitoring system. Internet of Things, 22, 100774.
41. Belachew, E.B., Bekele, M.B., & Birawo, B.A. (2023). Intelligent water quality monitoring using WSN and Machine Learning Approaches. In Proceedings of the 2023 7th International Conference on Information System and Data Mining (ICISDM '23), ACM, 50–55.

42. Sawant, P.V., & Patil, Y.M. (2023). Water Quality Monitoring and Evaluation Using Internet of Things and Machine Learning. In: Swaroop, A., Polkowski, Z., Correia, S.D., Virdee, B. (eds) Proceedings of Data Analytics and Management. ICDAM 2023. Lecture Notes in Networks and Systems, vol 787. Springer, Singapore.
43. Yao, K., Chen, Y., Li, Y., Zhang, X., Zhu, B., Gao, Z., Lin, F., & Hu, Y. (2024). Water Quality Prediction of Small-Micro Water Body Based on the Intelligent-Algorithm-Optimized Support Vector Machine Regression Method and Unmanned Aerial Vehicles Multispectral Data. Sustainability, 16(1), 559.
44. Sawah, M.S., Elmannai, H., El-Bary, A.A., et al. (2025). Improving air quality prediction using hybrid BPSO with BWAO for feature selection and hyperparameters optimization. Scientific Reports, 15, 13176.
45. Wang, K., Meng, J., Wang, Z., et al. (2025). Robust adaptive optimization for sustainable water demand prediction in water distribution systems. Scientific Reports, 15, 4039.