

A Review of Machine Learning Techniques for Water Quality Assessment and Prediction

¹ Ratika Aditya Bhegade, ² Nagesh Raikar, ³ Prashant Kumbharkar

¹ Ajeenkya DY Patil University, Pune, Maharashtra, India

² Shri Jagdishprasad Jambharmal Timbrawala University, Jhunjhun, India

³ Dean and Professor, Ajeenkya DY Patil University, Pune, Maharashtra, India

ABSTRACT

It is essential to forecast and examine water condition for public health. Because pollution is increasing, we need water condition prediction (WCP) regularly. Safe water is necessary for consumption. Traditional laboratory-oriented testing approach is accurate but slow and costly. Nowadays, machine learning (ML) and deep learning (DL) models have been used to water condition prediction (WCP) and classify water quality indices (WQI). These models have also been used for monitoring water in real time using Internet of Things (IoT) infrastructure by learning complex, non-linear relationships in data. We searched for papers on Google Scholar, IEEE Xplore, ScienceDirect, and MDPI. We used the keywords "water quality prediction", "machine learning", "water quality index", and "IoT water monitoring" to search for papers. This narrative review synthesizes the literature published between 2020 and 2026. It critically reviews machine learning (ML) and deep learning (DL) models for checking water condition, calculating water quality indices, and synthesizing real-time water condition checking using IoT.

Keywords: water quality prediction; water quality assessment; machine learning; deep learning; IoT monitoring; explainable AI

INTRODUCTION

Tracking the water condition is essential for human health, economic development, and agricultural productivity. Therefore, tracking the water condition is essential for environmental management and safe potable water for consumption [1]. Because physical, chemical, biological conditions change dynamically, regular tracking and examination are essential to maintain water condition [1][2]. In India, rapid urbanisation, industrialisation, and contamination of surface and groundwater sources have created a significant challenge in providing safe potable water through municipal supplies [3].

Traditionally, we require manual collection and laboratory testing for water condition checking. This

approach is time-consuming and costly due to manpower requirements [4]. There is a need for well-trained people to perform tests to calculate water quality parameters [5]. ML is increasingly used for water condition prediction because it can learn non-linear relationships between variables and provide target outcomes, such as water quality classes or water quality index values [1].

ML is used to calculate the WQI. Using an ML algorithm, we can classify water condition into excellent, good, and poor categories using physicochemical parameters; ML, which was used to predict the water quality index [6]. ML can use different models such as random forests, SVM, XGBoost, ANFIS, and neural networks to assess water condition [7]. The prediction of chemical amounts, WQI, and water classes are different tasks. High accuracy on one dataset does not guarantee that it will work everywhere [8][9].

DL models perform best when data are complete. This is useful when data vary with time and location. To study maps and satellite images, CNN models give the best results. LSTM and GRU are helpful to study how water condition changes as time change [10]. Combining CNN and LSTM gives better results for checking water condition. This model requires continuous sensor data and high computing power [11] [12][13]. When data are limited, incomplete, or have missing values and seasonal changes, simple methods can work similarly to complex methods [14]. Therefore, DL models require validation and careful testing with enough data and computing power before deployment [15].

We can combine ML and DL models with IoT, which can allow automated and continuous water condition surveillance [16]. The IoT-based systems have a microcontroller such as Arduino and an ESP32 connected with sensors for measuring different parameters. They utilized of various wireless communication protocols, to transfer physicochemical information to remote cloud and edge servers [16][17]. Hence, the robustness of the live monitoring system is dependent on the integrity of the sensor and communication protocol [1][18][19].

There are several limitations present. Water quality parameters (WQP) change over time and location, so the performance of the model also changes. In public datasets, data is collected from a particular place at a particular time, so results may vary for another place. Public datasets may not be able to show regional conditions. Further issue is many studies use different WQI formulations, WQP, and performance evaluation metrics, so comparing direct results is difficult.

In many studies, only physicochemical parameters are used to test the model. Limited studies have been conducted for chemical and biological parameters. This review paper summarize recent ML models applied for water condition analysis and also compare the datasets, input factors, and result evaluation used in existing studies to identify limitations.

Research Objective

RO1: To study and compare ML and DL methods used for water condition prediction.

RO2: To check the datasets, physicochemical parameters, data cleaning steps, and testing methods used in other studies.

RO3: To check IoT-based live tracking systems and determine if they provide live prediction, grouping, or rule-based checking.

Conventional water quality assessment

Continuously tracking the water condition is essential for human health. Changes in water can make it unsafe for consumption. Traditionally water condition is checked by manual sample collection from time to time and testing them in a lab. Lab tests give correct results but are time consuming and costly. It is not possible to check water condition all the time using traditional lab test methods [19][20]. Due to this ML is best to check water condition.

Water Quality Index (WQI)

Water quality is a highly dynamic and complex as it varies with space and time. Large datasets were analysed, which are complex to interpret. It is a mathematical techniques or composite index that aggregates multiple WQP into a single digit ranging between 0 to 100 [21].

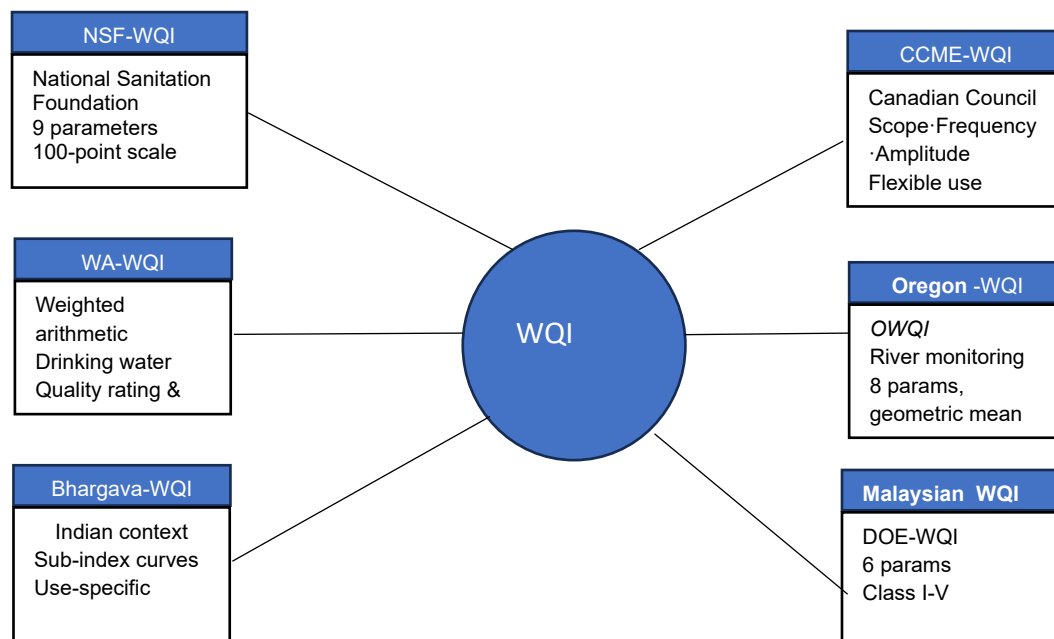


Fig1 Types of Water Quality Index (WQI)

Water Quality Parameters

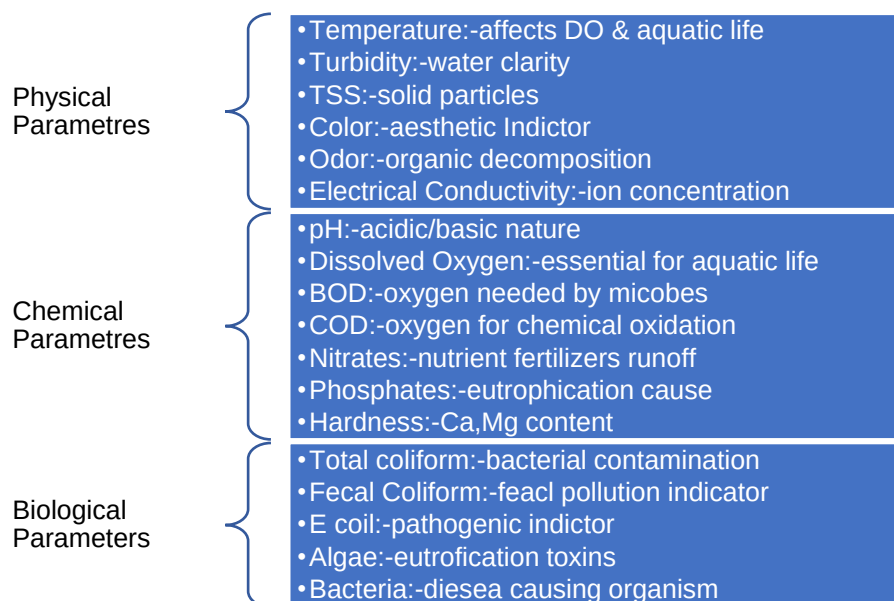


Fig 2 Water Quality Parameters

Literature Review

Zhu et al. [22] systematically reviewed 45 ML algorithms in diverse aquatic environments and concluded that ML effectively address complex nonlinear relationships.

Yan et al. [23] reviewed more than 170 papers for organising existing studies into predicting individual water-quality indicators and WQI. These studies show that data-driven approaches are important for water-quality research.

Hassan et al. [24] applied ML models on datasets collected from multiple sites in India. They considered general parameters (pH, electrical conductivity), organic pollution indicators (DO, BOD), nutrient (nitrate), and microbiological indicator (total coliforms) for the study. The results demonstrated that MLR achieved high performance.

Prasad et al. [25] investigated water-condition projection for Korattur Lake in Chennai, India using supervised ML algorithm. RF achieved approximately 95% accuracy, and it is also cost-effective. This study is relevant to Indian water-condition studies because they use commonly used physicochemical parameters for automated WQC.

Nasir et al. [26] developed a seven-WQC model and reported that stacking-based techniques were effective and CatBoost produced the best predictive performance in their study.

Nguyen Khoi et al. [27] compared 12 ML models and also estimate the WQI for River water, including boosting algorithms, tree-based and ANN-based approaches.

Shams et al. [28] used Exhaustive grid search with ensemble (RF, XGBoost, Gradient Boosting, AdaBoost), single (DT), instance-based (KNN), kernel-based (SVR), and ANN-based (MLP) for WQI and WQC. The 99.5% accuracy was obtained by the optimized GBM model for WQC.

C.Cheng et al. [29] developed a water condition tracking model with stacked-based ML. By use of drone-based multispectral imagery with high spatial resolution, they proposed a robust stacked Random Forest model (ST-RF) model for checking eutrophication indicators. They introduced a novel WQI formulated using Pearson correlation and grey relational analysis. The ST-RF model obtained an accuracy of 71%.

Pandya et al. [30]. proposed an advanced ensemble model by combining the strengths of different ML models predicting water potability. They developed a dashboard with forecasting and visualizing live input.

Chen et al. [31] reviewed ANN approaches and showed their wide application in water-condition prediction. They also investigated that LSTM is best because it can learn temporal dependencies when water condition data is sequential.

Li et al. [32] investigated LSTM-based water-condition forecasting for River water.

Zhang et al. [33] combined CEEMDAN with LSTM to forecast water condition in an urban reservoir. These studies suggest that time-dependent ML models may be useful when continuous historical water-quality data are available.

Q.Zhu et al. [34] proposed combination of CEEMDAN and CNN-LSTM enhanced by an attention mechanism for water condition projection. The authors collected data on physical parameter (turbidity) and chemical indicators (DO, pH, CODMn) in water condition checking and achieved high accuracy, strong generalisation, and robustness.

Tan et al. [35] integrated the CNN with LSTM model in which CNN learn relevant patterns and LSTM access temporal relationships. Their results showed improvements over conventional LSTM models in several error measures.

Haq et al. [10] proposed hybrid deep learning design by combining CNN with LSTM and GRU separately for water condition forecasting in smart aquaculture, wherein the combination of CNN with LSTM gives higher prediction accuracy and computational performance.

Yan et al. [36] proposed smart system with deep reinforcement learning and GCN to capture interdependencies among WQP for water condition prediction, for better water management and pollution control. The studies findings that hybrid DL architectures fcan be useful when data are large and continuous time-series.

Hu et al. [37] explored 27 studies that applied ML to forecast chemical contaminants in consumption water. Their review showed that RF was frequently used, particularly for groundwater contamination classification.

Özsezer and Mermer [38] used seven ML algorithms to estimate drinking water condition using a benchmark data containing physicochemical indicator. XGBoost and RF performed particularly well in their studies. Such results give support for developing ML models for drinking-water condition.

Essamlali et al. [39] reviewed the ML with IoT for water condition tracking and described the use of wireless technologies. The sensors are applied to collect data for

indicators and ML models can use this data to identify problems or predict future water condition.

Dharmarathne et al. [40] reviewed ML with IoT-based water-condition evaluation and highlighted the effectiveness of automated systems. New automated methods are useful for automated drinking water plants.

Ajakwe et al. [41] proposed the CIS-WQMS which combines AI and edge intelligence with IoT hardware to provide fast water-condition tracking and prediction based on threshold requirements.

Rahu et al. [42] proposed an infrastructure with IoT water condition analysis. Data were collected from the Rohri Canal which is located in Pakistan. The MLP and random forest models achieved 93% and 91% accuracy for the WQI and WQC prediction, respectively. They also concluded that ML models perform better with smaller datasets.

Bataineh et al. [43] proposed a hybrid ML model for WQI prediction. The author constructs a model by combining an SHAP-initialized neural network and the XGBoost boosting model to determine whether water is drinkable or not.

Makumbura et al. [44] integrated ML models with SHAP-based explainable AI (XAI) to recognize the contribution of different WQP to model predictions. Their work illustrates how XAI can provide additional information about which parameters influence water condition predictions.

Campos et al. [45] investigated AutoML for WQI prediction using a smaller number of parameters and demonstrated that WQI prediction can be automated.

No	Authors	Water Source	ML/DL Model(s)	Key Findings	Limitations Reported
1	Hassan et al. [24]	Surface water	RF, ANN, SVM, MLR and other ML models	MLR models showed strong capability for WQI prediction/classification;while nitrate,pH, conductivity, DO,total coliform,BOD were important variables	The performance depends on the dataset characteristics and parameter availability.
2	Prasad et al. [25]	Surface water	SVM,MDT,LR,RF,naive Bayesian	RF models yielded with 95% accuracy	Well trained datasets; only individual ML and DL;No hybrid architectures
3	Nasir et al. [26]	River/surface water	tree-based (DT, RF),boosting-based (XGBoost, CatBoost), statistical (LR, SVM), and neural network-based (MLP) models	Compared with seven classifiers; CATBoost achieved the highest accuracy	Performance may vary with geographical and dataset characteristics of the study.
4	Khoi et al. [27]	River/surface water	boosting-based ensemble models (AdaBoost, GB, HGB, LGBM, XGBoost), tree-based models (DT, ET, RF), and neural network-based models (MLP, RBF, DFFNN, CNN).	XGBoost gives 98% accuracy	The site-specific nature of the dataset limits direct generalization.
5	Cheng et al. [29]	River/surface water	stacked ensemble models	ST-RF improved accuracy,	The site-specific nature of the dataset limits direct generalization.

6	J. K. Pandya et al. [30]	surface water	ensemble combine with tree-based model (DT), ensemble model (RF), boosting-based models (GBM, XGBoost), and neural network-based model (NN) interactive dashboard for water potability prediction	Model gives 96.7% accuracy	Performance depends on dataset characteristics and parameter availability
7	Chen et al. [31]	Surface/river water and other aquatic systems	ANN/DNN	ANN models can capture nonlinear relationships and are widely applicable to water condition prediction.	ANN performance is significantly influenced by data integrity and model design
8	Li et al. [32]	Surface/river water ;Haihe River Basin, China	LSTM	LSTM is best for temporal water-quality checking.	It requires sufficient temporal data and careful model tuning.
9	Zhang et al. [33]	Surface/river water	CEEMDAN-LSTM	Coupling signal decomposition with LSTM improved the prediction of dynamic water condition.	A hybrid architecture is more complex and computationally demanding.
10	Zhu et al. [34]	Surface water	CEEMDAN-CNN-LSTM	The model achieves high accuracy with robustness against data noise	exploring other advanced DL techniques
11	Tan et al. [35]	Water environment	CNN-LSTM	CNN-LSTM captures local patterns and temporal dependencies and improves predictions over conventional approaches.	It requires substantial time-series data and model training.
12	Haq et al. [10]	Aquaculture water	CNN with LSTM and GRU separately	Hybrid DL provided effective prediction for smart aquaculture applications.	The model complexity and dataset size can affect deployment.

13	Yan M et al. [36]	River/surface water	GCAE-RL Network	DRL +Triple-Channel Adjacency-Attention GCN;Time-series Images conversion; Deep Clustering Autoencoder + Auto-Regression	unstable for the actual water
14	Özsezer & Mermer [38]	2,400 water bodies (physicochemical dataset)	RF, XGBoost, plus 5 others	RF and XGBoost performed best for drinking-water condition classification	Dataset-specific performance; limited to physicochemical data
15	Essamlali et al. [39]	General water condition tracking	ML + IoT integration	IoT sensors + ML support live abnormality detection	Sensor fouling, missing data, calibration errors, cybersecurity
16	Ajakwe et al. [41]	General (CIS- WQMS framework)	AI + edge computing	Edge-based ML reduced response time, improved availability	Practical deployment challenges (network reliability, fouling, calibration)
17	Rahu et al. [42]	Rohri Canal, Pakistan	MLP, RF	MLP achieved 93% accuracy for WQI , RF good with 91% for WQC	Small datasets improve performance; scalability issues
18	Bataineh et al. [43]	Secondary datasets	feature-informed neural network initialization +(XGBoost)	Model an accuracy of 86.9%	No framework to multiclass WQI classification and time-series forecasting ;No dynamic datasets

19	Makumbura et al. [44]	Surface/freshwater system	RF, LightGBM, XGBoost + SHAP	XGBoost achieved strong prediction performance; SHAP helped identify influential water-quality parameters.	High-performing models still require interpretability and validation across locations.
20	Campos et al. [45]	River/surface water	AutoML, CatBoost, RF, XGBoost	AutoML predicted WQI using a reduced parameter set; ensemble tree models performed strongly.	Reduced-parameter models require validation across other regions and datasets.

Table1: summary of reviewed literature

METHODOLOGY APPROACH

We searched for papers on Google Scholar, IEEE Xplore, ScienceDirect, and MDPI. We used the keywords "water quality prediction", "machine learning", "water quality index", and "IoT water monitoring" to search for papers. We considered papers from 2020 to 2026. This approach helps identify research gaps and future directions

CONCLUSION

From the review, we can conclude that ML can support water-quality prediction and WQI classification. It can also be used to detect contaminants and for real-time monitoring. There are several gaps. In many studies, water quality can be checked using water from rivers, lakes, groundwater, and reservoirs. There is no any research for predicting water quality for packaged drinking water. Many studies predict WQI or water quality, but fewer studies integrate ML predictions with regulatory compliance. Many models reported high accuracy without sufficient evaluation of the model and real-time deployment. Finally, most studies investigate individual water-quality prediction, IoT monitoring, or classification separately. But very few studies integrate them into a unified framework. In the future, we can focus on developing an ML model with real time prediction for water quality checking and BIS compliance assessment in packaged drinking water, with performance evaluation and real-time implementation

REFERENCES

- [1] Choudhary, R., Kumar, A., C. P., Naik, M. M., Choudhury, M., & Khan, N. A. (2025). Predicting water quality index using stacked

- ensemble regression and SHAP based explainable artificial intelligence. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-025-09463-4>
- [2] Talukdar, S., S., Ahmed, S., Naikoo, M. W., Rahman, A., Mallik, S., Ningthoujam, S., Bera, S., & Ramana, G. (2023). Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2023.136885>
- [3] Krishna kumar, S., Logeshkumaran, A., Magesh, N.S. et al. Hydro-geochemistry and application of water quality index (WQI) for groundwater quality assessment, Anna Nagar, part of Chennai City, Tamil Nadu, India. *Appl Water Sci* 5, 335–343 (2015). <https://doi.org/10.1007/s13201-014-0196-4>
- [4] Jan, Farmanullah & Min Allah, Nasro & Dustegor, Dilek. (2021). IoT Based Smart Water Quality Monitoring: Recent Techniques, Trends and Challenges for Domestic Applications. *Water* <https://doi.org/10.3390/w13131729>
- [5] Islam, M.A., Roy, S.C., Utsho, M.F.U. et al. An IoT-enabled intelligent water quality monitoring system for tourist safety using machine learning. *Discov Electron* 2, 28 (2025). <https://doi.org/10.1007/s44291-025-00070-1>
- [6] Bhowmik, V. (2025). *Assessing recreational water quality with machine learning and SHAP-based interpretability*. *World Journal of Advanced Research and Reviews*, 27(3), 922–930. <https://doi.org/10.30574/wjarr.2025.27.3.3166>
- [7] Durgun, Y. (2024). *Real-time water quality monitoring using AI-enabled sensors: Detection of contaminants and UV disinfection analysis in smart urban water systems*. *Journal of King Saud University - Science*, 36, 103409. <https://doi.org/10.1016/j.jksus.2024.103409>
- [8] Ahmed, A., Othman, F., Afan, H., Ibrahim, R. K., Fai, C., Hossain, S., Ehteram, M., & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2019.124084>
- [9] Mengistu, T., Chung, I.-M., & Chang, S. (2026). Machine Learning for Water Quality Prediction and Uncertainty Assessment. *Physics and Chemistry of the Earth, Parts A/B/C*. <https://doi.org/10.1016/j.pce.2026.104319>
- [10] Rasheed Abdul Haq, K. P., & Harigovindan, V. P. (2022). Water quality prediction for smart aquaculture using hybrid deep learning

- models. *IEEE Access*, 10, 60078–60098 <https://doi.org/10.1109/ACCESS.2022.3180482>
- [11] Pyo, J., Pachepsky, Y., Kim, S., Abbas, A., Kim, M., Kwon, Y., Ligaray, M., & Cho, K. H. (2023). Long short-term memory models of water quality in inland water environments. *Water Research X*, 21. <https://doi.org/10.1016/j.wroa.2023.100207>
- [12] Barzegar, R., Aalami, M., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, 34, 415–433. <https://doi.org/10.1007/s00477-020-01776-2>
- [13] Davoudi, S., & Roushangar, K. (2025). Innovative approaches to surface water quality management: advancing nitrate (NO₃) forecasting with hybrid CNN-LSTM and CNN-GRU techniques. *Modeling Earth Systems and Environment*, 11. <https://doi.org/10.1007/s40808-025-02291-5>
- [14] Choi, J., Lim, K., & Ji, B. (2023). Robust imputation method with context-aware voting ensemble model for management of water-quality data. *Water research*, 243, 120369 <https://doi.org/10.1016/j.watres.2023.120369>
- [15] A., Y., Cheng, Y., Wang, P., Wang, G., Wang, L., Xue, B., Wang, Y., & Wu, J. (2026). Continuous spatial prediction of river water quality based on a novel hybrid physical-data framework. *Applied Water Science*, 16. <https://doi.org/10.1007/s13201-025-02717-6>
- [16] Uddin, M., Kamruzzaman, S., Sharmin, N., Jamal, A., & Haque, M. (2024). IoT based real-time water quality monitoring system in water treatment plants (WTPs). *Heliyon*, 10. <https://doi.org/10.1016/j.heliyon.2024.e40746>
- [17] Kumar, J., Gupta, R., Sharma, S., Chakrabarti, T., Chakrabarti, P., & Margala, M. (2024). IoT-Enabled Advanced Water Quality Monitoring System for Pond Management and Environmental Conservation. *IEEE Access*, 12, 58156–58167. <https://doi.org/10.1109/access.2024.3391807>
- [18] Lin, J.-Y., Tsai, H.-L., & Lyu, W. (2021). An Integrated Wireless Multi-Sensor System for Monitoring the Water Quality of Aquaculture. *Sensors (Basel, Switzerland)*, 21. <https://doi.org/10.3390/s21248179>
- [19] Singh, Y., & Walingo, T. (2024). Smart Water Quality Monitoring with IoT Wireless Sensor Networks. *Sensors (Basel, Switzerland)*, 24. <https://doi.org/10.3390/s24092871>
- [20] Budda, R., Srinivasan, K., Chauhan, G., Jadon, R., Surya, V., & Gollapalli, T. (2024). Cloud-Integrated AI-Enhanced Software IoT Framework for Real-Time Water Quality Monitoring and E. Coli Prediction via Wireless Sensor Networks. *International Journal of Multidisciplinary and Current Research*. <https://doi.org/10.14741/ijmcr/v.12.1.12>
- [21] Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
- [22] Zhu, M., Wang, J., Yang, X., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*, 1, 107 - 116. <https://doi.org/10.1016/j.eehl.2022.06.001>
- [23] Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X., & Zhao, X. (2024). A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years. *Journal of Marine Science and Engineering*, 12(1), 159. <https://doi.org/10.3390/jmse12010159>
- [24] Hassan, M.M., Hassan, M.M., Akter, L. et al. Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. *Hum-Cent Intell Syst* 1, 86–97 (2021). <https://doi.org/10.2991/hcis.k.211203.001>
- [25] Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Soumya, K., & Poornema, A. J. (2021). Prediction on water quality of a lake in Chennai, India using machine learning algorithms. *Desalination and Water Treatment*, 218, 44–51 <https://doi.org/10.5004/dwt.2021.26970>
- [26] Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., & Al-Shamma'a, A. (2022). Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, 48, Article 102920. <https://doi.org/10.1016/j.jwpe.2022.102920>
- [27] Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. *Water*, 14(10), 1552. <https://doi.org/10.3390/w14101552>
- [28] Shams, M., Elshewey, A. M., El-Kenawy, E.-S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2023). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83, 35307 - 35334. <https://doi.org/10.1007/s11042-023-16737-4>
- [29] Cheng, C., et al. (2024). Urban fine-grained water quality monitoring based on stacked machine learning approach. *IEEE Access*, 12, 77152–77167. <https://doi.org/10.1109/ACCESS.2024.3404068>
- [30] J. K. Pandya, S. S. Khandelwal, R. K. Tipu and K. S. Pandya, "Advancing Water Quality Management: An Integrated Approach Using Ensemble Machine Learning and Real-Time Interactive Visualization," in *IEEE Access*, vol. 13, pp. 92406–92428, 2025, <https://doi.org/10.1109/ACCESS.2025.3573589>
- [31] Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A Review of the Artificial Neural Network Models for Water Quality Prediction. *Applied Sciences*, 10(17), 5776. <https://doi.org/10.3390/app10175776>
- [32] Li, Q., Yang, Y., Yang, L. et al. Comparative analysis of water quality prediction performance based on LSTM in the Haihe River Basin, China. *Environ Sci Pollut Res* 30, 7498–7509 (2023). <https://doi.org/10.1007/s11356-022-22758-7>
- [33] Zhang, L., Jiang, Z., He, S. et al. Study on Water Quality Prediction of Urban Reservoir by Coupled CEEMDAN Decomposition and LSTM Neural Network Model. *Water Resour Mnage* 36, 3715–3735 (2022). <https://doi.org/10.1007/s11269-022-03224-y>
- [34] Zhu, X. Yu, L. Zhao and L. Xu, "A Hybrid Deep Learning Model for Water Quality Prediction," in *IEEE Access*, vol. 13, pp. 106406–106415, 2025, <https://doi.org/10.1109/ACCESS.2025.3580741>
- [35] Tan, W., Zhang, J., Wu, J., Lan, H., Liu, X. et al. (2022). Application of CNN and Long Short-Term Memory Network in Water Quality Predicting. *Intelligent Automation & Soft Computing*, 34(3), 1943–1958. <https://doi.org/10.32604/iasc.2022.029660>
- [36] Yan M. & Wang Z. (2024). Water quality prediction method based on reinforcement learning graph neural network. *IEEE Access* 12 184421–184430. <https://doi.org/10.1109/ACCESS.2024.3509744>
- [37] Hu, X.C., Dai, M., Sun, J.M. et al. The Utility of Machine Learning Models for Predicting Chemical Contaminants in Drinking Water: Promise, Challenges, and Opportunities. *Curr Envir Health Rpt* 10, 45–60 (2023). <https://doi.org/10.1007/s40572-022-00389-x>
- [38] Özsezer, G., & Mermer, G. (2024). Prediction of drinking water quality with machine learning models: a public health nursing approach. *Public Health Nursing*, 41, 175–191. <https://doi.org/10.1111/phn.13264>
- [39] Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Advances in machine learning and IoT for water quality monitoring: A comprehensive review. *Heliyon*, 10(6), e27920. <https://doi.org/10.1016/j.heliyon.2024.e27920>
- [40] Dharmarathne, G., Abekoon, A. M. S. R., Bogahawaththa, M., Alawatugoda, J., & Meddage, D. P. P. (2025). A review of machine learning and internet-of-things on the water quality assessment: Methods, applications and future trends. *Results in Engineering*, 26, 105182. <https://doi.org/10.1016/j.rineng.2025.105182>
- [41] Ajakwe, S. O., Ajakwe, I. U., Jun, T., Kim, D.-S., & Lee, J.-M. (2023). CIS-WQMS: Connected intelligence smart water quality

- monitoring scheme. Internet of Things, 23, 100800. <https://doi.org/10.1016/j.iot.2023.100800>
- [42] Rahu, M. A., Chandio, A. F., Aurangzeb, K., Karim, S., Alhussein, M., & Anwar, M.S. (2023). Toward design of Internet of Things and machine learning-enabled frameworks for analysis and prediction of water quality. IEEE Access, 11, 101055–101086. <https://doi.org/10.1109/ACCESS.2023.3315649>
- [43] Bataineh et al. (2026). A hybrid machine learning framework for water quality index prediction using feature-based neural network initialization. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 19, 4887 - 4905. <https://doi.org/10.1109/JSTARS.2026.3654017>
- [44] Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. Results in Engineering, 23, Article 102831. <https://doi.org/10.1016/j.rineng.2024.102831>
- [45] Campos, D., Galvão, V., de Rezende, M.L. et al. Automated machine learning achieves accurate water quality prediction with reduced parameter requirements. Sci Rep 16, 4431 (2026). <https://doi.org/10.1038/s41598-025-34448-8>