Grenze International Journal of Engineering and Technology, June Issue

Evaluation of Water Quality using Machine Learning and Deep Learning Approaches

Savitha N[1] and Srinath S[2]
[1-2]Dept. of Computer Science and Engineering, JSS Science and Technology University, Mysuru, 570006
Email: savithan@jssstuniv.in

***Abstract*— Water quality estimation is one of the feasible and essential activities in protection of the popular health and sustainable development, which is important to any healthy ecosystem. The nonlinearity and complexity of a water system however is a major obstacle in trying to implement this objective. Such spatiotemporal complexities can be usually hard to be represented using traditional statistical models and autonomous machine learning regimens, especially in situations where there is a lot of noise or scarce information. More recent developments in the literature attempt to address the above limitations by developing integrated and hybrid deep learning frameworks that effectively trade off the advantages of state-of-the-art neural network designs such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU) and Convolutional Neural Networks (CNN) and Transformer-like networks with complex time-frequency decomposition algorithms (e.g., Wavelet Transform, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Complete Ensemble Empirical Mode Decomposition with Adaptive No These methods improve feature recovery, replica multi-scale temporal cycles and resilience. Explainable AI and data-driven mechanistic/coupling are also helpful because they enhance forecast accuracy and interpretability. Overall, such developments present a scalable simple paradigm of dependable water quality prediction enabling intelligent surveillance, timely warning, and dependable management of the aquatic environment.**

***Keywords*— water quality index, machine learning, real-time monitoring, GIS, decision support, environmental governance.**

I. INTRODUCTION

In the assessment of water quality, manual sampling and analysis in the laboratory were employed. Water samples in fixed monitoring stations were assessed by using standard analytical tools in evaluating physical, chemical, and biological parameters which included pH, dissolved oxygen, turbidity, and electrical conductivity, nutrients and microbial indicators. Waters Quality Indices (WQI) were frequently used to evaluate the results and that incorporated the factors into one measurement with a score of outstanding, good, fair, or awful. Mostly to identify trends and relationships, it was discovered that water quality was related to contributory factors such as temperature, discharge, and land use using simple statistical and deterministic models in several forms; multiple logistic regression and simple time-series models like ARIMA as well were found not to be so successful in real-time prediction and management in complex and changeable river systems.

Water quality forecasting is important in case of ecological preservation, public health and sustainable water resource management.

Grenze ID: 01.GIJET.12.2.24_25

The human, hydrological and climatic influences, however, affect the complex, nonlinear and dynamic nature of water systems. In some situations, these complex spatiotemporal relationships cannot be reliably modeled with basic statistical and off-the-shelf machine learning approaches, and especially in data-poor and noisy settings.

There are various machine learning (ML) methods due to their ability to process large and heterogeneous datasets and predict nonlinear relationships, thus becoming increasingly popular. ML, real-time monitoring, biological analysis, and sensors have enabled a world of variety of research. This survey has bundled these research streams giving an arranged outline of the methods, advantages and limitations. The main findings of this paper are: (i) a structured investigation of the means of assessing the quality of river water; (ii) a directed analysis of the supervised and unsupervised machine learning techniques of pollution classification and prediction; and (iii) the determination of gaps in open research and potential opportunities.

The hybrid deep learning paradigms based on time-frequency decomposition algorithms and complex neural networks such as CNN, LSTM and Transformer-based channels are used to address these issues. The techniques boost feature detection, multi-scale time-scale detection, and improve strength and understanding. The combination of them into a single implementation gives a scalable and useful paradigm of proper water quality predictions and smart administration of the aquatic environment.

In as much as the aspects of detection, analysis and comparative potential of the available water quality management methods are useful, they are begotten with inherent constraints in respect of timeliness, adaptability and interpretation. When all the gaps are considered one can see the importance of incorporated structures which are non opaque, forecasting and responsive to changing conditions in the river.

II. Classification of Anomalies

In as much as the aspects of detection, analysis and comparative potential of the available water quality management methods are useful, they are begotten with inherent constraints in respect of timeliness, adaptability and interpretation. When all the gaps are considered one can see the importance of incorporated structures which are non opaque, forecasting and responsive to changing conditions in the river.

A. Real-Time Monitoring and Anomaly Detection

To address the purpose of identifying unexpected cases of pollution and looking at associated environmental hazards, real-time monitoring systems have been of interest [1]. Continuous sensing and multivariate statistics of the anomaly events, causes of pollution, levels of danger, can be identified in highly stressed river systems. These techniques are particularly effective in detecting unintentional contamination and industrial discharge sporadic that happens.

However, though they have a number of advantages, many real-time monitoring systems do not have predictive intelligence and use statistical or threshold-based systems of detection. It has yet to be integrated on decision-support systems and adaptive machine learning models.

B. Data-Driven and Machine Learning Approaches for Water Quality

The ability of machine learning methods to detect complex nonlinear interactions has made them extremely popular in Water quality prediction and classification. The Adaptive Neuro-Fuzzy Inference Systems (ANFIS) that adjust with intelligence algorithms on unpredictable and varying environmental conditions have demonstrated superior accuracy in prediction [2]. In time series modeling, such time-series models as ARIMA have also been employed to estimate short-term water quality and quantity [3].

The methods of supervised and unsupervised water contamination predictors have been compared in the recent papers. As demonstrated by Zamri et al., supervised and unsupervised algorithms can be both used to improve the performance of pollution classification since multiple unsupervised and supervised algorithms perform better [4]. A similar study observed that distance selection plays an important role in classification accuracy and involved a number of distance measurements with the k-nearest neighbors (KNN) river quality classification algorithm [4].

The growing application of machine learning in water quality estimation is further confirmed by extensive review research, which incorporates all the wastewater management aspect through surface water and groundwater evaluation [5]. They have also shown that the performance of the Random Forest-based classifiers with the use of sampling methods increase the work of the unbalanced datasets consisting of river water quality [6]. Although the ML models are highly accurate, most of them are opaque, and are difficult to understand, and hence not applicable in judicial, governance and regulatory decision-making contexts.

C. Biological Indicators and Ecological Assessment

Biological methods of assessment especially the one pegged on macroinvertebrate communities is a long term measure of river health as it entails cumulative ecological stress [3]. Biotic indices can be used to allow streams to be classified into the water-quality category and are commonly applied in ecological surveillance and controls. But biological analyzes are resource demanding, time consuming and cannot be used in real time detection of pollution. They have frequently been used without the use of chemical monitoring and predictive modeling frameworks.

D. Remote Sensing, GIS, and Geomorphological Analysis

The use of remote sensing and GIS methods allows performing a very large amount of spatially explicit research on river water quality and the factors that regulate it [4]. The research of geomorphological evolution combined with water quality indexes proved that the channel morphology, sediment transportation, and dispersion of pollution have significant connections. These methods are both scalable and spatially descriptive, but are mostly retrospective and descriptive, as well as having fewer predictive capabilities of real-time and the ability to make an adaptive decision.

E. Multi-Criteria Decision-Making Approaches

On complex and less-data-rich environments, multi-criteria decision-making has been used to determine the quality of surface water using the Fuzzy Analytic Hierarchy Process (FAHP) [5]. These methods combine various pointers and professional decision-making in the face of uncertainty assisting in comparative analysis and prioritization of the degree of pollution.

However, MCDM models have been typically fixed and subjective, restricting them to dynamic rivers and data streams.

III. LITERATURE REVIEW

In the prediction of water quality, the literature has evidently adopted a paradigm shift, whereby traditional statistical and standalone machine learning algorithms have been replaced by the sophisticated hybrid and deep learning models.

Recent findings suggest that signal decomposition methods (Wavelet Transform, CEEMD, EWT and STL) are beneficial when used together with deep models (LSTM, GRU, and Transformer-based architectures) in order to capture nonlinear and multi-scale dynamics of time-related signals. The attention mechanisms and spatiotemporal learning enable the model to be stronger in highlighting salient features and relationships and outperform the traditional approaches based on LSTM.

Mechanistic-data-coupled accuracy and interpretability can be achieved through the use of physical processes and sources of pollution, e.g. SWAT coupled with deep learning. Although hybrid models that integrate autoencoders, recurrent networks, and ensemble learners provide excellent performance in wastewater and industrial settings, transfer learning works effectively in cases of data deficits at ill-monitored locations. The general tendency is that integrated, scalable, and interpretable deep learning structures are the core towards an accurate and intelligent prediction of water quality, whereas more simple models such as SVM and Random Forest may be used in less complex cases.

TABLE I: COMPARATIVE SUMMARY OF RELATED WORK ON WATER QUALITY ASSESSMENT

Citation	Water Source / Parameters	Algorithm / Model Used	Key Results	Major Findings
Unified DL framework for WQ prediction (Lijiang River, China)	River water quality: DO, CODMn + meteorology & hydrology	Wavelet Transform + Informer Encoder + Stacked BiLSTM	RMSE: DO 0.329, CODMn 0.121; MAE: DO 0.217, CODMn 0.057; Max relative error: DO 12.4%, CODMn 20.7%	Time–frequency decomposition + global–local DL learning yields high accuracy and robustness in complex rivers.
SWAT–XGBoost–LSTM under non-point pollution	Rivers: NH_3-N, Total Nitrogen (TN)	SWAT (mechanistic) + LSTM + XGBoost	R^2 ↑ 6.3% (NH_3-N), 7.5% (TN); RMSE ↓ up to 20.5%	Coupling physical models with interpretable DL improves prediction and quantifies non-point source impacts.
Transfer Learning for data-scarce sites	Surface water at sparse stations (multiple WQ params)	Neural Networks + Transfer Learning with SMIs	RMSE reduced up to 79.9%	TL enables reliable forecasting at new or data-poor stations; source-domain similarity is crucial.

LSTMAE–XGBoost (Pulp & Paper effluent)	Industrial wastewater effluent quality	Autoencoder + LSTM + XGBoost	R^2 +40%, RMSE −35% vs PLS	Hybrid feature extraction + sequence learning + boosting handles noise and nonlinearity effectively.
CEEMD + LTSF-Linear (Korea)	Water supply system: pH, EC	CEEMD + LTSF-Linear	High accuracy across short/mid/long term	Multi-scale decomposition enhances linear forecasters for seasonal and climate-driven variability.
Large-scale EWT–DL assessment (China, 350 sites)	Rivers: TN, TP	Empirical Wavelet Transform + GRU	Avg. RMSE ↓ 47.7%; best at 88% sites	Decomposition–DL hybrids scale well and remain robust under data sparsity.
BiLSTM–KAN for WWTPs	WWTP effluent: COD, others	BiLSTM + Kolmogorov–Arnold Network + feature fusion	COD RMSE ↓ 7.7–45.2%; R^2 ↑ up to 14.8%	Captures nonlinear interactions; interpretable and robust for real-time plant control.
Surrogate + VMD–CPO–LSTM (Nanming Basin)	Reclaimed water plants & river sections: COD, NH_3-N, TP	MIKE 11 + XGBoost surrogate + VMD–CPO–LSTM	RMSE < 1; R^2 > 0.8; section error < 12%	Fast surrogate + DL enables downstream short-term forecasting with low computation cost.
Dual-Attention Encoder–Decoder	Rivers: multi-parameter WQ	Encoder–Decoder LSTM + Dual Attention	Out-of-sample accuracy ≈ 80%	Joint attention over features and time improves future trend prediction.
ML for WQI (Aquaculture)	Water bodies: WQI & classes	LightGBM, XGBoost, RF, SVM + SHAP	Accuracy up to 99.65%; XGBoost R^2 = 0.9685	Ensemble ML with explainability supports early warning and disease-risk management.
DKCTAN + Poplar Optimization (India)	Indian WQ dataset: WQI & classes	Dual-Key Conv-Transformer Autoencoder + POA	≈99% accuracy	Advanced DL + meta-optimization achieves near-perfect WQI prediction.
ELSTM–EBP (Daling River Basin)	Rivers: Total Nitrogen (TN)	Enhanced LSTM + BP Network	Multi-step error within ±0.4 mg/L (7 steps)	Captures spatiotemporal TN dynamics and key drivers.
CNN–BiLSTM–SA (Henan)	Industrial water: Total Hardness	CNN + BiLSTM + Self-Attention	R^2 = 0.945; MAPE = 0.65%	Hybrid DL excels for long-sequence, real-time hardness prediction.
MV-Online-LSTM (Multi-point sources)	Multi-site river WQ indicators	Multi-view Online LSTM	R^2 > 0.96 across indicators	Online learning adapts to dynamic environments and spatial dependencies.
Stacked DL in Cloud (STP)	Sewage treatment WQI	LSTM/GRU/BiLSTM + Stacked Ensemble	R^2 = 0.93; RMSE = 0.0655	Cloud-enabled ensembles provide scalable, real-time monitoring.
Grey Models (Anhui Cities)	Urban WQI	Fractional Grey Multivariate Model + GCA	Improved accuracy; future WQI trends	Suitable for limited data; agriculture is dominant pollution driver.
Feature-driven Hybrid Attention (Yellow River)	Rivers: multi-parameter WQ	AHP + GAT + Informer	MAE 0.22; RMSE 0.09; 67.9% gain	Feature selection + hybrid attention captures complex spatiotemporal dynamics.
Classical ML (Tireh River, Iran)	River WQ components	ANN, GMDH, SVM	SVM best; lowest DDR	Kernel-based ML remains competitive for traditional WQ tasks.
DeepTCN–GRU (Yellow River Source)	pH, TN	TCN + GRU	RMSE ↓ ≥29.7%; R^2 > 0.9	Combines CNN temporal features with RNN memory for strong generalization.
ARIMA (River Pinios, Greece)	Multiple WQ params + discharge	ARIMA + cross-correlation	Clear lag relationships	Statistical models reveal physical controls (T, Q) on chemistry.
Ensemble STL–Attention (River sites)	DO, TP, NH_3-N	STL + Temporal & Spatio-Temporal Attention	Long-step error ↓ up to 22%	Learning from trend components improves long-horizon forecasts.
Stacking for WQI (Southern Bug, Ukraine)	WQI	11 Stacking ML models	GPR best; stacking improves others	Meta-learner choice is critical for accuracy vs. cost.
ML Comparison (Zayandeh Rood, Iran)	EC, TDS, SAR, TH	Lasso, RF, GB, XGBoost, SVM + PCA	R^2 up to 0.999 (SAR)	Best model varies by parameter; ensembles & SVM dominate long-term prediction.

To forecast water quality, the latest studies reveal the shift in the traditional models and the adoption of hybrid and deep learning models and networks, which combine decomposition, attention, physical modeling, and transfer learning. The operational WQ forecasts that are represented by hybrid and interpretable AI are becoming the new standard due to the greatly increased accuracy, resistance to challenging situations, and practical as well as real-time decision support provided by these methods.

A. Machine learning approaches

How machine learning can be utilized to control water quality is highlighted in the paragraph. In aquaculture, ensemble models have extremely high accuracy (up to 99.65) with interpretable output via SHAP that can be used to make informed decisions in real-time.

TABLE II: STUDY ON WATER QUALITY ASSESSMENT USING MACHINE LEARNING APPROACHES

Study/ Application Area	Data& Region	Machine Learning Approach	Key Techniques Used	Main Outcomes
WQI & Classification (Aquaculture)	Real-time water quality data	LightGBM, XGBoost, RF, SVM	Ensemble machine learning, SHAP-based interpretability	Accuracy up to 99.65%
Urban Water Quality (China)	16 cities	Grey Multivariate Model	Grey correlation analysis	Improved long-term prediction accuracy

In the meantime, the Grey Multivariate Model demonstrates that machine learning may be utilized in both short-term monitoring and the long-term prediction of urban water quality under uncertainty.

B. Deep learning approaches

The paragraph demonstrates the way in which various deep learning models improve the quality of water prediction under different hydrologic conditions. These hybrid methods are accurate, have low error rates, are more interpretable, and can do strong generalization by utilizing decomposition, attention mechanisms, CNNRNN fusion, and physical modeling. Everything under consideration, it can be stated that the deep learning is able to predict and control the water quality accurately and capture the most intricate spatiotemporal variations.

TABLE III: STUDY ON WATER QUALITY ASSESSMENT USING DEEP LEARNING APPROACHES

Study / Application Area	Data & Region	Deep Learning Approach	Key Techniques Used	Main Outcomes
Unified DL Framework (Lijiang River, China)	River WQ + meteorological–hydrological data	WT–Informer–SBiLSTM	Wavelet decomposition, global & local temporal learning	RMSE: DO 0.329, COD Mn 0.121; high robustness
SWAT–Interpretable DL Model	River basins (China)	SWAT–LSTM (hybrid DL)	Mechanistic coupling, interpretable DL	R^2 ↑ 6–8%, RMSE ↓ up to 20%
WWTP Real-Time Control	Effluent sensors	BiLSTM–KAN	Hybrid DL, correlation fusion, SHAP	RMSE ↓ up to 45%, $R^2 > 0.92$
Out-of-Sample River Prediction	Multi-parameter rivers	Dual-Attention Encoder–Decoder	Temporal & feature attention	Accuracy ≈ 80%
Indian WQI Dataset	National WQ records	DKCTAN (CNN–Transformer–AE)	Deep hybrid architecture + optimization	99% accuracy
Yellow River Source	pH, TN	DeepTCN–GRU	CNN–RNN hybrid	RMSE ↓ ≥ 29%, $R^2 >$ 0.9

As presented in the following table, the traditional machine learning and statistical models have turned into hybrid deep learning, attention-based, transfer learning, and mechanistic-data-driven models. It has demonstrated continuous improvements in the accuracy, stability, and scale of various water systems.

The gathered papers show a clear improvement of the water quality research towards more complex hybrid and deep learning architectures that would be capable of dealing with nonlinear and multiple-scale and spatiotemporal dynamics. Although the earlier methods of ARIMA, ANN, SVM, and regression models are still useful in simple or long-term cases, clearly, it performs poorly in complex and data-sparse situations. Recent studies are composing deep networks such as as LSTM, GRU, TCN, Transformer, and Informer with signal decomposition methods (Wavelet, CEEMD, EWT, STL, and VMD) to enhance the representation of features and multi-frequency temporal behaviors. Architecture Hybrid architectures comprising of attention mechanism, spatiotemporal learning and ensemble techniques have been shown to significantly improve the accuracy and resilience of prediction in river basins, wastewater treatment plants, reclaimed water systems, and urban networks. Mechanistic -data-driven coupling (e.g., SWAT and MIKE 11 with ML/DL) improves the analysis of pollution sources and provides physical interpretability, whereas transfer learning has proven effective to address the lack of sufficient data in new monitoring sites. Explainable AI tools (SHAP, LIME) enhance decision-making transparency and thoughtful evaluations are used to verify the scalability of decomposition-deep learning models. On the whole, studies have shown that integrated, explainable, and scalable deep architectures are the best paradigm of accurate, intelligent and real-time prediction of water quality.

IV. Challenges and Future Research Directions

Although hybrid versions of deep learning models work well to predict the quality of water, they suffer interpretation, out-of-region ability, data dependency, and computational expenditure. Their non-openness reduces the trust and their fractions often drop to below data inadequacy and between basins. These problems underscore the need to have lightweight, lightweight and legible models that can be implemented in the real world.

Further studies are needed to develop lightweight and real-time models of edge and Internet of Things implementation and transferable frameworks, which function under different hydrological conditions. Composing mechanics models with learning physics based on physics principles will enhance reliability but automated feature selection and adaptive attention can enhance durability. The extension of transfer and federated learning will enable collaborative modeling within limited data. This will allow long-term forecasting due to integrating remote sensing, climate projections, and socioeconomic information, as well as improve early warning and decision-making through the creation of explainable AI.

V. Conclusion

The analysis indicates that models with hybrid and deep learning are always superior to the traditional statistical and independent machine learning models in predicting the water quality. These methods significantly minimise errors, enhance accuracy and increase robustness on a variety of hydrological and operational systems by integrating decomposition, attention systems, mechanistic coupling, and transfer learning. They can be good in capturing multi-scale time patterns, scanty data, providing real-time control, and interpretable predictions. They have the advantage that their RMSE can be reduced by up to 79.9, accuracy is likewise enhanced by more than 60, and their classification performance is close to 100. The new integrated deep learning structures offer a more recent, more scalable and accurate system to predict water quality, although older methods remain useful in the case of simple or short term solutions.

Challenges and Future Research Directions

Based on the literature it is understandable that the conventional statistical and standalone machine learning techniques have been substituted with integrated hybrid and deep learning models of water quality prediction. One can promote the clearness of the signal as well as the multi scale pattern extraction utilizing the decay methodologies, though the deep models (LMST, GRU, CNN, and Transformer) are advantageous in finding nonlinear and long-range relations. The consistency of information is also enhanced with the help of attention mechanisms and spatially temporal learning, which help to relieve salient features and spatial associations. Mechanistic-data-coupling which goes a step further with physical knowledge combining with data-driven models offers better forecasting and interpretation of sources. Transfer learning makes the model more applicable to places with inadequate surveillance due to the fact that the lack of data will be overcome. Despite these developments, there are problems with the processing cost, inter-region generalization, operational interpretability and dependence on data.

The future research should aim at the development of lightweight, real-time deployment models of edges and Internet of things and also deployable transferable frameworks that can perform well in various hydrological environments. Mechanistic models can be combined with physics-informed learning, which would lead to increased reliability, and automated feature selection and adaptive attention can be used to enhance stability. Federated learning and Transfer extension will support collaborative modeling in the conditions of data scarcity. Explainable AI creation will improve the variety of predictive technologies and the decision-making process, the implementation of remote sensing, climate prediction, and socioeconomic information will make long-term forecasting possible.

References

[1] J. A. Hernández-García et al., “Detection, provenance and associated environmental risks of water quality pollutants during anomaly events in River Atoyac, Central Mexico: A real-time monitoring approach,” Science of the Total Environment, vol. 689, pp. 1221–1236, 2019.

[2] A. Ghorbani, M. R. Kisi, and A. Shiri, "Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (Case study: Gorganrood River)," Journal of Hydrology: Regional Studies, vol. 47, 2024.
[3] D. Koutsoyiannis et al., "Monitoring, modeling, and assessment of water quality and quantity in River Pinios using ARIMA models," Sustainable Water Resources Management, vol. 10, 2024.
[4] N. Zamri et al., "A comparison of unsupervised and supervised machine learning algorithms to predict water pollutions," Procedia Computer Science, vol. 204, pp. 172–179, 2022.
[5] M. Zhu et al., "A review of the application of machine learning in water quality evaluation," Eco-Environment & Health, vol. 2, pp. 107–116, 2022.
[6] R. Fadhilah, H. Kuswanto, and D. D. Prastyo, "Performance analysis of Random Forest with sampling for river water quality classification," in Proc. Int. Conf. Informatics and Computational Sciences (ICICoS), 2024.
[7] A. Nigatu et al., "Microplastic pollution in river systems: A case study of Megech River, Ethiopia," Regional Studies in Marine Science, vol. 68, 2023.
[8] J. Melad et al., "Spatial profiling of river water quality using GIS techniques: A case study of the Mananga River, Philippines," Environmental Monitoring and Assessment, vol. 195, 2023.
[9] R. Englónér and G. Németh, "Impact of climate and hydrological variability on river water quality: Insights from the Danube River," Hydrology Research, vol. 54, pp. 145–160, 2023.
[10] S. Crespo et al., "A hydro-economic modeling framework for integrated river basin management: Application to the Colorado River Basin," Water Resources Research, vol. 59, 2023.
[11] S. Kumar et al., "Evolution of river geomorphology to water quality impact using remote sensing and GIS technique," Egyptian Journal of Remote Sensing and Space Science, vol. 27, 2024.
[12] M. S. Islam et al., "Assessment of surface water quality using Fuzzy Analytic Hierarchy Process (FAHP): A case study of Piyain River's sand and gravel quarry mining area in Jaflong, Sylhet," Journal of Environmental Management, vol. 215, pp. 394–407, 2018.
[13] M. Zhu et al., "A review of the application of machine learning in water quality evaluation," Eco-Environment & Health, vol. 2, pp. 107–116, 2022.
[14] Y. Wang et al., "Development of biological water quality categories for streams using a biotic index of macroinvertebrates in the Yangtze River Delta, China," Ecological Indicators, vol. 117, 2020.
[15] R. Xu et al., "A unified deep learning framework for water quality prediction based on time-frequency feature extraction and data feature enhancement," Journal of Environmental Management, Feb. 2024.
[16] J. Huan et al., "Deep learning model based on coupled SWAT and interpretable methods for water quality prediction under the influence of non-point source pollution," Computers and Electronics in Agriculture, Apr. 2025.
[17] Z. Liao et al., "Enhancing surface water quality prediction in data-scarce sites using transfer learning and neural networks," Journal of Water Process Engineering, Jun. 2025.
[18] K. Zhang et al., "Enhancing water quality prediction with advanced machine learning techniques: An extreme gradient boosting model based on long short-term memory and autoencoder," Journal of Hydrology, Nov. 2024.
[19] J. Shin et al., "Development of water quality prediction model using LTSF-Linear and complete ensemble empirical mode decomposition," Desalination and Water Treatment, Jul. 2025.
[20] J. Shi et al., "Water quality prediction at large-scale: Systematic assessment of decomposition–deep learning models," Journal of Cleaner Production, Aug. 2025.
[21] S. Liu and Z. Wang, "Real-time effluent water quality prediction model based on BiLSTM and KAN for wastewater treatment plants," Journal of Water Process Engineering, Oct. 2025.
[22] J. Feng et al., "Short-term water quality prediction of reclaimed water plant effluent and key measurement sections based on a surrogate prediction model," Journal of Environmental Management, Apr. 2025.
[23] Z. Zheng et al., "Research on out-of-sample prediction method of water quality parameters based on dual-attention mechanism," Environmental Modelling & Software, May 2024.
[24] M. A. A. M. Hridoy et al., "Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management," Science of The Total Environment, Dec. 2025.
[25] R. Maruthamuthu et al., "Enhanced water quality index prediction using a dual-key convolutional transformer autoencoder network integrating with poplar optimization," Franklin Open, Dec. 2025.
[26] Y. Liu and Y. Wang, "Water quality prediction method based on a combined machine learning model: A case study of the Daling River Basin," Journal of Contaminant Hydrology, Jan. 2026.
[27] H. Xu et al., "Study on short-term total hardness prediction of water quality based on CNN-BiLSTM-SA model: A case study of Henan Province," Journal of Water Process Engineering, Jun. 2025.
[28] D. Wang, C. Zhang, and F. Wu, "Time-series prediction of water quality at multiple points in water sources," Environmental Modelling & Software, Jan. 2026.
[29] Riya et al., "An intelligent cloud-enabled system for water quality index prediction of sewage treatment plant using deep learning models," Journal of Water Process Engineering, Oct. 2025.
[30] W. Wu et al., "Water quality prediction based on different kinds of water consumption in urban areas in China," Applied Mathematical Modelling, Dec. 2025.
[31] X. Yao et al., "Feature-driven hybrid attention learning for accurate water quality prediction," Expert Systems with Applications, Jun. 2025.

[32] A. H. Haghiabi, A. H. Nasrolahi, and A. Parsaie, “Water quality prediction using machine learning methods,” Water Quality Research Journal, vol. 53, no. 1, 2018.
[33] Q. Tian et al., “Water quality prediction in the Yellow River source area based on the DeepTCN-GRU model,” Journal of Water Process Engineering, Mar. 2024.