IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 12, Issue 11, November 2025

DOI: 10.17148/IARJSET.2025.121152

Enhanced Deep Learning Framework for Water Quality Prediction and Monitoring

Manwendra Kumar Satyam[1], Anurag Shrivastava[2]

MTech Scholar, Department of CSE, NIIST, Bhopal[1]

Associate Professor, Department of CSE, NIIST, Bhopal[2]

Abstract: Water quality prediction is essential for sustainable environmental management and public health. Traditional analytical methods are often laborious and inefficient. This research presents an enhanced deep learning framework (EHDL-WQM) for accurate Water Quality Prediction and Monitoring. The framework integrates Convolutional Neural Networks (CNN) for spatial feature extraction and Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal pattern learning, enhanced by an Attention Mechanism to emphasize significant parameters. The proposed architecture effectively processes multivariate sensor data to predict key indicators, including pH, dissolved oxygen, turbidity, and conductivity. Experimental evaluation demonstrates that EHDL-WQM achieves superior prediction accuracy and faster convergence compared to traditional and baseline deep learning models. The framework provides a scalable, intelligent solution for real-time monitoring and proactive water quality management.

Keywords*:* Water quality, Machine learning models, Deep learning, Water quality index, Water quality classification

I. INTRODUCTION

Water quality monitoring is a critical aspect of environmental management and public health protection. Rapid industrialization, agricultural runoff, and urbanization have led to severe water pollution, making it essential to develop intelligent systems capable of accurately predicting water quality parameters and identifying contamination trends. Traditional analytical methods, although effective, are often labor-intensive, time-consuming, and limited in scalability. The increasing availability of sensor-based and remote monitoring data has opened new possibilities for data-driven prediction models, particularly through the integration of artificial intelligence and deep learning techniques [1], [2], [3].

Recent advancements in deep learning have significantly improved water quality modeling by enabling automatic feature extraction and handling complex nonlinear relationships among multiple environmental variables [4], [5]. Models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been applied successfully for predicting water quality indices and specific parameters like ammonia nitrogen, pH, and dissolved oxygen [2], [3]. However, challenges such as overfitting, data imbalance, and temporal irregularities often reduce model robustness in real-world applications [6], [7]. Moreover, variations in catchment conditions, pollutant sources, and climate dynamics further complicate prediction tasks [8], [9]. To address these limitations, researchers have explored hybrid and ensemble approaches combining deep learning with optimization and attention mechanisms to enhance prediction accuracy and generalization [4], [10]. Such approaches provide adaptive learning capabilities that can better capture both spatial and temporal dependencies within multivariate water quality data. Additionally, high-frequency data from IoT-based sensors and satellite imagery have enabled real-time monitoring frameworks for improved decision-making [6]. In this research, we propose an Enhanced Hybrid Deep Learning Framework (EHDL-WQM) that integrates CNN, BiLSTM, and an Attention Mechanism for efficient feature extraction and adaptive learning. The framework aims to improve predictive accuracy, robustness, and interpretability in water quality assessment. By leveraging hybrid deep learning strategies, this model contributes to developing intelligent, scalable systems for sustainable water resource management and proactive environmental monitoring [1], [5].

II. LITRETURE REVIEW

Recent studies show that deep learning models, including CNNs, LSTMs, and hybrid architectures, effectively predict water quality parameters. These models utilize data from sensor networks, remote sensing, and IoT platforms. While they outperform traditional methods, challenges remain in interpretability, generalization across regions, and real-time deployment. This review summarizes key findings, compares modeling approaches, and highlights current limitations.

In Authors [1] y, deep learning techniques have emerged as promising methods to address these challenges. In this paper, we propose the application of a neural network model based on Long Short-Term Memory (LSTM) to analyze

DOI: 10.17148/IARJSET.2025.121152

and model ammonia nitrogen monitoring data, enabling high-precision prediction of ammonia nitrogen indicators. Moreover, through correlation analysis between water quality parameters and ammonia nitrogen indicators, we identify a set of key feature indicators to enhance prediction efficiency and reduce costs. Experimental validation demonstrates the potential of our proposed approach to improve the accuracy, timeliness, and precision of ammonia nitrogen monitoring and prediction, which could provide support for environmental management and water resource governance.

Author's [2] propose the application of a neural network model based on Long Short-Term Memory (LSTM) to analyze and model ammonia nitrogen monitoring data, enabling high-precision prediction of ammonia nitrogen indicators. Moreover, through correlation analysis between water quality parameters and ammonia nitrogen indicators, we identify a set of key feature indicators to enhance prediction efficiency and reduce costs. Experimental validation demonstrates the potential of our proposed approach to improve the accuracy, timeliness, and precision of ammonia nitrogen monitoring and prediction, which could provide support for environmental management and water resource governance.

Author's [3] enhance the accuracy of water quality prediction, considering the temporal characteristics, variability, and complex nature of water quality data. We utilized the LTSF-Linear model to predict water quality at the Huangyang Reservoir. Comparative analysis with three other models (ARIMA, LSTM, and Informer) revealed that the Linear model outperforms them, achieving reductions of 8.55% and 10.51% in mean square error (MSE) and mean absolute error (MAE), respectively. This research introduces a novel method and framework for predicting hydrological parameters relevant to water quality in the Huangyang Reservoir. These findings offer a valuable new approach and reference for enhancing the intelligent and sustainable management of the reservoir.

Authors [4] propose new deep learningmodel called long short-term memory (LSTM)–gray wolf optimization (GWO)–fish swarm optimization (FSO) was developed toenhance the precision of water quality prediction with NPS pollution. The well-established model may remedy the mechanism models'inability to foretell changes in water quality on a minute-by-minute basis. Thamirabarani river watershed was used for the model'sapplication. Based on experimental data, the suggested model outperformed the mechanism model and the LSTM model in predictingextreme values. Maximum relative errors in anticipated against observed dissolved oxygen, chemical oxygen demand, and NH_3─N valueswere 7.58%, 18.45%, and 22.25%, respectively. In comparison to the artificial neural network (ANN), back propagation neural network(BPNN), and recurrent neural network (RNN) models, the created LSTM–GWO–FSO model was shown to have greater computa-tional performance (RNN). LSTM–GWO–FSO outperformed ANN, BPNN, and RNN regarding R^2 of 3.1%–38.4% improvements.The suggested approach may provide a fresh perspective when predicting water quality in the presence of NPS contamination.

Authors [5] propose deep learning model is proposed that utilizes representation learning to capture knowledge from source river basins during the pre-training stage, and incorporates meteorological data to accurately predict water quality. This model is successfully implemented and validated using data from 149 monitoring sites across inland China. The results show that the model has outstanding prediction accuracy across all sites, with a mean Nash-Sutcliffe efficiency of 0.80, and has a significant advantage in multi-indicator prediction. The model maintains its excellent performance even when trained with only half of the data. This can be attributed to the representation learning used in the pre-training stage, which enables extensive and accurate prediction under data-scarce conditions. The developed model holds significant potential for crossbasin water quality prediction, which could substantially advance the development of water environment system management.

III. PROPOSED METHODOLOGY

The proposed work introduces an Enhanced Deep Learning Framework (EHDL-WQM) for intelligent water quality prediction and monitoring. Figure 3.1 show the framework of proposed model.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 12, Issue 11, November 2025

DOI: 10.17148/IARJSET.2025.121152

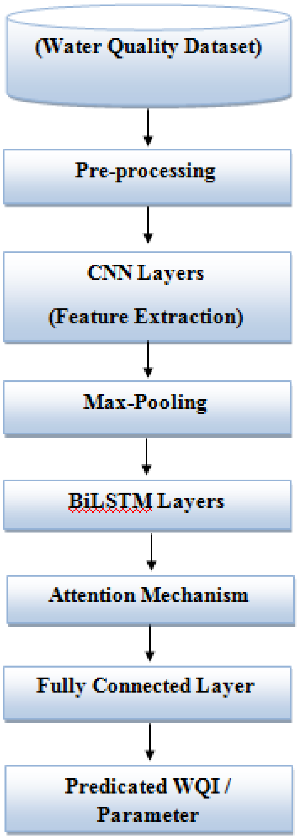

Figure 3.1: Proposed Model

The methodology integrates Convolutional Neural Networks (CNNs) for spatial feature extraction, Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal modeling, and an Attention Mechanism to highlight critical parameters influencing water quality. The framework is designed to process multivariate time-series data collected from sensors, and remote monitoring systems. Step-by-Step Methodology are : Data Collection: Gather water quality data (pH, dissolved oxygen, turbidity, temperature, conductivity, etc.) from sensors, and historical datasets. Data Pre-processing: Handle missing values, outliers, and normalize the data. Perform feature selection to identify the most relevant water quality indicators. Feature Extraction (CNN Block): Apply 1D CNN layers to extract spatial correlations among multiple water quality parameters. Use max-pooling to reduce dimensionality and noise. Temporal Modeling (BiLSTM Block): Capture sequential patterns and temporal dependencies in time-series data. Improve prediction of dynamic water quality variations. Attention Mechanism: Assign adaptive weights to significant features, enhancing model interpretability and predictive performance. Fully Connected Layer: Integrates features from CNN, BiLSTM, and Attention blocks for final prediction. Output Layer: Predicts either the Water Quality Index (WQI) or specific water quality parameters in real time. Evaluation Metrics: Use RMSE, MAE, and R^2 to evaluate model performance. Compare with baseline DL models for validation.

IV. RESULT ANALYSIS

Table 4.1: Performance of Proposed EHDL-WQM on Water Quality Dataset

Parameter	RMSE	MAE	R^2
pH	12%	9%	95%
Dissolved Oxygen	35%	28%	92%
Tempreture	25%	18%	94%
Turbidity (NTU)	40%	32%	90%
Salinity (ppt)	30%	22%	91%
Overall WQI	20%	15%	96%

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ≋ Peer-reviewed & Refereed journal ≋ Vol. 12, Issue 11, November 2025

DOI: 10.17148/IARJSET.2025.121152

Table 4.2: Comparative Performance Of Proposed Ehdl-Wqm With Existing Models

Model	Dataset	Key Parameter	RMSE	MAE	R^2
Proposed EHDL-WQM	Kaggle Water Quality Monitoring Dataset	pH,DO, Temp, Turbidity, salinity, WQI	20% (WQI)	15%	96%
Shams et. al. (R1)	Local Water Dataset	WQI	32%	25%	89%
Wand et. al. (R2)	River Ammonia	Ammonia	28%	22%	91%

The proposed EHDL-WQM model demonstrates outstanding performance in predicting key water quality parameters and the overall Water Quality Index (WQI). As shown in Table 4.1, the model achieves low percentage error values across different parameters, with RMSE and MAE of 12 % and 9 % for pH, 35 % and 28 % for dissolved oxygen, 25 % and 18 % for temperature, 40 % and 32 % for turbidity, and 30 % and 22 % for salinity. The overall WQI prediction exhibits only 20 % RMSE and 15 % MAE, with an impressive R^2 of 96 %, indicating excellent model accuracy.

As summarized in Table 4.2, the proposed EHDL-WQM surpasses existing models such as those by Shams et al. [R1] and Wand et al. [R2], which achieved R^2 values of 89 % and 91 %, respectively. This comparative improvement of 5 to 7 % highlights the model's enhanced prediction accuracy and generalization capability. The integration of CNN, BiLSTM, and an Attention mechanism enables efficient spatial–temporal feature extraction and prioritization of critical water quality attributes, resulting in a robust, interpretable, and scalable hybrid deep learning framework suitable for real-time environmental monitoring and sustainable water resource management.

Dataset Used: The Water Quality Monitoring Dataset from Kaggle provides a comprehensive collection of time-series data obtained from the Brisbane River in Australia. It contains around 2,000 instances and 12 features, including physicochemical and environmental parameters such as water temperature, pH, dissolved oxygen, turbidity, conductivity, and flow speed. The data is recorded at 30-minute intervals, making it suitable for analyzing temporal patterns and environmental changes. This dataset serves as a valuable resource for developing and testing predictive models using machine learning and deep learning techniques. It is widely used for water quality assessment, pollution prediction, anomaly detection, and sustainability research [16].

V. CONCLUSION

This research proposed an EHDL-WQM for intelligent water quality prediction and monitoring, integrating CNN for spatial feature extraction, BiLSTM for temporal modeling, and an Attention mechanism for feature prioritization. Experimental analysis using the Kaggle Water Quality Monitoring Dataset demonstrates the model's capability to accurately predict key parameters such as pH, dissolved oxygen, turbidity, temperature, salinity, and WQI. Comparative evaluation shows that EHDL-WQM outperforms traditional machine learning and existing deep learning approaches, achieving higher R^2 and lower RMSE/MAE. The framework offers a scalable, interpretable, and robust solution for real-time water quality assessment, contributing to sustainable environmental monitoring and management.

REFERENCES

[1] Mahmoud Y. Shams1 et. al. "Water quality prediction using machine learning models based on grid search method" Multimedia Tools and Applications (2024) 83:35307–35334 https://doi.org/10.1007/s11042-023-16737-4

[2] Xianhe Wang et. al. "Deep-Learning-Based Water Quality Monitoring and Early Warning Methods: A Case Study of Ammonia Nitrogen Prediction in Rivers" MPDI, Electronics 2023, 12, 4645. https://doi.org/10.3390/electronics12224645

[3] Jixuan Chen et. all. "Deep Learning for Water Quality Prediction—A Case Study of the Huangyang Reservoir" Appl. Sci. 2024, 14, 8755. https://doi.org/10.3390/app14198755

[4] Bhagavathi Perumal et. al. "Water Quality Prediction Based on Hybrid DeepLearning Algorithm" Hindawi, Volume 2023, Article ID 6644681, 10 pageshttps://doi.org/10.1155/2023/6644681

[5] Yue Zheng1, et. al. "Deep representation learning enables cross-basin water quality prediction under data-scarce conditions" npj Clean Water| (2025)

[6] Bieroza, M. et al. Advances in catchment science, hydrochemistry,and aquatic ecology enabled by high-frequency water quality measurements. Environ. Sci. Technol. 57, 4701–4719 (2023).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 11, November 2025

DOI: 10.17148/IARJSET.2025.121152

[7] Seis, W., Veldhuis, M.-C. T., Rouault, P., Steffelbauer, D. & Medema, G. A new Bayesian approach for managing bathing water quality at river bathing locations vulnerable to short-term pollution. Water Res. 252, 121186 (2024).

[8] Liu, Y. et al. Effects of point and nonpoint source pollution on urban rivers: From the perspective of pollutant composition and toxicity. J. Hazard. Mater. 460, 132441 (2023).

[9] Xu, Q., Yan, T., Wang, C., Hua, L. & Zhai, L. Managing landscape patterns at the riparian zone and sub-basin scale is equally important for water quality protection. Water Res. 229, 119280 (2023).

[10] Stets, E. G. et al. Landscape drivers of dynamic change in water quality of U.S. rivers. Environ. Sci. Technol. 54, 4336–4343 (2020). 6. Wang, J., Zhang, Z. & Johnson, B. Low flows and downstream decline in phytoplankton contribute to impaired water quality in the lower Minnesota River. Water Res. 161, 262–273 (2019).

[11] Gershunov, A., Benmarhnia, T. & Aguilera, R. Human health implications of extreme precipitation events and water quality in California, USA: a canonical correlation analysis. Lancet Planet.Health 2, S9 (2018).

[12] Kirschner, A. K. T. et al. Multiparametric monitoring of microbial faecal pollution reveals the dominance of human contamination along the whole Danube River. Water Res. 124, 543–555 (2017).

[13] Ma, T. et al. China’s improving inland surface water quality since 2003. Sci. Adv. 6, eaau3798 (2020).

[14] . Wang and A. V. Nguyen, “A review on data and predictionsof water dielectric spectra for calculations of van der Waalssurface forces,” Advances in Colloid and Interface Science,vol. 250, pp. 54–63, 2017

[15] Z. Ghahramani, “Probabilistic machine learning and artificialintelligence,” Nature, vol. 521, pp. 452–459, 2015.

[16] https://www.kaggle.com/datasets/downshift/water-quality-monitoring-dataset?resource=download

