

A Review on Deep Learning Prediction for Water Quality Monitoring

Manwendra Kumar Satyam¹, Anurag Shrivastava²

¹*MTech Scholar, CSE, Department, NIIST, Bhopal*

²*Associate Professor, Department of CSE, NIIST, Bhopal*

Abstract: The increasing scarcity and contamination of freshwater resources have made water quality monitoring an essential global concern. Traditional approaches to water quality assessment, which often rely on manual sampling and laboratory testing, are time-consuming, labor-intensive, and lack real-time responsiveness. In recent years, the emergence of deep learning (DL) has offered a transformative potential for predictive analytics in environmental monitoring. This review paper presents a comprehensive overview of the current research landscape on deep learning techniques applied to water quality prediction. Various models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid architectures, are analyzed for their performance in forecasting key water quality parameters such as pH, turbidity, dissolved oxygen, and chemical contaminants. The review also explores datasets, sensor integration, feature engineering methods, and evaluation metrics used across studies. Furthermore, it highlights the challenges related to data scarcity, model interpretability, and deployment in resource-constrained environments. By synthesizing recent advances, this paper aims to guide researchers and practitioners toward developing efficient, accurate, and scalable DL-based solutions for real-time water quality monitoring, with implications for smart water management and public health protection.

Keywords— Water quality, Machine learning models, Deep learning, Water quality index, Water quality classification

I. INTRODUCTION

Water is a fundamental resource for life, ecosystems, agriculture, and industry. As global populations rise and industrial activities expand, the degradation of water quality has become a pressing concern worldwide. Poor water quality not only affects human health but also has severe implications for biodiversity, economic development, and sustainable

resource management. Traditional water quality monitoring techniques involve manual sampling followed by laboratory analysis, which, although accurate, are time-consuming, costly, and lack scalability and real-time capability. With the advent of advanced sensing technologies and the Internet of Things (IoT), large volumes of data can now be collected from diverse aquatic environments, laying the groundwork for automated, intelligent water monitoring systems. In this context, deep learning (DL)—a subset of artificial intelligence—has emerged as a powerful tool for processing complex, high-dimensional data. Unlike conventional machine learning algorithms, DL models can automatically learn spatial and temporal patterns from raw sensor inputs without requiring extensive manual feature extraction. Techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks have demonstrated considerable success in environmental data analysis, particularly in time-series prediction tasks relevant to water quality parameters [1, 2].

Water quality indicators such as pH, dissolved oxygen (DO), turbidity, temperature, and nitrate concentration are dynamic and influenced by a multitude of natural and anthropogenic factors. Predicting these parameters in real time is critical for early detection of pollution events, regulatory compliance, and proactive water management. Deep learning models, when trained on large and diverse datasets, have the potential to provide accurate, real-time predictions, thereby enabling decision-makers to respond promptly to environmental risks.

This review aims to synthesize current research efforts in applying deep learning techniques for water quality prediction. It discusses the various types of DL

architectures used, the datasets and sensors employed, and the performance metrics considered in different studies. The review also identifies existing challenges such as data sparsity, interpretability of deep models, and limitations in deployment across varying environmental contexts. Through this examination, the paper highlights research gaps and offers insights for future developments toward smart, scalable, and sustainable water quality monitoring systems using deep learning.

II. LITRETURE REVIEW

Recent studies have demonstrated the efficacy of deep learning models such as CNNs, LSTMs, and hybrid architectures in predicting water quality parameters with high accuracy. Researchers have applied these models to diverse datasets collected from sensor networks, remote sensing, and IoT platforms. While several works highlight improvements in prediction performance over traditional methods, challenges remain in terms of model interpretability, generalization across regions, and real-time deployment. This review consolidates key findings, model comparisons, and limitations from the current body of research.

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 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 learning model called long short-term memory (LSTM)-gray wolf optimization (GWO)-fish swarm optimization (FSO) was developed to enhance 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's application. Based on experimental data, the suggested model outperformed the mechanism model and the LSTM model in predicting extreme values. Maximum relative errors in anticipated against observed dissolved oxygen, chemical oxygen demand, and NH₃-N values were 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 computational performance (RNN). LSTM-GWO-FSO outperformed ANN, BPNN, and RNN regarding R² 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.

FINDINGS OF THE SURVEY

The survey of recent literature reveals that deep learning has significantly enhanced the predictive capabilities of water quality monitoring systems. Various architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid models, have been successfully implemented to forecast parameters like pH, turbidity, dissolved oxygen, and temperature. These models outperform traditional machine learning algorithms by automatically learning temporal and spatial features from complex, non-linear environmental data. Studies also show that combining DL models with real-time sensor data from IoT frameworks enables continuous monitoring and early detection of anomalies or pollution events. Moreover, researchers have started exploring transfer learning and ensemble techniques to improve generalizability across different water bodies and geographic locations. Despite their promise, deep learning models still face challenges such as the need for large, high-quality datasets, difficulties in interpreting black-box predictions, and limited real-world deployment in

rural or low-resource areas. In this survey find out some important points:

Deep learning models, especially CNNs, LSTMs, and hybrid frameworks, outperform traditional approaches in water quality prediction.

These models effectively capture non-linear relationships and temporal dependencies in environmental data.

IoT-integrated systems using deep learning support real-time monitoring and early pollution detection.

Transfer learning and ensemble methods improve the robustness and accuracy across varied datasets and geographical areas. Studies highlight the superiority of deep learning in predicting key parameters like pH, DO, turbidity, and nitrate levels. There is a growing trend of deploying DL models on edge devices for real-time analytics. Key challenges include: Requirement of large labeled datasets, Model interpretability, Lack of scalable deployment in rural or remote environments.

The literature suggests the need for standardized benchmarks and frameworks for water quality prediction tasks.

CONCLUSION

Deep learning has emerged as a powerful approach for water quality prediction, offering significant improvements in accuracy, automation, and real-time monitoring over traditional methods. Through the analysis of diverse studies, it is evident that models like CNNs, LSTMs, and hybrid networks can effectively predict key water parameters and support early detection of contamination events. However, challenges such as data scarcity, interpretability, and limited deployment persist. Continued research is essential to address these gaps and to develop scalable, explainable, and energy-efficient deep learning systems for sustainable water resource 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] [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] [3] Jixuan Chen et. al. “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] [4] Bhagavathi Perumal et. al. “Water Quality Prediction Based on Hybrid DeepLearning Algorithm” *Hindawi, Volume 2023, Article ID 6644681, 10 pages*
<https://doi.org/10.1155/2023/6644681>
- [5] [5] Yue Zheng1, et. al. “Deep representation learning enables cross-basin water quality prediction under data-scarce conditions” *npj Clean Water* (2025)
- [6] [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).
- [7] [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] [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] [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][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).
- [11]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).
- [12][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).
- [13][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).
- [14][13] Ma, T. et al. China’s improving inland surface water quality since 2003. *Sci. Adv.* 6, eaau3798 (2020).
- [15][14] . Wang and A. V. Nguyen, “A review on data and prediction of water dielectric spectra for calculations of van der Waals surface forces,” *Advances in Colloid and Interface Science*, vol. 250, pp. 54–63, 2017
- [16][15] Z. Ghahramani, “Probabilistic machine learning and artificial intelligence,” *Nature*, vol. 521, pp. 452–459, 2015.