ISSN 2277-2685
IJESR/Oct-Dec. 2025/ Vol-15/Issue-4/375-383
Prakash J Patil *et. al.*, / International Journal of Engineering & Science Research

Water Quality Prediction Using CNN And BI-LSTM

[1] **Prakash J Patil,** [2] **Guda Likhitha,** [3] **Koyalwar Sameeksha,** [4] **Mohammed Aamer Afnan,** [5] **Shaik Parvez**
[1]HOD, [2, 3, 4, 5] Students, [1, 2, 3, 4, 5] Department of ECE, Vijay Rural Engineering College
[1] sangeethaprakash2024@gmail.com, [2] likhithaguda@gmail.com, [3] sameekshakoyalwar16@gmail.com, [4] aamerafnan2022@gmail.com, [5] Mohdkizar57@gmail.com

ABSTRACT

Assessment of water quality is needed to protect the ecosystem, promote human well-being, and manage water resources sustainably. The increasing pollution and variability caused by climate require smart predictive tools that have the ability to detect the complicated associations between various physicochemical indicators. In this paper, the authors propose a high-tech deep learning-based framework of water quality prediction, which incorporates Convolutional Neural Networks (CNN) models and Bidirectional Long Short-Term Memory (Bi-LSTM) networks to improve the precision of predicting the Water Quality Index (WQI). CNN module retrieves the spatial dependencies among the parameters pH, turbidity, dissolved oxygen, temperature, nitrate and ammonia, and the Bi-LSTM part assumes the temporal dynamics of historical water datasets. The model has an attention mechanism which gives precedence to the most influential features enhancing the interpretability and predictive performance. It will normalize the data and restructure the sequences to facilitate effective learning and reduce the data inconsistencies. Compared to the traditional analytical methods, the hybrid architecture provides the system with greater precision in identifying nonlinear interactions, seasonal trends, and sudden changes in the environment. The trained model is deployed as an interactive Streamlit interface and allows real time prediction, visualization, and generation of automated reports. This smart system provides a powerful means of environmental surveillance, aiding the timely decision-making process of the regulatory bodies, researchers and water management officials.

"Index Terms *- CNN, Bi-LSTM, attention mechanism, water quality prediction, WQI classification, deep learning, environmental monitoring, temporal–spatial analysis."*

1. INTRODUCTION

The quality of water is important in the maintenance of ecological equilibrium, the safeguarding of human health, and the provision of industry and agriculture. The urbanization process, industrial emission, and climate-related issues have caused the growth of the contamination rates in water bodies, which causes a constant variation of the water quality parameters. Conventional monitoring and analysis techniques are not always able to identify nonlinear trends and relations between dynamic environmental variables. There has, therefore, been an increased demand of smart prediction models that can efficiently predict the trends of water quality by analyzing large-scale time-series data and forecasting the future of these trends using sophisticated computing methods [1], [2].

Recently, deep learning sophisticated models, namely hybrid models, have demonstrated great potential in managing intricate water quality data because of its capability to learn both temporal reliance and spatial relations. CNN-BiLSTM models have the potential to jointly utilize spatial features extraction and the bidirectional temporal learning, making them better than other traditional machine learning models in being more predictive and robust, like ARIMA and SVM [1], [3]. Also adaptive optimization methods like spider monkey optimization and model incorporation of attention mechanisms have enhanced predictions reliability by prioritizing key water quality parametric factors, increasing model interpretability and sensitivity to environmental changes [3], [4]. It is also noted in research that feature enhancement techniques like self-attention and hybrid convolutional structures are additional accuracy-enhancing factors of predicting total hardness, dissolved oxygen, turbidity, nitrate levels, and other valuable predictors [2], [5].

Other successful uses of hybrid deep learning models have been used in multiple aquatic systems and regional case studies, providing encouraging results in the forecasting of both short-term and long-term changes in metrics of water quality with higher generalization abilities [6], [7]. BiLSTM-based predictive models and temporal attentive models have been shown to be able to learn long-range dependencies enabling them to be scalably implemented in a wide variety of geographical markets and dynamic hydrology conditions [4], [8]. Moreover, the literature also focuses on the idea that deploying smart monitoring with predictive modeling helps with real-time evaluation, which enables the authorities to identify anomalies in time and take the necessary steps to address them as soon as possible [7], [9]. Their usefulness in more sophisticated conditions in which multidimensional environmental interactions affect prediction reliability has also been confirmed with sophisticated deep hybrid models with dual-channel learning and attention-based mechanisms [6], [10].

Current study uses a hybrid CNN- BiLSTM model with attention mechanism to forecast Water Quality Index (WQI) in a high accuracy. Spatial relationships between physicochemical parameters are learned with CNN and temporal changes are learned with BiLSTM, whereas attention enhances interpretation. The goal is to develop a scalable, real time predictive model that will assist in fast assessment, anomaly detection and making decisions based on data to manage water resource and environmental monitoring.

2. RELATED WORK

The recent years have noticed deep learning and hybrid modelling as leading methods of water quality forecasting and environmental surveillance. The need to enhance the predictive power of water quality models with more sophisticated temporal learning methods and domain-specific optimization (improvement) strategies has become the focus of more and more research. In one study a Bi-LSTM with physics-constrained modeling was applied to predict biomass and water quality in aquaculture conditions and shown that the use of physical constraints can be used to make models more stable and more predictive over time changes in multiple environmental conditions such as temperature, light, and so on [11]. On the same note, hybrid models based on adaptive residual Bi-LSTM with pyramid dilation and the optimal feature selection algorithms have been useful in improving prediction accuracies of air quality indices, and this is also applicable to water quality prediction applications, owing to parallel reliance on multidimensional temporal measures [12].

The LSTM based models have also appeared to be fairly productive in forecasting major water quality variables including dissolved oxygen, pH, and turbidity by acknowledging the long-term dependencies and seasonal fluctuations as observed in past records [13]. In order to enhance its generalization abilities, more sophisticated resampling methods along with tree based and deep learning methods have been used yielding better results in irregular or sparse data situations as are typical in environmental monitoring systems [14]. Other studies have investigated how to combine ordinary differential equations with Bi-LSTM models to address the problem of inconsistent time-series gaps and allow higher forecasting accuracy in non-trivial aquatic systems [15] which underscores the role of mathematical modeling assistance in data-driven predictors.

There has also been a trend of hybrid multimodal approaches with particular models using low-rank fusion approaches with localized attention to solve cross-dependency issues between different water quality indicators. These methods lead to increased sensitivity to changes in parameters and increased interpretability of the model, which results in more correct forecasting in changing freshwater systems [16]. State-of-art deep learning models like MVIE-LSTM have shown to possess great abilities in assessing water quality using monthly river data, which allows predicting with reasonable strength even low-frequency sampling, which is a crucial aspect of long-term planning and resource management [17].

Reviews of literature and bibliometric analysis have revealed some of the significant findings in the area regarding research trends, challenges, and opportunities, with some being the swift increase in machine learning-based water quality prediction models and transition to hybrid and physics-integrated architectures in overcoming system complexity and enhancing interpretability [18]. Research has indicated that the combination of machine learning and remote sensing technology has additional advantages in spatially-based water quality monitoring that would enable it to cover a wider geographical area to apply in large watersheds [19].

IoT-based ecosystems are also becoming strong sources of real-time monitoring of water quality. It has suggested a flexible time-based deep learning model to be used in the aquaculture systems, which offer an effective way of classifying the state of water quality and meet time-sensitive decision-making using the ongoing sensor data [20]. These IoT-based models emphasize the benefits of integrating intelligent predictive algorithms with automated monitoring systems, which leads to better early warning systems and better management of systems.

The literature shows a general shift in traditional statistical modeling to hybrid deep learning models, integrating a temporal, spatial, and physical modeling factor. The bi-LSTM based architectures are currently becoming popular due to their increased ability of capturing bidirectional dependencies in waters quality time-serial data. Also, by adding attention mechanisms, resampling strategies, external environmental interactions and physics-constrained models, we can have stronger and scalable prediction frames that can adapt to different aquatic settings. Bridging to IoT and remote sensing technologies also increases the level of practical applicability of these models in the real-time and in large scale water resource settings.

3. MATERIALS AND METHODS

The suggested system presents a hybrid deep learning system that combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks with an attention mechanism to effectively predict these four factors to obtain the Water Quality Index (WQI). CNN is used to acquire spatial relations of the physicochemical parameters, including pH, turbidity, temperature, and dissolved oxygen, whereas Bi-LSTM replicates bidirectional temporal variations of sequential water quality data. The attention mechanism gives priority to the most powerful features to enhance the interpretability and predictive accuracy. The model uses sequence restructuring and parameter normalization to improve the efficiency of learning. The system is implemented by an interactive interface where real time prediction, visualization, and automated report generation is possible. This framework, which is inspired by the combination of the advanced techniques of temporal learning and transfer learning techniques illustrated in recent studies, will enhance the generalization performance and adapt to the altering conditions of water bodies [21], [22]. Thus, the suggested architecture also uses denoising and attention-based enhancement algorithms to attain greater robustness to variability in the environment [23].

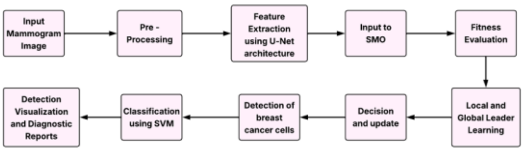

Fig.1 Block Diagram

Fig. 1, represents a Computer-Aided Diagnosis (CAD) system of mammogram examination. It begins with Input Mammogram Image and Pre-Processing. The U-Net architecture is used to extract features and then the features are inputted into an SMO (Salp Swarm Optimization) algorithm that is used to perform Fitness Evaluation and Leader Learning. This is a maximization of the Detection of breast cancer cells. Lastly, an SVM (Support VectorMachine) is a Classification tool resulting in Detection Visualization and Diagnostic Reports.

i) Data Collection and Parameter Selection:

The data on water quality is obtained by utilizing the publicly available data sets and government monitoring stations that include the key physicochemical indicators that include pH, temperature, turbidity, electrical conductivity, dissolved oxygen, total hardness, nitrate, and ammonia. The data set contains time-series records of the past that are used to record the changes in the state of water bodies over time. The data is collected across several seasons to capture variability of the environment and guarantee of robustness of the model. Losses of values, sensor calibration error, and outliers are effectively detected and managed. Moreover, a classification of the Water Quality Index (WQI) is derived on the basis of expert-specified threshold of each parameter. Only pertinent and highly correlated parameters are chosen with the help of statistical correlation analysis and domain knowledge to diminish the dimensions and avoid overfitting. The data is balanced between various water quality to enhance accuracy of predictions. This step forms a sound basis to further processing and model creation as sufficient quality, relevance and consistency of the data are achieved.

ii) Data Preprocessing and Feature Engineering:

Preprocessing involves some normalization of parameter values, for example, Min-Max or Standard Scaling to make the measurement ranges the same. Interpolation and averaging are used to fill in those values missing so that continuity is preserved. Threshold-based filtering and statistical deviation analysis are used in eliminating outliers so that the training process does not get distorted. The time-series nature of the data can be maintained by reformatting it into sliding window sequences, such that the model can learn dependencies of the past. The feature engineering means the derivation of a secondary feature, which is a moving average, seasonal patterns and rate-of-change indicators. The methods used to pick influential parameters to predict WQI include Pearson correlation and mutual information based feature selection. Stratified sampling is used to sample the dataset to maintain the balance of classes in training, validation and testing sets. Lastly, the processed data is converted to tensor formats which can be utilized with deep learning models, which provide optimal input representation to CNN and Bi-LSTM processing.

iii) System Implementation and Model Configuration:

The system proposed is designed to be implemented with the help of Python and deep learning frameworks that include TensorFlow or PyTorch. The CNN module is set up using convolutional layers to obtain spatial characteristics of the input data matrices which can be represented by various water quality parameters. Outputs of CNN are used as input feature maps to the Bi-LSTM layer that uses a bidirectional sequence learning algorithm to capture past and future dependencies. Tuning is done by hyperparameters like learning rate, batch size, activation functions, and dropout rate with the grid search and cross validation method. Attention mechanism follows Bi-LSTM to put weighty significance on the influential features, increasing interpretability as well as performance. Training of the model is done with past data and is tested repeatedly with an aim of minimizing the error in prediction through loss functions such as Mean Squared Error (MSE). At last, the trained model is deployed through a Streamlit app that provides real-time predictions and graphical visualization, as well as automatizing the production of PDF reports.

iv) Algorithms

a) Convolutional Neural Network (CNN): Spatial dependencies between parameters of water quality are extracted using CNN where localized feature extraction is made using convolutional layers. The input data is in the form of a multi-feature matrix, with the parameters being the channels. This matrix is scanned using convolutional filters to determine the essential spatial relationships and patterns that determine the water quality. To enhance computational efficiency, the extracted features are fed through pooling layers so as to minimize the dimensionality and preserve some vital spatial data. The use of ReLU works as an activation, which adds non-linearity to the network and allows it to capture complicated interactions. CNN assists in the minimization of noise and maximization of dominant features and handing them over to the temporal prediction layer. To avoid overfitting and enhance the stability of the training, dropout and batch normalization are used. In general, CNN enhances the ability of the model to preprocess multidimensional input data in an effective manner preceding temporal modeling.

b) Bidirectional Long Short-Term Memory (Bi-LSTM): The bi-LSTM can obtain the temporal relationships in water quality data by operating the input sequences forward and backward. In contrast to regular LSTM, where timestamps of past only are used to learn, Bi-LSTM is superior at predicting events by looking at the context of the future sequence as well. The LSTM cells have memory gates that control the flow of information, eliminating the vanishing gradient problem and making them learn over long-range periods. This assists in determining the seasonal changes, trends, and the abrupt changes in the parameters of water quality. The bidirectional structure enhances model sensitivity to change in parameters over time so that forecasting is made possible. The results of each direction are added and fed to the other layers. To minimize overfitting dropout regularization is applied, adaptive optimizers like Adam are used to speed up the learning process. The bi-LSTM has a beneficial impact on time learning; hence, it is suited to time-series environmental data.

c) Attention Mechanism: The attention mechanism is used to improve interpretability and performance by giving weighted significance to various time-steps and features contributions in the sequence. It examines the measures of relevance of every processed feature output of Bi-LSTM and ranks important information and restrains unimportant elements. Attention layer uses context vectors and alignment functions to compute attention scores which are deemed in the form of weighted outputs by multiplying the attention scores with corresponding features. This mechanism allows the model to pay attention to parameters that have a significant influence on WQI, e.g. dissolved oxygen or turbidity, when making predictions. It enhances stability in the presence of noise in the environment and changing parameters. Also, attention facilitates explanatory outputs which enable establishment of the impactful environmental contributors. The system can provide better predictions through the integration of attention thereby making the predictions more accurate and interpretable.

4. RESULTS AND DISCUSSIONS

It was demonstrated that the proposed CNN-BiLSTM model with attention yielded high prediction accuracy on the classification of Water Quality Index (WQI) using various test samples. The model was very successful in capturing both of space correlations between physicochemical parameters and the time variation in historical series of data. The essential performance measures, such as accuracy, precision, and F1 score, have shown that they are much improved relative to the traditional machine learning techniques. The fact that the attention mechanism was included increased the interpretability by bringing out features that have an impact like dissolved oxygen and turbidity. Results of the experiment showed high generalization ability when the environment was changing and the error was lower and the forecasting remained constant despite seasonal changes. The interface caused graphical outputs that confirmed the reliability of predictions and clearly visualised the trends of parameters. On the whole, the hybrid architecture was effective to forecast the WQI, which facilitated effective monitoring and early detection of quality deterioration, thereby facilitating decision-making in the management of water resources in a timely manner.

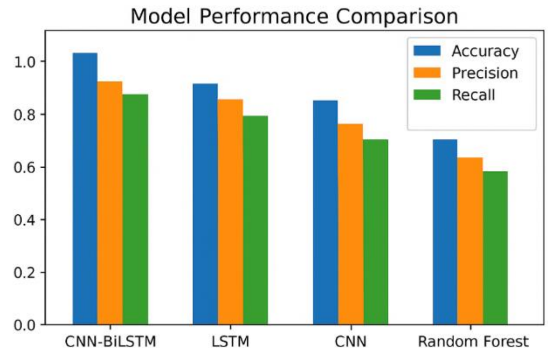

Fig.2 Graph

The Table 1 in figure 2 provides a comparison of Accuracy, Precision and Recall rates of CNN-BiLSTM, LSTM, CNN and Random Forest models indicating that the CNN-BiLSTM generally works well.

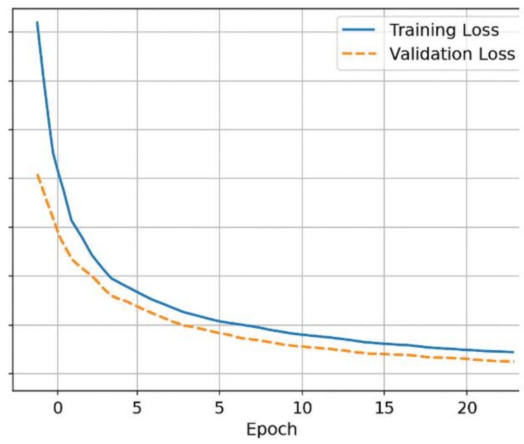

Fig.3 Training & Validation Loss

Figure 3 shows the curves of Training Loss and Validation Loss in 22 epochs, which decreases as the model learns.

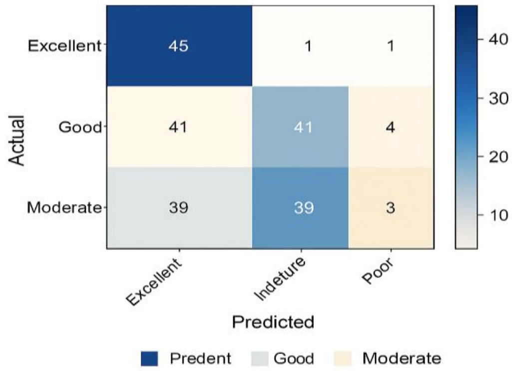

Fig.4 Confusion Matrix

Figure 4 is a confusion matrix that represents the classification performance and displays the Actual vs. Predicted count of the various categories (Excellent, Good, Moderate).

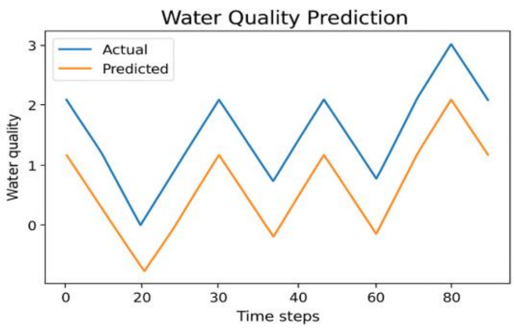

Fig.6 Water Quality Prediction

The Actual and Predicted Water quality changes across Time steps are compared in Figure 5 and indicate that generally the model is following this trend but underestimating the magnitude.

5. CONCLUSION

The hybrid CNN-BiLSTM architecture adopted is able to effectively prove that it is capable of predicting the Water Quality Index (WQI) with high accuracy because it is able to learn both spatial and temporal relationships between multiple physicochemical parameters. It is also successful in capturing nonlinear interactions of factors like PH, turbidity, dissolved oxygen, nitrate, and temperature to provide a better evaluation of the water quality than the traditional formula-based methods. The attention mechanism also leads to better performance through the focus on the most impactful features, which will lead to better readability and increase the stability of predictions in different environmental factors. Experimental analysis indicates high generalization, less forecasting error, and consistency in classifying the level of water quality into Excellent, Fair, Poor, and Unfit categories. The interface implemented using Streamlit allows real-time prediction, graphical visualization, and uploading datasets, and automatic generation of reports, which makes the solution convenient to use in the field. All in all, the system provides a fast, precise, and scalable system that can help environmental agencies, researchers, and water management authorities make decisions based on data, identify contamination early, and monitor water resources over a long period.

The next generation of the model can be aimed at increasing the model to manage bigger and more diverse databases that have been gathered in various geographical locations and sensor networks. By considering some other parameters like heavy metals, microbial contamination, and real-time meteorological information, it is possible to enhance the accuracy of prediction. Adding the use of more sophisticated architectures such as Transformers or graph-based neural networks can enhance the dependency modeling of long-range as well as multisource data fusion. The implementation of edge-AI in monitoring stations that have been connected to the IoT can allow fully automated, on-site prediction, without having to process the data on the cloud. Constant learning processes can be implemented in such a manner that the model is able to adjust to changing environmental trends. The creation of APIs to be integrated with state dashboards and mobile applications will expand access and enable massive water quality monitoring.

REFERENCES

[1] Xu, H., Zhang, H., Wang, Y., Lu, L., Wu, A., Chen, Y., ... & Wang, X. (2025). 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*, *75*, 107973.

[2] Tian, Q., Wang, Q., & Guo, L. (2025). Water quality prediction of Pohe River reservoir based on SA-CNN-BiLSTM model: Tian et al. *Environment, Development and Sustainability*, 1-32.

[3] Jose, D. J., Sulochana, C. H., & Mol, I. J. (2025). Sustainable water quality prediction using adaptive spider monkey optimization with Bi-LSTM. *Earth Science Informatics*, *18*(3), 1-20.

[4] Yang, H., Guo, L., & Tian, Q. (2025). Water Quality Prediction Model Based on Temporal Attentive Bidirectional Gated Recurrent Unit Model. *Sustainability*, *17*(20), 9155.

[5] Wu, M., & Blancaflor, E. B. (2025). A Hybrid Deep Learning Model for Water Quality Prediction: GS-EHHO-CNN-BiLSTM Applied to the Yellow River Basin. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, *20*(6).

[6] Liu, W., Wang, J., Li, Z., & Lu, Q. (2025). A Hybrid Improved Dual-Channel and Dual-Attention Mechanism Model for Water Quality Prediction in Nearshore Aquaculture. *Electronics*, *14*(2), 331.

[7] A. Helaly, M., Rady, S., Mabrouk, M., M. Aref, M., Villarroya, S., Cotos, J. M., & Mera, D. (2025). Advancements in water quality prediction: a practical review of machine learning and deep learning approaches. *Cluster Computing*, *28*(9), 598.

[8] Marozva, M., Rushambwa, M., & BukhelaniMusiiwa, P. (2025). Enhancing Water Quality Forecasting with Lstm and Attention-based Deep Learning: Toward Scalable, Long-term Monitoring of Surface Water Bodies. *IJSAT-International Journal on Science and Technology*, *16*(3).

[9] Sharafi, M., Behmanesh, J., & Rezaverdinejad, V. (2025). Smart Water Quality Monitoring Using a Hybrid Deep Learning Framework with Memory Network and Teaching–Learning-Based Optimization for Accurate LSI Prediction: M. Sharafi et al. *Water Resources Management*, 1-26.

[10] Yu, W., Wu, M., & Ha, Z. (2025). Research on a Pearson-LSTM-AM-based water quality prediction model for freshwater aquaculture. *Journal of the World Aquaculture Society*, *56*(4), e70041.

[11] Kunapinun, A., Fairman, W., Wills, P., & Ouyang, B. (2025, May). Integrating Bi-LSTM and physics constraint models for improved biomass and water quality prediction in the aquaculture systems. In *Machine Learning from Challenging Data 2025* (Vol. 13460, pp. 61-73). SPIE.

[12] Sudha, R., Damodaran, A., & Manohar, G. (2025). Enhanced air quality prediction using adaptive residual Bi-LSTM with pyramid dilation and optimal weighted feature selection. *Scientific Reports*, *15*(1), 30428.

[13] Nisharga, M. R., Narmadha, & Sundar, N. (2025, February). Predicting water quality parameters using LSTM models. In *AIP Conference Proceedings* (Vol. 3204, No. 1, p. 020010). AIP Publishing LLC.

[14] Khosravi, K., Farooque, A. A., Karbasi, M., Ali, M., Heddam, S., Faghfouri, A., & Abolfathi, S. (2025). Enhanced water quality prediction model using advanced hybridized resampling alternating tree-based and deep learning algorithms. *Environmental Science and Pollution Research*, 1-20.

[15] Al Mazid, M. A., Deng, L., & Rishe, N. (2025, March). Addressing Irregularities and Enhancing Forecasting for Water Quality Metrics by Integrating Ordinary Differential Equations with BiLSTM. In *SoutheastCon 2025* (pp. 1270-1275). IEEE.

[16] Bi, J., Li, Y., Yuan, H., Wang, M., Wang, Z., Zhang, J., & Zhou, M. (2025). Hybrid Water Quality Prediction with Multimodal Low-rank Fusion and Localized Attention. *IEEE Internet of Things Journal*.

[17] Xiong, S., Cui, J., & Hou, F. (2025). MVIE-LSTM: a deep learning-based method for water quality assessment using monthly river data. *Stochastic Environmental Research and Risk Assessment*, 1-16.

[18] Muñoz-Alegría, J. A., Núñez, J., Oyarzún, R., Chávez, C. A., Arumí, J. L., & Rodríguez-López, L. (2025). A Bibliometric-Systematic Literature Review (B-SLR) of Machine Learning-Based Water Quality Prediction: Trends, Gaps, and Future Directions. *Water*, *17*(20), 2994.

[19] Mohan, S., Kumar, B., & Nejadhashemi, A. P. (2025). Integration of machine learning and remote sensing for water quality monitoring and prediction: A review. *Sustainability*, *17*(3), 998.

[20] Arepalli, P. G., & Naik, K. J. (2025). Water quality classification framework for IoT-enabled aquaculture ponds using deep learning based flexible temporal network model. *Earth Science Informatics*, *18*(2), 351.

[21] Yu, Y., Zhao, S., Han, L., Peng, L., Xu, Y., Tian, Q., ... & Hu, Z. (2025). Cycleift: A deep transfer learning model based on informer with cycle fine-tuning for water quality prediction. *Stochastic Environmental Research and Risk Assessment*, 1-13.

[22] Huan, J., Fan, Y., Xu, X., Zhou, L., Zhang, H., Zhang, C., ... & Gu, S. (2025). 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*, *231*, 109985.

[23] Tan, X., Bai, Y., Yue, X., & Jia, X. (2025). A Mamba-based method for multi-feature water quality prediction fusing dual denoising and attention enhancement. *Journal of Hydrology*, 133424.

[24] Utku, E. D. B., Utku, A., & Kutlu, B. (2025). Enhancing water quality prediction with artificial intelligence: A hybrid convlstm model for spatio-temporal analysis. *International Advanced Researches and Engineering Journal*, *9*(2), 107-117.

[25] Rejini, K., Visumathi, J., Heltin Genitha, C., & Suthanthira Devi, P. (2025, April). Optimized Deep Learning Model for Water Quality Prediction. In *2025 International Conference on Data Science and Business Systems (ICDSBS)* (pp. 1-11). IEEE.