Deep Learning-Based Real-Time Water Quality Forecasting And Disaster Reporting System For The Ganga River Basin

Akshara Sharma, Nitya Pillai

Amity School Of Engineering And Technology,
Amity University, Noida, Uttar Pradesh

Abstract
The Ganga River is an essential source of water and livelihood for millions of people across India, yet it continues to face serious pollution problems that threaten both environmental sustainability and public health. To help address this issue, this study presents an AI-enabled decision support system designed to monitor and predict water quality conditions along the Ganga River in real time. The system uses an Exogenous Long Short-Term Memory (LSTM) model to predict key water quality parameters such as pH, dissolved oxygen, and biochemical oxygen demand, based on historical records and real-time data collected from monitoring stations along the river. The model demonstrates a prediction accuracy of 96.86 percent and is able to forecast water quality trends up to five days in advance, while effectively capturing seasonal variations, pollution levels, and the influence of external factors such as industrial discharge. Alongside the prediction model, the system includes a user-driven disaster reporting dashboard where pollution incidents, floods, and industrial spills can be reported for specific locations. These reports are analyzed and visualized as location-wise disaster hotspots, helping identify areas that require immediate attention. By combining predictive analysis with real-time reporting and alerts, the proposed system supports timely decision-making, strengthens pollution control efforts, and improves preparedness for environmental risks. The framework can also be adapted for other rivers facing similar challenges, contributing to more sustainable water resource management and better protection of public health.
Keywords: *Alert System, Forecasting, Ganga River, Exogenous LSTM, Pollution*

I. Introduction

The Ganga River is far more than just a river in India. Flowing for over 2,500 kilometres from the Himalayas to the Bay of Bengal, it supports the daily lives of millions of people by providing water for drinking, farming, and industry. For many, the Ganga also carries deep cultural and spiritual meaning, symbolising purity, faith, and renewal. Despite this immense importance, the health of the river has been steadily declining. Years of industrial discharge, agricultural runoff, and untreated domestic waste have severely polluted its waters, placing immense stress on the river's ecosystem and the communities that depend on it. In recent years, advances in artificial intelligence and machine learning have opened new possibilities for addressing these challenges. In particular, Exogenous Long Short-Term Memory (LSTM) models have shown strong potential for understanding complex, time-dependent environmental systems like river water quality. By learning from historical patterns while also incorporating external factors such as seasonal changes and pollution sources, these models enable more accurate and timely predictions of key water quality parameters. Such predictive capabilities make it possible not only to assess the current state of the river, but also to anticipate future risks, support informed decision-making, and strengthen efforts to protect both the Ganga and the millions of lives connected to it.

II. Literature Review

AI-based models, particularly deep learning architectures such as Long Short-Term Memory (LSTM) networks [1] and Prophet-based forecasting frameworks [2], have been widely applied to predict critical water quality indicators. For instance, Kogekar et al. [3] utilized time-series learning approaches to forecast the water quality of the Ganga River, while Ye et al. [4] demonstrated that LSTM-based hybrid models outperform conventional methods in water quality prediction. Additionally, Zhou et al. [5] employed the fruit fly optimization algorithm and particle filtering techniques to improve estimation accuracy under uncertain environmental conditions, and Gopi and Naik [6] used convolutional neural networks (CNNs) to analyse IoT-based water quality data. Kogekar et al. [7] further extended this work by developing a hybrid CNN–LSTM–SVR model for Ganga water quality forecasting and management. The emergence of hybrid AI models in water quality prediction has

been particularly promising, as these approaches combine the temporal learning capability of LSTM networks with complementary machine learning techniques to enhance predictive accuracy. Zhang et al. [8] employed decomposition-assisted deep learning models to achieve high precision in water quality prediction, while Krtolica and Medojević [9] demonstrated the superior performance of hybrid AI-based techniques. Mittal et al. [10], in a comprehensive review, examined multiple methodologies for water quality prediction, highlighting the strengths and limitations of various approaches. Meanwhile, Nanjundan et al. [11] and Tung and Yaseen [12] provided a broader overview of recent AI-driven advancements in this domain. More recently, Amudha and Lakshmi [13] applied a deep ensemble learning approach for predicting the water quality index, while Wang et al. [14] explored real-time AIoT-based systems for continuous freshwater monitoring.

Building upon these developments, the present study introduces a novel AI-enabled Decision Support System (DSS) that integrates an LSTM-based prediction model with IoT sensor data and historical water quality records. This work is distinct in its use of a multi-feature Exogenous LSTM framework capable of capturing both long-term temporal trends and sudden pollution spikes with high accuracy, achieving a prediction accuracy of 96.86%. Unlike earlier LSTM-based studies that rely on a limited number of input variables, the proposed approach incorporates eight key water quality indicators, including dissolved oxygen (DO), biochemical oxygen demand (BOD), pH, turbidity, and coliform levels, thereby providing a more comprehensive and practical understanding of water quality dynamics. This diverse feature set enables a holistic assessment of the Ganga River's water quality and improves predictive reliability compared to existing models.

The primary contribution of this work is the development of an advanced AI-powered Decision Support System (DSS) aimed at proactive monitoring and management of the Ganga River's water quality. At the core of the system is an Exogenous LSTM architecture that significantly enhances the prediction accuracy of essential water quality parameters such as DO, BOD, and pH. With an achieved accuracy of 96.86%, the model delivers reliable real-time predictions and clearly outperforms traditional approaches that typically depend on simpler algorithms and limited feature sets. A key strength of the proposed system is its ability to generate five-day water quality forecasts, enabling early warnings that help prevent pollution-related issues before they escalate. This forecasting capability is particularly valuable for managing industrial discharges and agricultural runoff, allowing authorities and stakeholders to take timely preventive actions to protect public health and the river ecosystem. In addition to predictive modelling, the system incorporates a dedicated disaster management dashboard that enables real-time reporting and visualization of pollution incidents, floods, and industrial spills across different river locations. These user-reported and system-detected events are mapped as location-wise disaster hotspots, assisting decision-makers in rapidly identifying high-risk zones and prioritizing interventions. The DSS also includes an integrated alert mechanism that automatically notifies relevant authorities whenever water quality parameters exceed permissible thresholds. Supported by a user-friendly, cloud-based dashboard, the system facilitates continuous monitoring and rapid response to emerging environmental threats. By integrating predictive analytics with disaster reporting, alerts, and visualization, this work offers a comprehensive and practical framework for long-term river management, environmental sustainability, and enhanced resilience against water-related disasters.

The paper is organized as follows: Section III details the proposed methodology; Section IV presents a comprehensive analysis of the results; and the final section concludes by summarizing the key findings and discussing future directions for practical application.

III. Methodology

A. Data Description

The dataset utilized in this study consists of comprehensive water quality measurements compiled and curated as part of our previously published work [15]. The dataset spans a period of four years, from 01/01/2020 to 01/09/2025, and contains daily records of key water quality parameters that offer critical insights into the condition and safety of aquatic ecosystems. These parameters capture both temporal variations and pollution dynamics across the river system, forming a reliable foundation for water quality analysis and prediction in this study. The key features of the dataset are as follows:

1) *Biochemical Oxygen Demand (BOD)*: This parameter quantifies the amount of organic matter in water that is subject to decomposition by microorganisms. High BOD values indicate significant levels of pollution, as organic matter consumes oxygen during the decomposition process, leading to oxygen depletion in aquatic environments. Measured in mg/L.

2) *Dissolved Oxygen*: DO is a critical indicator of water quality, as it represents the amount of oxygen available for aquatic organisms. Low DO levels can signal poor water quality and can be detrimental to fish and other aquatic life, often leading to hypoxic conditions. Measured in mg/L.

3) *Faecal Coliform & Streptococci:* These are key indicators of sewage contamination in water. The presence of these bacteria suggests the potential for microbial pathogens, indicating a risk to human health and the ecosystem. Measured in CFU/100 mL.

4) *Nitrate & Ammonium Nitrogen:* These nutrients are essential for aquatic plant growth but can become pollutants when present in elevated concentrations. Their presence often results from agricultural runoff, indicating potential nutrient loading in the water body. Measured in mg/L.

5) *Turbidity:* This parameter measures water clarity and is affected by the presence of suspended particles. High turbidity levels can suggest increased sedimentation or pollution, affecting aquatic habitats and photosynthesis in aquatic plants. Measured in NTU.

6) *pH*: This parameter indicates the acidity or alkalinity of water, which can significantly affect the biological and chemical processes in aquatic ecosystems. Extreme pH values can harm aquatic organisms and disrupt ecological balance. Measured in NTU.

7) *Rainfall*: Rainfall is an essential parameter for understanding hydrological cycles and can affect water quality by influencing runoff, sediment transport, and nutrient loading in water bodies. Rainfall is typically measured in millimeters (mm)

8) *Temperature*: This parameter is crucial as it affects the physical and biological processes in aquatic ecosystems. Temperature influences dissolved oxygen levels, metabolic rates of organisms, and the solubility of various substances in water. Water temperature is measured in degrees Celsius (°C).
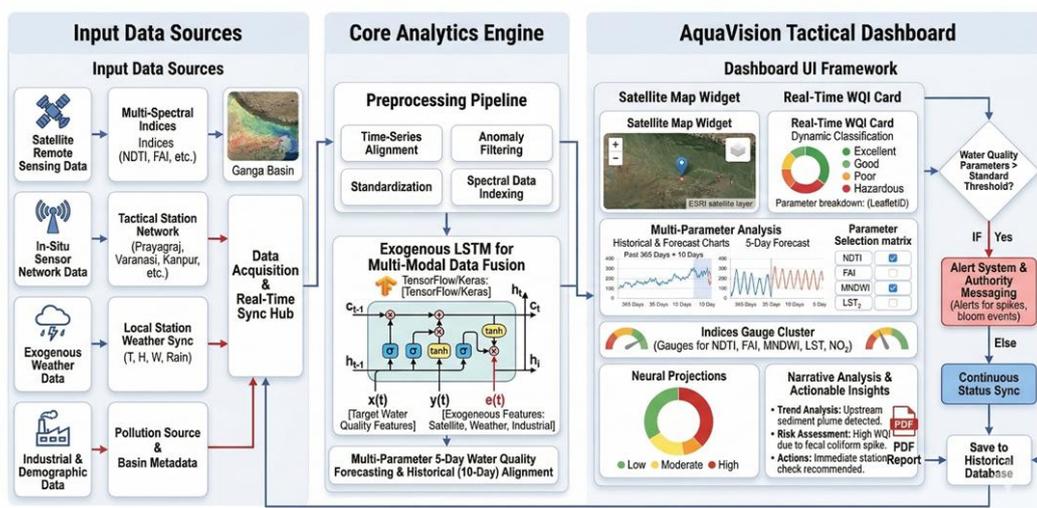

Fig.1. LSTM Model Architecture

B. Proposed Architecture

The proposed methodology for predicting water quality indicators along the Ganga River, as illustrated in Fig. 1, is based on an Exogenous Long Short-Term Memory (Exogenous LSTM) network. Unlike conventional LSTM models that rely solely on historical water quality data, the Exogenous LSTM integrates external environmental factors such as rainfall, temperature, seasonal variations, and anthropogenic activities, enabling a more realistic representation of river dynamics. The model architecture comprises an Exogenous LSTM layer followed by a fully connected output layer, where the LSTM learns long-term temporal dependencies while adjusting predictions according to external drivers. The tanh activation function is employed for updating the cell state, introducing non-linearity and ensuring stable training. Model optimization is carried out using the Adam optimizer with a learning rate of 0.001 and the mean squared error (MSE) loss function, which is well suited for continuous water quality prediction tasks. Training is conducted with a batch size of 16 over 100 epochs, while validation data are used to monitor performance and mitigate overfitting. This carefully designed architecture enables accurate real-time prediction and reliable multi-day forecasting of key water quality parameters, making it suitable for practical deployment in river monitoring and decision support systems.

In the Exogenous LSTM model the forget gate determines which information from the previous state should be discarded or retained for the current cell state, which is given by (3). The input gate of the Exogenous LSTM network is responsible for deciding which values from the input water quality parameters 'x' will be updated in the cell state, given by (4). In these equations, f_t is the forget gate output, σ represents the sigmoid activation function, W_f and W_c are the weight matrices for the forget and input gates, respectively, h_{t-1} is the previous hidden state, x_t is the input at the current time step, b_f and b_c are bias terms for the forget and input gates, respectively, and $\tilde{C}$ is the candidate cell state.

$$ft = \sigma(W_f[h_{t-1}, x_t, z_t] + b_f) \tag{3}$$

$$\tilde{C} = tanh\,tanh\,(W_c[h_{t-1}, x_t] + b_c)$$

(4)

While the candidate cell state represents the potential water quality parameter values that can be added to the cell state, based on the input and the previous hidden state values as shown in (5). The cell state of the Exogenous LSTM network is updated by combining the previous cell state and the newly scaled candidate values, given by (6). In these equations, $\tilde{C}_t$ is the candidate cell state, W_c is the weight matrix for the candidate state, h_{t-1} is the previous hidden state, x_t is the current input, b_c is the bias term for the candidate cell state, Ct is the updated cell state, f_t is the forget gate output, C_{t-1} is the previous cell state, and i_t is the input gate output.

$$C{\sim}t = tanh(Wc[h_{t-1}, x_t, z_t] + b_c)$$

5)

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

6)

IV.   Results & Discussion

The performance of the proposed system was rigorously evaluated using real-world data collected from historical records and real-time sensors, covering multiple critical water quality parameters. The forecasting performance of the model was assessed using standard statistical metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The detailed error analysis for each parameter is presented in Table 1. The evaluation metrics are defined as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}{\square}\left|\frac{y_i - \hat{y}}{y_i}\right| \times 100$$
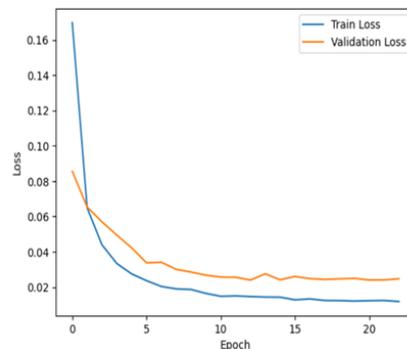
12)

Fig. 2 Model Loss Curve

Fig. 2 illustrates the model loss curve during training, indicating stable convergence and effective learning behaviour. The forecasting model demonstrates strong predictive performance across most parameters, with MAPE values consistently remaining below 10%. This reflects reliable accuracy in predicting key environmental indicators such as BOD, DO, turbidity, pH, nitrate, and total ammonium nitrogen. Although Faecal Coliform and Faecal Streptococci exhibit relatively larger absolute errors due to their high variability and wide numerical range, the low percentage errors indicate that the model remains effective when performance is evaluated on a relative scale.

The proposed Decision Support System (DSS), built around an Exogenous LSTM model and trained on four years of daily water quality data, achieves an average prediction accuracy of 96.86% across multiple monitoring locations along the Ganga River. This level of accuracy highlights the robustness of the model in capturing complex temporal patterns, seasonal variations, and external environmental influences affecting river water quality.

Fig. 3 Water quality meter for Faecal Coliform

Table 1. Results and Metrics			
Features	**MAE**	**MSE**	**MAPE (%)**
Biochemical Oxygen Demand (BOD)	0.15	0.03	3.1
Dissolved Oxygen (DO)	0.22	0.06	3.2
Faecal Coliform	480	230400	3.0
Faecal Streptococci	390	152100	3.1
Turbidity	5.4	29.2	2.9
pH	0.58	0.34	2.8
Nitrate	0.020	0.00040	3.0
Total Ammonium Nitrogen	0.0008	0.00000064	3.1

To enhance interpretability and usability, the DSS categorizes daily water quality into three predefined classes, Good, Neutral, and Bad, based on regulatory thresholds for each parameter. A water quality meter provides a clear visual summary of these classifications, making the results easily understandable for non-technical stakeholders. This classification framework enables rapid assessment of river health and supports informed decision-making by water resource managers. An example of the water quality meter for Faecal Coliform is shown in Fig. 3.

Beyond predictive analytics, the system integrates a dedicated disaster management dashboard that allows users to report pollution incidents, floods, and industrial spills at specific river locations. These user-generated and system-detected reports are analysed in real time and visualized as location-wise disaster hotspots, enabling rapid identification of high-risk areas that require immediate intervention. This functionality strengthens the practical utility of the DSS by linking predictive modelling with real-world incident reporting.

A key advantage of the proposed DSS is its capability to model the interconnectedness of monitoring stations by considering the downstream flow of contaminants. This is particularly important in a river system like the Ganga, where upstream pollution directly affects downstream water quality. By combining high-accuracy prediction, real-time disaster reporting, and intuitive visualization, the proposed framework enhances preparedness against environmental risks and supports sustainable water resource management and improved protection of public health

V.   Conclusion & Future Works

The proposed water quality forecasting system presents an innovative and comprehensive solution for the continuous monitoring and prediction of the Ganga River's water quality. By leveraging an advanced Exogenous LSTM model, the system delivers accurate and reliable forecasts for critical water quality parameters. The low error rates observed across multiple evaluation metrics validate the robustness and predictive capability of the proposed approach, making it a valuable tool for stakeholders responsible for maintaining river health. The system's real-time monitoring and forecasting capability provides decision-makers with an effective mechanism to anticipate water quality trends and implement informed, proactive interventions. Through the generation of five-day forecasts, the system enables early identification of potential pollution threats, supporting timely preventive actions to reduce environmental damage. Furthermore, the model's ability to capture the downstream flow of pollutants across monitoring stations ensures a holistic understanding of the river's overall condition.

In addition to predictive analytics, the integrated disaster management dashboard enhances the practical impact of the system by allowing users to report pollution incidents, floods, and industrial spills at specific river locations. These reports are analysed and visualized as location-wise disaster hotspots, enabling rapid identification of high-risk areas and supporting timely response measures. When combined with real-time forecasting and alert mechanisms, this dashboard strengthens pollution control efforts and improves preparedness for environmental emergencies.

Despite the high accuracy achieved by the Exogenous LSTM model, certain challenges remain. Parameters influenced by complex external factors, such as nitrate levels and faecal coliform concentrations, exhibit higher variability in predictions. This limitation can be addressed by incorporating additional contextual data, including land use patterns and detailed meteorological information, to further enhance model performance. Overall, the proposed DSS demonstrates strong potential for supporting water quality management in large river systems like the Ganga. Its combination of accurate forecasting, real-time monitoring, user-friendly visualization, alert mechanisms, and disaster reporting capabilities makes it a powerful tool for promoting long-term sustainability of water resources. Future enhancements may include extending the forecasting horizon and integrating finer spatial resolution data to improve the detection and prediction of localized pollution events.

Looking ahead, the system can be expanded to support simultaneous monitoring of multiple river systems, thereby increasing its environmental impact at a regional or national scale. The incorporation of more

advanced AI techniques and adaptive learning strategies would further improve predictive accuracy and strengthen proactive environmental management efforts.

References

[1] Xu, Rui, Qingyu Xiong, Hualing Yi, Chao Wu, And Jianxin Ye. "Research On Water Quality Prediction Based On SARIMA-SARIMAX: A Case Study Of Beilun Estuary." In 2019 IEEE 21st International Conference On High Performance Computing And Communications; IEEE 17th International Conference On Smart City; IEEE 5th International Conference On Data Science And Systems (HPCC/Smartcity/DSS), Pp. 2183-2188. IEEE, 2019. Doi: 10.1109/HPCC/Smartcity/DSS.2019.00302

[2] Papacharalampous, Georgia A., And Hristos Tyralis. "Evaluation Of Random Forests And Prophet For Daily Streamflow Forecasting." Advances In Geosciences 45 (2018): 201-208. Doi: 10.5194/Adgeo-45-201-2018

[3] Kogekar, A. P., Nayak, R., & Pati, U. C. (2021). Forecasting Of Water Quality For The River Ganga Using Univariate Time-Series Models. In 2021 8th International Conference On Smart Computing And Communications (ICSCC) (Pp. 52-57). IEEE. Doi: 10.1109/ICSCC51209.2021.9528216.

[4] Ye, Q., Yang, X., Chen, C., & Wang, J. (2019). River Water Quality Parameters Prediction Method Based On SARIMAX-RNN Model. In 2019 Chinese Control And Decision Conference (CCDC) (Pp. 3024-3028). IEEE. Doi: 10.1109/CCDC.2019.8832885

[5] Zhou, Y., Zhao, Z., Wang, X., & Sun, Q. (2020, July). Water Quality Evolution Mechanism Modeling And Estimation Based On Fruit Fly Optimization Algorithm And Particle Filter. In 2020 39th Chinese Control Conference (CCC) (Pp. 1178-1183). IEEE. Doi: 10.23919/CCC50068.2020.9188465

[6] Gopi, A. P., & Naik, K. J. (2021, December). A Model For Analysis Of Iot Based Aquarium Water Quality Data Using CNN Model. In 2021 International Conference On Decision Aid Sciences And Application (DASA) (Pp. 976- 980). IEEE. Doi: 10.1109/DASA53625.2021.9682251

[7] Kogekar, A. P., Nayak, R., & Pati, U. C. (2021, December). A CNN-Bisarimax-SVR Based Deep Hybrid Model For Water Quality Forecasting Of The River Ganga. In 2021 IEEE 18th India Council International Conference (INDICON) (Pp. 1-6). IEEE. Doi: 10.1109/INDICON52576.2021.9691532

[8] Zhang, D., Chang, R., Wang, H., Wang, Y., Wang, H., & Chen, S. (2021, May). Predicting Water Quality Based On EEMD And SARIMAX Networks. In 2021 33rd Chinese Control And Decision Conference (CCDC) (Pp. 2372- 2377). IEEE. Doi: 10.1109/CCDC52312.2021.9602800

[9] Krtolica, I., & Medojević, M. (2024, June). Enhancing Precision In Artificial Intelligence–Based Water Quality Prediction: The Advantages Of Hybrid Modeling Approaches–Review. In 2024 9th International Conference On Smart And Sustainable Technologies (Splitech) (Pp. 1-5). IEEE. Doi: 10.23919/Splitech61897.2024.10612379

[10] Mittal, A., Patwal, S., Adhikari, M., & Manu, M. (2023, August). A Review Of Various Water Quality Prediction Models And Techniques. In 2023 5th International Conference On Inventive Research In Computing Applications (ICIRCA) (Pp. 614-620). IEEE. Doi: 10.1109/ICIRCA57980.2023.10220687

[11] Nanjundan, P., George, J. P., & Vij, A. (2022, July). A Reliable Method Of Predicting Water Quality Using Supervised Machine Learning Model. In 2022 IEEE International Conference On Data Science And Information System (ICDSIS) (Pp. 1-5). IEEE. Doi: 10.1109/ICDSIS55133.2022.9915819

[12] Tung, T. M., & Yaseen, Z. M. (2020). A Survey On River Water Quality Modelling Using Artificial Intelligence Models: 2000–2020. Journal Of Hydrology, 585, 124670. Doi: 10.1016/J.Jhydrol.2020.124670

[13] Amudha, V., & Lakshmi, S. V. (2023, August). Deep Ensemble-Based Water Quality Index Prediction And Classification Using Randomized Low-Rank Approximation. In 2023 Second International Conference On Augmented Intelligence And Sustainable Systems (ICAISS) (Pp. 434-439). IEEE. Doi: 10.1109/ICAISS58487.2023.10250465

[14] Wang, Y., Ho, I. W. H., Chen, Y., Wang, Y., & Lin, Y. (2021). Real-Time Water Quality Monitoring And Estimation In Aiot For Freshwater Biodiversity Conservation. IEEE Internet Of Things Journal, 9(16), 14366-14374. Doi: 10.1109/JIOT.2021.3078166

[15] Pillai, Nitya, Akshara Sharma, Amisha Krishna Gupta, Anoushka Ishi Gupta, Pavel Sikora, Kamil Riha, And Malay Kishore Dutta. "AI-Driven Decision Support System For Real-Time Ganga River Water Quality Monitoring And Forecasting Using Iot Sensors." In ICUMT 2024; 16th International Congress On Ultra Modern Telecommunications And Control Systems And Workshops, Pp. 88-93. VDE, 2024."