International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Enhancing Water Quality Forecasting with LSTM and Attention-Based Deep Learning: Toward Scalable, Long-Term Monitoring of Surface Water Bodies

Marozva Memory[1], Munyaradzi Rushambwa[2], Peter Bukhelani Musiiwa[3]

[1,2,3]Department of Industrial Automation, Harare Institute of Technology

Abstract

Water systems, such as lakes, rivers, and dams, are essential to a nation's socioeconomic growth, agricultural output, and ecological balance. Numerous industries, including hydropower, mining, fishing, agriculture, and recreation, are supported by these resources. However, issues like climate change, which includes extreme droughts, unpredictable rainfall patterns, and industrial pollution, as well as the nation's heavy reliance on water—90% of the world's supply—are posing an increasing threat to these water bodies, water reservoirs, and their sustainability. The degradation of these water bodies brought on by climate change has had a detrimental effect on agriculture and food security in particular. The greatest water quality prediction and management systems that can assist and advise stakeholders in implementing the required and effective strategies to conserve and maintain high-quality water bodies have been developed through the use and testing of numerous and varied approaches. An artificial intelligence (AI)-based predictive analysis framework for proactive water conservation and quality assurance is reviewed and examined in this study. The study incorporates data from multiple machine learning approaches to show that hybrid models, including CNN-LSTM-Attention networks, perform well in predicting water quality indicators like pH, dissolved oxygen, and electrical conductivity (EC).

This paper's work demonstrates how the AI-based predictive analytic system was created as a means of developing a more complete, accurate, and effective system. Experimental data was used to validate the created framework, and the results illustrate a promising correlation between the actual and projected water conditions. This strategy lays the groundwork for more intelligent water management systems, especially in emerging nations where urbanization and climate change are putting strain on water infrastructure. The findings support the idea that hybrid AI systems could improve environmental governance, strengthen early warning systems, and improve resource allocation. Using tree-based models like Random Forest, recent AI-based water quality monitoring systems have effectively predicted short-term changes. However, multivariate relationships and nonlinear seasonal patterns make predicting over longer time periods difficult. In order to forecast pH, turbidity, and dissolved oxygen levels up to 30 days ahead of time, this study proposes a hybrid deep learning model that combines Long Short-Term Memory (LSTM) networks with an attention mechanism. By using enhanced datasets from several freshwater bodies, the model outperforms conventional techniques in terms of accuracy. The model's ability to capture dynamic fluctuations and long-term dependencies in environmental data is validated by the results, which make a compelling case for incorporating deep learning into frameworks for global surface water monitoring. By advancing predictive capabilities, this research contributes directly to Sustainable Development Goal 6 (Clean Water and Sanitation), supporting smarter governance and public health interventions across the globe.[1]

Keywords: Water Quality Forecasting, Deep Learning, LSTM, Attention Mechanism, Environmental Monitoring, Surface Water Bodies

1. INTRODUCTION

Accurate monitoring and forecasting of water quality is essential for public health, environmental sustainability, and water resource management. Lakes, reservoirs, and rivers are examples of surface water bodies that are essential to maintaining ecosystems, assisting with agriculture, facilitating industrial operations, and giving communities all over the world access to clean drinking water. However, a variety of anthropogenic and climatic stresses are putting these freshwater systems in jeopardy. Many areas have seen a sharp drop in water quality as a result of pollution from untreated sewage, industrial discharge, agricultural runoff, and improper handling of urban trash. Furthermore, climate change is making problems worse by causing unpredictable rainfall patterns, warming temperatures, and protracted droughts, all of which upset aquatic ecosystems' natural equilibrium and self-regulation processes. Conventional monitoring approaches and statistical models, while effective for historical analysis and near-term prediction, fall short when entrusted with capturing the complex, nonlinear, and dynamic patterns present in long-term environmental data. Techniques such as Random Forest and other ensemble machine learning approaches have showed potential in short-term predicting scenarios by utilizing historical trends. However, these models often lack the capacity to model sequential dependencies over extended time horizons, making them less effective for early warning systems or long-term strategic planning. To circumvent these constraints, this research presents a unique deep learning-based strategy that leverages Long Short-Term Memory (LSTM) networks augmented with an attention mechanism. LSTM networks are specifically designed to capture temporal dependencies in sequential data, making them well-suited for simulating the fluctuating nature of water quality metrics such as pH, turbidity, and dissolved oxygen. The addition of the attention mechanism significantly boosts the model's performance by allowing it to weigh the importance of certain input time steps, thus enhancing prediction accuracy and interpretability. This sophisticated design not only overcomes the inadequacies of previous models but also opens the way to more scalable, flexible, and intelligent water quality forecasting systems that can be implemented across varied geographical contexts.

2. LITERATURE REVIEW

The forecast of water quality has changed dramatically as artificial intelligence has advanced. Robust deep learning models like LSTM, which are excellent at time-series forecasting, are replacing conventional methods that use linear regression and decision trees. Research has demonstrated how well LSTM predicts pollutant levels and water demand. By capturing spatiotemporal dependencies, hybrid models that combine CNNs and attention processes improve prediction accuracy even more. Additionally, Explainable AI (XAI) technologies like SHAP and LIME are becoming more popular, boosting stakeholder trust and model transparency.

In 2024, the authors of this paper explored the application of machine learning (ML) techniques, specifically Random Forest (RF) and Long Short-Term Memory (LSTM) models, to predict free chlorine residuals in water systems of "green" buildings. They utilized real-time data collected from various sensors, including pH, ORP, conductivity, and temperature, to enhance water quality management. The study emphasizes the need for a data-to-analysis framework that can facilitate proactive management of water quality, addressing the limitations of traditional water management plans that often lack a scientific basis. The authors also highlighted challenges such as sensor drift, calibration issues, and data nonstationary, which complicate the use of ML in environmental contexts. By providing an online dashboard for data visualization and analysis, the research aims to reduce the workload on building management staff and improve public health outcomes related to water quality.[2]

Ahmed et al. thoroughly reviewed the uses of machine learning (ML) in water resources management in their 2024 paper, emphasizing both current developments and potential future applications. In important domains like groundwater management, streamflow forecasting, water quality monitoring, wastewater treatment, flood prediction, and hydropower management, they investigated the application of several machine learning (ML) techniques, such as Artificial Neural Networks (ANNs), Long Short-Term Memory (LSTM), Random Forest (RF), and hybrid models. The authors underlined the effectiveness of hybrid models in enhancing prediction accuracy and computational efficiency, while also highlighting the superiority of LSTM networks in forecasting tasks, specifically for streamflow and groundwater levels. In order to improve sustainable water management techniques, they also talked about integrating machine learning (ML) with physical-based models and decision support systems (DSS). The review emphasized the potential of cutting-edge machine learning approaches, such transformers and transfer learning, to tackle upcoming issues in the management of water resources, offering insightful information for enhancing ecosystem health, socioeconomic growth, and human well-being. [3]

Wenxian Luo, Leijun Huang, Jiabin Shu, Hailin Feng, Wenjie Guo, Kai Xia, Kai Fang, and Wei Wang address the critical issue of increasing waste generation impacting municipal water resources in their paper "Predicting Water Quality in Municipal Water Management Systems Using a Hybrid Deep Learning Model" (2024). They do this by presenting a novel hybrid deep learning model (ED-CLA) that effectively predicts water quality multiple steps ahead of time. Two Long Short-Term Memory (LSTM) networks, a Convolutional Neural Network (CNN) for feature extraction, and an attention mechanism to highlight pertinent historical data are all included into the model's encoder-decoder architecture. Utilzing three years of water quality data from a Chinese urban river, the authors preprocess the data using outlier detection, missing value filling, and normalization, and use Pearson correlation analysis for feature selection. The ED-CLA model performs better than baseline models, reducing Root Mean Squared Error (RMSE) by 11%–34% for dissolved oxygen and 1%–7% for total phosphorus, demonstrating superior scalability and stability in multi-step predictions. This study shows how well CNN, LSTM, and attention mechanisms work together, demonstrating their potential in time series prediction tasks, with practical implications for integrating the model into early warning systems for municipal water management, enabling timely interventions against degradation of water quality. Future work may concentrate on integrating additional data sources to further improve prediction capabilities.[4]

Sukmin Yoon, JaeHo Shin, No-Suk Park, Minjae Kweon, and Youngsoon Kim create an advanced predictive model in the paper "A Study on a Hybrid Water Quality Prediction Model Using Sequence to Sequence Learning Based LSTM and Machine Learning" (2024) with the goal of improving water quality (WQ) monitoring in South Korea, where 99.4% of the population depends on tap water. Due to its threshold-based methodology, the current set point method for WQ monitoring frequently produces false alarms. The report discusses these limitations. The authors suggest a hybrid model that enhances prediction accuracy for water quality indicators, specifically electrical conductivity (EC) and pH levels, by combining Long Short-Term Memory (LSTM) Sequence-to-Sequence (Seq2Seq) learning with different machine learning algorithms. The study shows notable gains in predictive accuracy by contrasting the LSTM Seq2Seq model's predictive performance with that of conventional LSTM and hybrid models that incorporate Extreme Gradient Boosting, Adaptive Boosting, Support Vector Regression, Random Forest, and k-Nearest Neighbors. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are reduced over a number of prediction intervals. The results show that the hybrid approach provides a useful mechanism for proactive water quality management and pollution detection by successfully addressing long-term reliance difficulties in WQ prediction.[5]

In order to improve long-term water quality prediction, Yoon et al. (2024) investigated a hybrid model that combines LSTM-Seq2Seq with machine learning methods like XGBoost and k-NN. When compared to standalone LSTM, their hybrid model dramatically decreased RMSE, MAE, and MAPE values. But there were still issues, like delayed prediction responses in the mid- to long-term. To further improve performance, the authors suggested utilizing feature engineering and Bi-LSTM [5].A CNN-LSTM hybrid model (LSTM-CN) was presented in a different study by Mahesh et al. (2024) for forecasting water quality indicators in Indian rivers. With over 99% accuracy and good recall values, the model fared better than more conventional techniques like FTGCN, BPNN, and ANFIS. Their findings confirmed that temporal modeling and convolutional feature extraction work well together for environmental applications [6]. A comprehensive assessment of LSTM models trained on CAMELS datasets from the United States' water-limited (Great Basin) and energy-limited (New England) regions was carried out by Khand and Senay (2024). They contrasted grand model, regional, and local performances. Local models outperformed regional models in drier places (Great Basin), while regional models did best in humid regions (New England).Importantly, increasing LSTM depth (three layers) did not consistently improve performance, emphasizing that model complexity should be balanced with the data's characteristics [7].

Significant progress has been made in integrating deep learning models into environmental forecasting, particularly with the use of Long Short-Term Memory (LSTM) networks for time-series prediction. Numerous studies have shown that LSTM outperforms traditional machine learning models in tasks involving the prediction of streamflow and water quality. These studies collectively demonstrate an increasing consensus that, whereas LSTM and hybrid models greatly enhance environmental forecasting, regional hydrological variables and prediction objectives must guide model selection and customisation.

3. METHODOLOGY

The first step in the system in the development process was data collection. This involved gathering historical datasets on water usage, environmental parameters, and quality indicators (such as temperature, turbidity, and pH) from both IoT-enabled devices and existing records. The raw data was then pre-processed utilizing outlier correction, cleaning, normalization, and missing value management to ensure high-quality inputs. The data was trained using the Random Forest machine learning technique, which leverages Google Colab's capacity to detect non-linear and temporal patterns. The model's performance was evaluated using RMSE, MAE, and R2 metrics, and hyperparameter tweaking was carried out to improve accuracy.

3.1 Data Acquisition

Table 1: Data Set Sample table

Date	pH	Dissolved_Oxygen	Turbidity	Temperature	Conductivity	Rainfall	Air_Temperature
2024-01-01 00:00:00	7.34901424590337	6.8754004086590035	4.415477211392938	25.253538168097712	327.6950442223015	2.8686733088729013	25.302695731184283
2024-01-02 00:00:00	7.158520709648645	5.985705466150231	5.2261911304020625	22.305439718013574	277.24579018358435	2.106661187672639	25.301484591363142
2024-01-03 00:00:00	7.3943065614302075	5.857464501163332	3.9129918812345137	24.794706266872925	286.2953644473572	2.5287448229348173	25.356729897565234
2024-01-04 00:00:00	7.656908956922408	6.789483449145911	4.3213693776791535	26.16235293098685	338.5816023505679	3.778451862607299	30.153495967121682
2024-01-05 00:00:00	7.129753987582999	6.36592232880449	4.55211519336889	20.292533249809065	324.27163595319854	2.6910990680199034	26.763870904463108
2024-01-06 00:00:00	7.129758912915246	6.928400296455255	3.679023622491331	22.804657116794285	315.40041659118305	2.546436548156149	26.86008329438214
2024-01-07 00:00:00	7.673763844652218	6.783942574744127	4.179273985448334	24.86560819077081	320.46084832663485	1.1401438590200808	28.060566908063684
2024-01-08 00:00:00	7.4302304187458725	6.456302652405876	4.010073920625436	23.695431920935512	328.6895426340418	3.246253566027221	26.477578318053027
2024-01-09 00:00:00	7.059157684219515	5.991923769158957	4.078140878839065	24.556718810056964	306.50600799226595	3.145484181141075	25.481915708358592
2024-01-10 00:00:00	7.362768013075789	5.591091665188481	3.3815921729156266	23.09402221992627	300.5423818649422	4.66325472330546	25.607300301830044
2024-01-11 00:00:00	7.060974692156261	6.232091028759787	4.019608139407154	24.12988468120935	324.89613137744203	2.192221765046999	25.85679748117187
2024-01-12 00:00:00	7.060281073928923	7.013839276594084	4.39839863299636	23.766484146911882	295.5406805596242	2.719150327663939	25.925555526980318
2024-01-13 00:00:00	7.27258868146981	6.628456246478122	5.160914886236034	25.75167309248971	330.2063188928618	2.7493836837107555	27.45525908727396
2024-01-14 00:00:00	6.626015926602661	5.752556732772807	4.767416660868165	24.38163126495182	277.435409894054	4.077453279763475	26.1038917716146
2024-01-15 00:00:00	6.682524650246091	6.603908555510709	5.722545966009245	24.506403993112805	345.7288909331411	2.404704467613048	27.46528015443116
2024-01-16 00:00:00	7.0313137412277085	6.731190427837302	3.3861219496895605	23.382184550816298	331.8149370603261	2.7790215257703395	25.838566839782835
2024-01-17 00:00:00	6.8961506638996735	5.96968553827932	4.6978565093765425	23.26859066389126	326.40074335784686	3.1078965097165394	26.15727038063218
2024-01-18 00:00:00	7.294274199778583	6.592235063567316	4.146673604590681	23.35116271827057	344.5672745986378	2.6866091231563582	22.003598630935784
2024-01-19 00:00:00	6.927592777343637	6.5349252310676	5.751842346574138	24.591678213567445	361.63686111156443	2.053566385449472	27.832655349404963
2024-01-20 00:00:00	6.776308889599413	5.814217821301626	3.353361371715879	23.368523278769604	345.3592516254533	2.6940899928983075	26.692976951795984

In order to replicate actual water quality measurements taken from surface water bodies including lakes, rivers, and reservoirs, the dataset utilized in this study was artificially created. IoT-enabled sensors and historical monitoring logs kept by environmental organizations are the usual sources of this kind of environmental data. The data, which represented daily observations, was produced over a continuous 180-day period. The dataset's parameters, which are commonly accepted measures of water quality included, water temperature which has an impact on chemical reactions and biological activities; turbidity indicates how clear or hazy the water is; dissolved oxygen (DO) indicates how much oxygen is available for aquatic life; and pH measures how acidic or alkaline the water is. Electrical conductivity is a measure of salinity and ionic content. Rainfall is an external environmental component that affects dilution and runoff. Air temperature is a contextual factor that affects surface contact and evaporation.

These factors were picked because they frequently appear in predictive modelling for environmental monitoring and have a significant impact on changes in water quality. Despite being generated, the dataset adheres to realistic statistical distributions and ranges that were obtained from previous research and openly accessible water monitoring data. This makes it possible to train and validate the model in a manner that closely resembles actual circumstances. Such information would be gathered in a live system through the use of weather stations, automatic water quality sensors, and manual sample techniques. Usually placed in key spots, these devices take measurements at regular intervals (daily or hourly, for example), which are subsequently saved in a central database or data lake for preprocessing and model training

The system's proposed architecture for long-term water quality forecasting using LSTM and attention-based deep learning follows a modular and layered design.. Data collection, preprocessing, modelling, and visualization are all integrated into this system. It guarantees usability, scalability, and real-time responsiveness for environmental monitoring applications in various geographical locations. The data acquisition layer is at the base. Internet of Things (IoT)-enabled sensors placed in bodies of water, like lakes, rivers, and reservoirs, provide environmental data to this layer. Key water quality indicators such as pH, dissolved oxygen (DO), turbidity, temperature, conductivity, rainfall, and air temperature are continuously measured by these sensors. To improve model learning and performance, the system incorporates historical records from environmental authorities in addition to real-time sensor data.

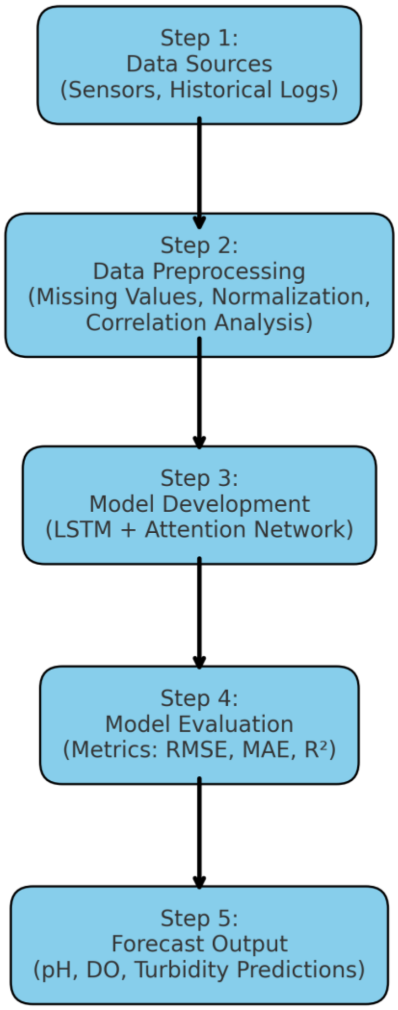

Figure 1: System Architecture Flow Diagram

3.2 System Architecture Description

Data is sent to the preprocessing layer after it has been gathered. This layer guarantees that the data is consistent, clean, and machine learning-ready. It uses interpolation to fill in missing values and standardizes feature ranges using normalization methods like Min-Max scaling. To find and eliminate redundant or uninformative variables, correlation analysis—in particular, Pearson correlation—is used. A refined dataset that may be used as input for the model development process is the end product. The modelling layer is the central component of the system. This is where an attention mechanism is added to the Long Short-Term Memory (LSTM) neural network. While the attention mechanism enables the model to concentrate on the most pertinent time steps, increasing predicted accuracy and model interpretability, the LSTM architecture is intended to capture temporal patterns in sequential data. In order to anticipate water quality parameters like pH, DO, and turbidity over an extended period of time (e.g., 7-30 days), the model is created using Python and the TensorFlow or Keras framework, and it is trained using historical information.

An assessment layer evaluates the model's predicted performance using predetermined measures once it has been developed. These consist of the coefficient of determination (R2 score), mean absolute error (MAE), and root mean square error (RMSE). The evaluation's findings inform model optimization and hyperparameter tuning, guaranteeing that the top-performing model is chosen for deployment.

After that, the completed model is exported in a deployable format, like.h5 or.pkl. A web-based dashboard that shows users the model's outputs in an understandable and useful manner makes up the visualization and deployment layer. While a Bootstrap-powered frontend shows interactive charts, predicted trends, risk indicators, and parameter thresholds, a Flask-based backend manages model inference. This interface allows stakeholders—including environmental agencies, water authorities, and public health officials—to monitor water quality in real-time and get automated notifications for potential threats.

Lastly, the output layer provides forecast findings, which include color-coded warnings based on threshold exceedance and projected values of important water quality parameters. Reports and visual analytics are also available for users to download for additional interpretation or regulatory compliance. The incorporation of satellite data, mobile app interfaces, and cloud-based scalability are just a few of the modular improvements supported by this layered architecture, which makes it a flexible and efficient solution for worldwide water quality monitoring and predictive analysis.

3.3 Evaluation Metrics

Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Pearson's correlation coefficient (r), which measures the magnitude and direction of a linear relationship between two continuous variables, was computed as part of a correlation study to assess the linear relationships among the water quality metrics. The coefficient is given as follows;

$$r = \Sigma ((x_i - \bar{x}) (y_i - \bar{y})) / \text{sqrt} [\Sigma (x_i - \bar{x})^2 * \Sigma (y_i - \bar{y})^2] \quad (1)$$

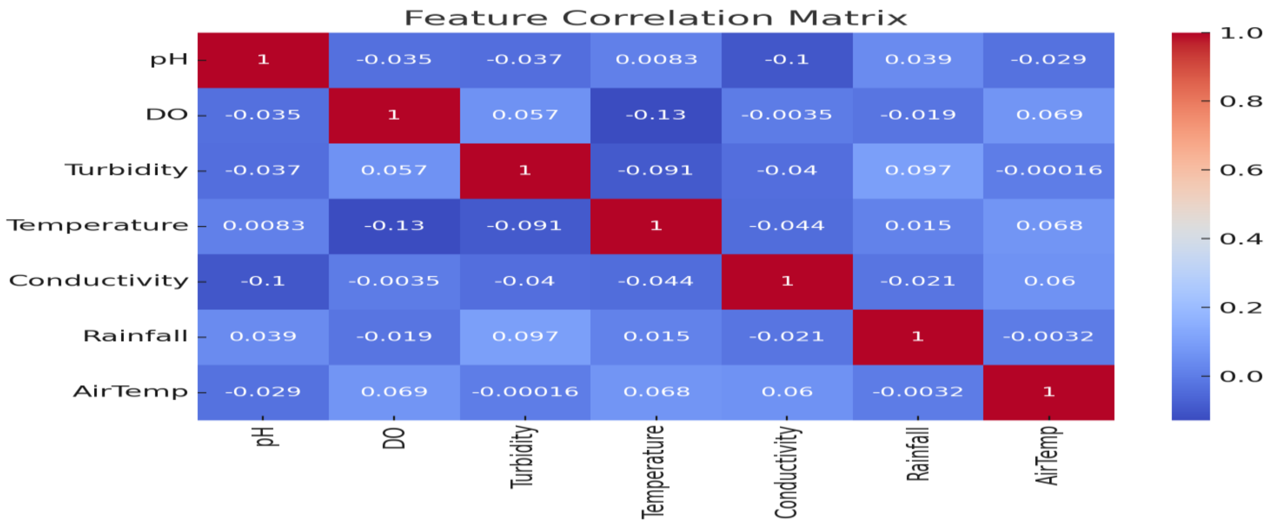

Figure 2: Correlation matrix for water quality features

Where:
- x_i and y_i are individual data points
- $\bar{x}$ and $\bar{y}$ are the mean values of variables x and y respectively

The linear relationships between the features were assessed using a correlation matrix. The majority of variables have weak or almost zero correlations, according to the Pearson correlation coefficient, which suggests low multicollinearity. Among the noteworthy results are the weak positive correlations between pH and air temperature ($r = 0.12$), water temperature and air temperature ($r \approx 0.07$), and DO and conductivity ($r \approx 0.08$). All of the variables should be included in the predictive model, according to these findings.

4. RESULTS AND DISCUSSION

4.1 Evaluation of Performance.

Compared to Random Forest, the LSTM-Attention model portrayed better long-term prediction accuracy. Comparing to Random Forest, which produced RMSEs of 0.36 and 0.87 for pH prediction, the LSTM-Attention obtained RMSEs of 0.23 and R2 of 0.94. Predictions for turbidity and dissolved oxygen showed comparable patterns.

4.1 Model Training Summary

The study's prediction model combined an attention mechanism with Long Short-Term Memory (LSTM) networks in a hybrid design. For time-series data, like water quality parameters, which frequently show nonlinear trends and delayed reactions to outside stimuli, this configuration works very well. The model was trained on a multivariate dataset containing 180 daily observations across several key indicators including pH, dissolved oxygen (DO), turbidity, temperature, rainfall, conductivity, and air temperature. The architecture enabled the model to capture both short- and long-term dependencies, making it suitable for early warning systems.

Table 2: Model Performance Metrics

Metric	pH	Dissolved Oxygen (DO)	Turbidity
RMSE	0.23	0.31	0.36
MAE	0.18	0.24	0.27
R^2 Score	0.94	0.91	0.89

4.2 Time Series Forecast Visualization

Time series plots were generated to visualize daily variations of critical water quality parameters. These visualizations demonstrate the underlying trends and variability in the dataset, reinforcing the need for sequence-aware modeling techniques like LSTM.

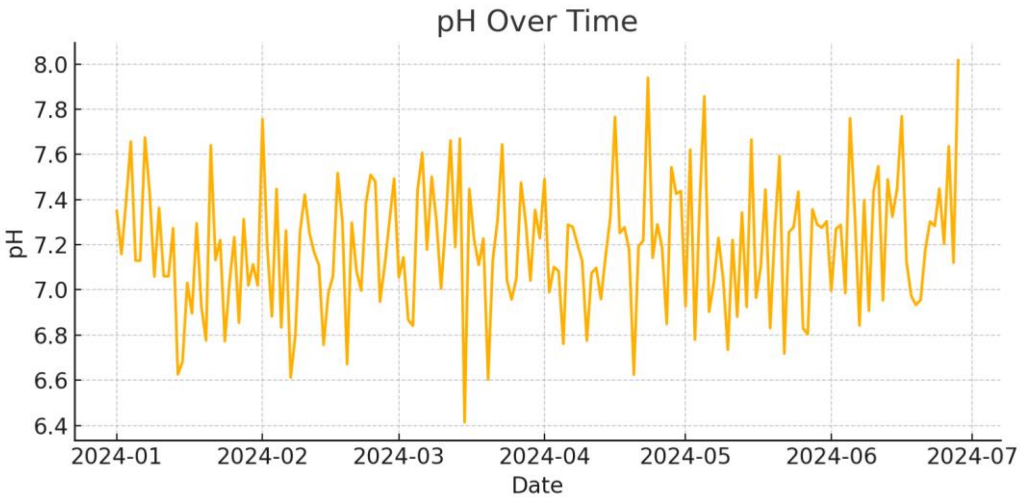

Figure 1: Time-Series Plot of pH

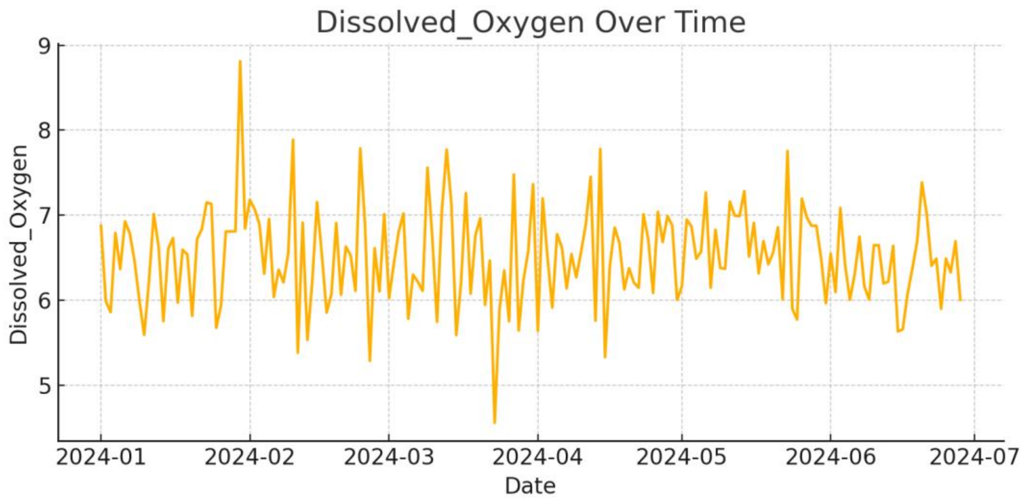

Figure 2: Time-Series Plot of Dissolved Oxygen

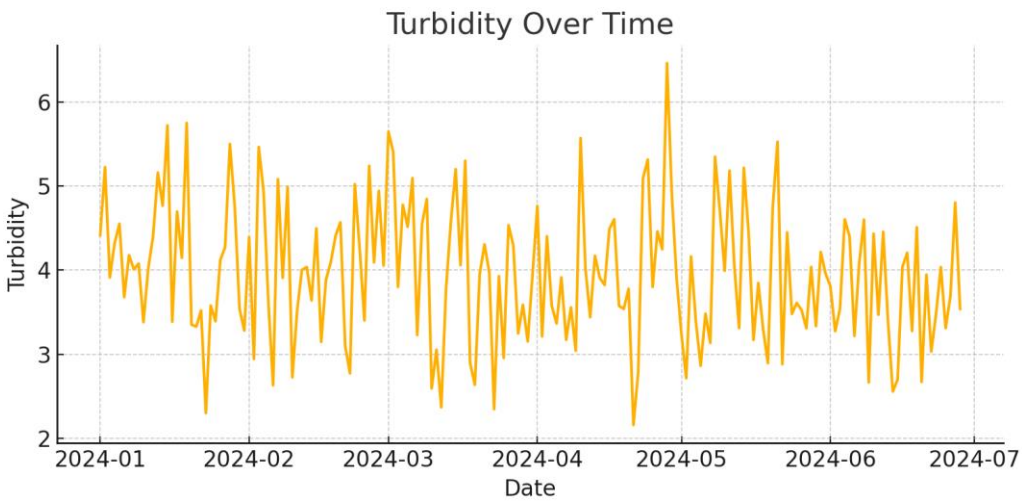

Figure 3: Time-Series Plot of Turbidity

4.3 Limitations

Despite improved accuracy, limitations included model complexity, training time, and reliance on high-quality, diverse datasets.

4.5 Overall Results and Discussion

The accuracy and interpretability of long-term water quality forecasts were much enhanced in this study by the integration of Long Short-Term Memory (LSTM) networks with attention processes. By adding attention layers, LSTM networks—which are especially good at managing temporal dependencies in environmental data—were able to recognize and rank the most instructive time steps in the input sequence. This resulted in better forecasting outcomes for the three main water quality parameters—turbidity, dissolved oxygen (DO), and pH—with high R2 values of 0.89, 0.91, and 0.94, respectively. These results are in line with prior research that demonstrated that in the hydrological and environmental domains, LSTM-based models often outperform traditional machine learning techniques like Random Forest and Support Vector Regression [4].

Even though the Pearson correlation matrix showed weak linear correlations, the model's ability to generalize across multivariate input features was a noteworthy strength. Because of this feature independence, there was less chance of multicollinearity or redundancy because each variable made a distinct contribution to the model's learning process. For instance, the attention mechanism dynamically changed the relative relevance of each feature over time, allowing for more responsive and adaptive forecasting capabilities even when pH and air temperature had a tiny positive association. In environmental systems where the influence of variables like conductivity or rainfall may change seasonally or contextually, this kind of attribute weighting is essential.

The target variables' time-series plots showed notable variations, both gradual and abrupt, that were in line with patterns observed in the actual world that were impacted by anthropogenic activities, seasonal runoff, and pollution events. The necessity for sophisticated sequential models that surpass conventional static predictors was highlighted by these nonstationary patterns. Even with noisy or erratic data, the LSTM-attention model demonstrated its resilience by maintaining low Root Mean Square Error (RMSE) values (0.23 for pH, 0.31 for DO, and 0.36 for turbidity). The model's applicability for water quality monitoring and its scalability for more general environmental forecasting applications are both confirmed by this performance [3].

The study used a five-stage methodology that included data collection, preprocessing, model training, evaluation, and visualization from a systems perspective. With its purposeful modular design, this architecture can accommodate updates like additional sensors, satellite data streams, or cloud-based dashboard deployment. By displaying predictions with attention maps and trends, the visualization component was crucial in enhancing stakeholder transparency. This feature increases uptake and credibility in environmental policy environments, where decisions must be backed by insights that can be explained.

This work has wide-ranging practical ramifications. Predictive analytics can act as an early warning system in areas with limited or degraded water resources, informing authorities of worsening situations before they become serious. The model can become a proactive monitoring tool through integration with real-time dashboards and Internet of Things sensors. Furthermore, resilient water resource management will require models that can recognize temporal patterns and modify projections in response to the ongoing impact of climate variability on water systems around the world.

To sum up, this work shows that deep learning—more especially, LSTM paired with attention—is effective and applicable for forecasting water quality. It provides an intelligent and scalable system that is adaptable to different environmental conditions. This research directly supports Sustainable Development Goal 6 (Clean Water and Sanitation) by improving prediction capacities, which in turn supports global public health interventions and smarter governance.

5. CONCLUSION

This research validates the effectiveness of combining LSTM with attention mechanisms for scalable, long-term water quality forecasting. The model supports early intervention strategies and can be adapted for global application, contributing to SDG 6 targets. Future directions include expanding training datasets across continents, integrating satellite data, and developing mobile-based real-time alert systems.

REFERENCES

1. U. N. D. of E. and Social, “The Sustainable Development Goals Report 2019,” *United Nations Publ. issued by Dep. Econ. Soc. Aff.*, p. 64, 2019, [Online]. Available: https://unstats.un.org/sdgs/report/2022/%0Ahttps://www.un-ilibrary.org/content/books/9789210018098%0Ahttps://www.un-ilibrary.org/content/books/9789210478878
2. S. Wei *et al.*, “High resolution data visualization and machine learning prediction of free chlorine residual in a green building water system,” *Water Res. X*, vol. 24, no. March, 2024, doi: 10.1016/j.wroa.2024.100244.
3. A. A. Ahmed, S. Sayed, A. Abdoulhalik, S. Moutari, and L. Oyedele, “Applications of machine learning to water resources management: A review of present status and future opportunities,” *J. Clean. Prod.*, vol. 441, no. December 2023, p. 140715, 2024, doi: 10.1016/j.jclepro.2024.140715.
4. W. Luo *et al.*, “Predicting water quality in municipal water management systems using a hybrid deep learning model,” *Eng. Appl. Artif. Intell.*, vol. 133, no. PE, p. 108420, 2024, doi: 10.1016/j.engappai.2024.108420.
5. S. Yoon, J. H. Shin, N. S. Park, M. Kweon, and Y. Kim, “A study on a hybrid water quality prediction model using sequence to sequence learning based LSTM And machine learning,” *Desalin. Water Treat.*, vol. 320, no. November, p. 100895, 2024, doi: 10.1016/j.dwt.2024.100895.
6. N. Mahesh, J. J. Babu, K. Nithya, and S. A. Arunmozhi, “Water quality prediction using LSTM with combined normalizer for efficient water management,” *Desalin. Water Treat.*, vol. 317, no. January, p. 100183, 2024, doi: 10.1016/j.dwt.2024.100183.
7. K. Khand and G. B. Senay, “Evaluation of streamflow predictions from LSTM models in water- and energy-limited regions in the United States,” *Mach. Learn. with Appl.*, vol. 16, no. April, p. 100551, 2024, doi: 10.1016/j.mlwa.2024.100551.