

An AI-Driven Environmental Informatics Framework for Real-Time Water Quality Monitoring and Anomaly Detection

Godfrey Perfectson

`godfrey.oise@physci.uniben.edu`

Wellspring University

Felix Oshiozenoya ULOKO

Veritas University

Immunhierokene Clinton OBRORINDO

Petroleum Training Institute

Roli Lydia OSHASHA

Petroleum Training Institute

Research Article

Keywords: Environmental Informatics, AI-Driven Monitoring Systems, Water Quality Monitoring, Anomaly Detection, Internet of Things (IoT), Time-Series Analysis, Water Quality Index (WQI), Decision Support Systems, Real-Time Data Analytics

Posted Date: April 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9237239/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

An AI-Driven Environmental Informatics Framework for Real-Time Water Quality Monitoring and Anomaly Detection

Godfrey Perfectson OISE, Department of Computer Science, University of Benin, Edo State
godfrey.oise@physci.uniben.edu, <https://orcid.org/0009-0006-4393-7874>.

Felix Oshioyenoye ULOKO, Department Computer Science, Veritas University, Abuja,
ulokof@veritas.edu.ng, Orcid ID: 0009-0009-5622-3645.

Immunhierokene Clinton OBRORINDO, Department of computer science and information technology,
Petroleum Training Institute, Effurun Delta State, Obrorindo_ci@pti.edu.ng

Roli Lydia OSHASHA, Department of computer science and information technology, Petroleum Training
Institute, Effurun Delta State, arenyekaroli@pti.edu.ng.

Corresponding author: Godfrey Perfectson OISE

Abstract

The degradation of freshwater resources caused by industrialization, urban expansion, and climate change necessitates intelligent, data-driven monitoring systems that address the limitations of conventional laboratory-based methods. This study proposes an AI-driven environmental informatics framework for real-time water quality monitoring and anomaly detection by integrating Internet of Things (IoT) sensor data with advanced machine learning models. The framework employs a dual-model architecture that combines a Long Short-Term Memory Autoencoder (LSTM-AE) for unsupervised anomaly detection and a Random Forest Regressor for Water Quality Index (WQI) prediction, forming a unified analytics pipeline for continuous environmental assessment. A synthetic multivariate time-series dataset of 10,000 records from five monitoring stations was used, incorporating parameters such as pH, dissolved oxygen, turbidity, temperature, electrical conductivity, and nitrate concentration. The LSTM-AE, trained exclusively on normal patterns using 24-hour sliding windows, achieved a high ROC-AUC score of 0.9763, demonstrating strong capability in identifying abnormal water quality events. Statistical features derived from the same sequences were used to train the Random Forest model, which achieved a Mean Squared Error (MSE) of 2.3573, indicating accurate WQI prediction. Comparative evaluation with an Isolation Forest baseline confirmed the superior performance of the proposed model, though challenges persist in detecting rare contamination events. The framework supports real-time data ingestion, preprocessing, and analysis, enabling early warning systems and reliable assessment. Its scalable design allows integration with monitoring infrastructures and decision-support platforms, supporting proactive water management, regulatory compliance, and sustainable governance.

Keywords: Environmental Informatics, AI-Driven Monitoring Systems, Water Quality Monitoring, Anomaly Detection, Internet of Things (IoT), Time-Series Analysis, Water Quality Index (WQI), Decision Support Systems, Real-Time Data Analytics

Introduction

Access to safe and reliable freshwater resources is fundamental to human health, ecological sustainability, and socio-economic development. However, accelerating pressures from industrialization, urban expansion, agricultural intensification, and climate variability have significantly degraded aquatic environments, increasing the complexity and unpredictability of water quality dynamics. These challenges are particularly pronounced in resource-constrained regions, where limited infrastructure, fragmented data systems, and weak regulatory mechanisms hinder effective monitoring and management [1]. Within this context, achieving

sustainable water governance requires not only accurate measurement but also continuous, data-driven situational awareness supported by intelligent analytical systems.

Conventional water quality monitoring paradigms are predominantly based on manual sampling and laboratory-based analysis [2]. While these approaches provide high-precision measurements, they are inherently constrained by low temporal frequency, limited spatial coverage, high operational costs, and delayed feedback cycles. Such limitations render them inadequate for capturing the rapidly evolving conditions of dynamic water systems, where contamination events can emerge and propagate within short time intervals [3]. From an informatics perspective, these traditional methods lack the capability to support real-time data acquisition, continuous processing, and adaptive decision-making, all of which are essential for proactive environmental management [4]. Recent advances in environmental informatics have enabled the integration of distributed sensing technologies with intelligent data analytics. The convergence of Internet of Things (IoT) infrastructures and Artificial Intelligence (AI) has facilitated the development of real-time monitoring systems capable of ingesting, processing, and interpreting high-frequency multivariate data streams [5]. IoT-enabled sensor networks provide continuous observations of key physicochemical parameters, while machine learning models offer the ability to extract latent patterns, detect anomalies, and generate predictive insights [6]. This shift from episodic measurement to continuous data-driven monitoring represents a fundamental transformation in how environmental systems are observed and managed [7].

Despite these advancements, several critical challenges remain. Existing approaches often treat anomaly detection and water quality prediction as independent tasks, resulting in fragmented analytical pipelines that limit their operational effectiveness [8]. Furthermore, the inherent imbalance in environmental datasets where abnormal contamination events are rare relative to normal conditions poses significant challenges for reliable anomaly detection. Many models also exhibit limited generalizability across different environmental contexts due to variations in hydrological, climatic, and pollution characteristics [9]. Additionally, the need for low-latency processing in real-time applications introduces constraints on computational efficiency, particularly in edge or resource-limited deployment environments. These gaps highlight the need for integrated, scalable, and context-aware informatics frameworks that unify detection, prediction, and decision support within a single architecture [10]. In response to these challenges, this study proposes an AI-driven environmental informatics framework for real-time water quality monitoring and anomaly detection. The framework adopts a dual-model architecture that combines a Long Short-Term Memory Autoencoder (LSTM-AE) for unsupervised anomaly detection with a Random Forest-based regression model for Water Quality Index (WQI) prediction. The LSTM-AE is designed to learn normal temporal patterns from multivariate time-series data and identify deviations indicative of potential contamination events [11], while the Random Forest model leverages statistical feature representations to provide accurate and interpretable WQI estimates [12]. By decoupling detection and prediction tasks within a unified pipeline, the framework enhances sensitivity to rare anomalies while maintaining high predictive accuracy.

The proposed system is evaluated using a multivariate time-series dataset representing IoT-based water quality observations across multiple monitoring stations. Experimental results demonstrate strong performance in both anomaly detection and predictive modeling, with high discriminative capability and low estimation error. The architecture is designed to support real-time data ingestion, preprocessing, and analytics, enabling continuous monitoring, early warning generation, and integration with decision-support systems. Its modular and scalable design allows deployment across diverse environments, ranging from urban water treatment facilities to remote monitoring scenarios. Generally, this research advances the field of environmental informatics by developing an integrated AI-driven framework that unifies anomaly detection and water quality prediction within a real-time, data-centric monitoring system. The study demonstrates how

the combination of temporal deep learning and ensemble regression techniques can address key challenges in environmental data analysis, including class imbalance, real-time processing, and system scalability. By enabling continuous environmental intelligence and supporting proactive decision-making, the proposed framework contributes to the development of sustainable, technology-enabled water resource management systems.

Related Works

Advances in environmental informatics have significantly reshaped water quality monitoring by enabling the integration of intelligent data analytics with distributed sensing infrastructures. The convergence of Artificial Intelligence (AI), Machine Learning (ML), and Internet of Things (IoT) technologies has facilitated the transition from static, laboratory-based assessment toward continuous, data-driven environmental intelligence [13]. Within this evolving landscape, existing research can be broadly categorized into four interrelated domains: anomaly detection in environmental time-series data, predictive modeling of water quality indices, IoT-enabled sensing and data acquisition systems, and hybrid frameworks that integrate detection and prediction capabilities [14]. Research on anomaly detection has primarily focused on unsupervised and semi-supervised learning approaches designed to identify deviations from normal environmental behavior without requiring extensive labeled datasets [15]. Techniques such as autoencoders, isolation-based methods, and clustering algorithms have been widely applied to detect irregular patterns in multivariate water quality data [16]. These approaches are particularly suited to environmental contexts characterized by data imbalance, where contamination events are rare relative to normal observations [17]. However, despite their effectiveness in modeling normal system dynamics, many of these methods exhibit reduced sensitivity to rare anomalies, often resulting in high false-negative rates and limited reliability in critical monitoring scenarios.

In parallel, predictive modeling of water quality has advanced through the application of regression-based machine learning algorithms, ensemble methods, and deep learning architectures [18]. Neural network-based models, including recurrent and feedforward architectures, have demonstrated strong capability in capturing complex nonlinear relationships among physicochemical parameters and in forecasting water quality indicators [19]. Ensemble learning approaches, particularly those based on decision tree aggregation, have shown robustness in handling heterogeneous environmental datasets and improving predictive accuracy [20]. More recently, transformer-based architectures have emerged as a powerful alternative for long-term forecasting, leveraging attention mechanisms and sequence segmentation strategies to model temporal dependencies with high computational efficiency [15]. Comparative evaluations across multiple monitoring sites have shown that such architectures can achieve superior performance, particularly for extended prediction horizons [21]. Comprehensive reviews of deep learning applications in water quality management highlight the growing importance of data preprocessing, model optimization, and hybridization techniques in improving predictive performance [22]. These studies emphasize the role of AI-driven systems in supporting sustainable development objectives, particularly in enhancing water resource management, reducing environmental risks, and enabling data-informed policy decisions [15]. Hybrid modeling approaches, which combine multiple algorithms or integrate statistical and machine learning techniques, have been shown to outperform individual models, particularly in tasks such as dissolved oxygen prediction and multivariate water quality forecasting [23]. These findings underscore the increasing reliance on integrated analytical strategies to address the complexity of environmental systems. Beyond predictive modeling, artificial neural networks have been extensively applied in simulating and optimizing water and wastewater treatment processes [24]. These models have demonstrated strong capability in representing both linear and nonlinear system dynamics, enabling accurate prediction of treatment performance and supporting operational decision-making [15].

However, practical implementation challenges persist, including limited data availability, suboptimal model design, lack of continuous system updates, and issues related to reproducibility and scalability. Classification-based approaches have also been explored for categorizing water quality into potable and non-potable classes, though their performance remains moderate in many cases, indicating the need for more robust modeling strategies [25], [26]. Recent developments have also focused on enhancing real-time monitoring through the integration of AI with signal processing and filtering techniques. Systems that combine neural networks with filtering mechanisms have demonstrated improved accuracy and reduced false alarm rates, enabling rapid decision-making in time-sensitive scenarios [27], [28]. Similarly, IoT-based monitoring and control systems have been developed to regulate environmental conditions in specialized applications such as aquaculture, where maintaining optimal water quality parameters is critical [29]. These systems enable continuous measurement, automated control, and adaptive responses, highlighting the potential of intelligent sensing infrastructures in supporting domain-specific environmental management.

The application of AI in water treatment and resource management has also been explored from a broader systems perspective, encompassing process optimization, efficiency prediction, and performance evaluation. While these approaches demonstrate significant potential, challenges related to model interpretability, data requirements, and integration with existing operational frameworks remain key barriers to adoption [30]. Additionally, the importance of data governance, interdisciplinary collaboration, and ethical considerations has been emphasized as critical factors in the successful deployment of AI-driven environmental systems [31]. Despite substantial progress across these domains, several persistent limitations can be identified. First, many existing systems treat anomaly detection and predictive modeling as independent tasks, resulting in fragmented analytical pipelines that limit their effectiveness in real-time monitoring scenarios. Second, the inherent imbalance in environmental datasets continues to undermine the performance of anomaly detection models, particularly in detecting rare but high-impact contamination events [32]. Third, the generalizability of models across different environmental contexts remains limited, as variations in hydrological, climatic, and anthropogenic factors introduce significant uncertainty. Finally, practical deployment challenges, including latency constraints, computational limitations in edge environments, sensor drift, and data quality issues, hinder the scalability and reliability of current solutions.

These limitations highlight the need for integrated environmental informatics frameworks that unify anomaly detection and predictive modeling within a single, scalable architecture. In response, the present study proposes a dual-model AI-IoT pipeline that combines a Long Short-Term Memory Autoencoder for anomaly detection with a feature-engineered Random Forest model for Water Quality Index prediction. By separating yet coordinating detection and prediction tasks, the framework enhances sensitivity to rare anomalies while maintaining high predictive accuracy [33]. Furthermore, its design supports real-time data ingestion, low-latency processing, and deployment across diverse environmental settings, positioning it as a robust and practical solution for next-generation water quality monitoring and decision-support systems.

3. Methodology

This study proposes an AI-driven environmental informatics framework for real-time water quality monitoring, integrating Internet of Things (IoT) sensing with advanced machine learning techniques. The framework introduces a dual-model architecture that combines unsupervised temporal anomaly detection with supervised predictive modeling within a unified analytical pipeline. Unlike conventional approaches that treat these tasks independently, the proposed system enables coordinated analysis, improving detection sensitivity, predictive accuracy, and real-time applicability in resource-constrained environments.

3.1 Data Acquisition and Preprocessing

To emulate a real-world IoT-enabled monitoring environment, a synthetic multivariate time-series dataset

comprising 10,000 hourly observations was generated from five distributed monitoring stations (ST1–ST5). Each station records six key physicochemical parameters: temperature, pH, dissolved oxygen, turbidity, electrical conductivity, and nitrate concentration[34] . These variables collectively capture the dynamic behavior of aquatic systems and key indicators of water contamination.

The dataset reflects realistic environmental conditions, including inherent class imbalance, with 9,445 normal observations and 555 anomalous events. To ensure numerical stability and facilitate model convergence, all features were standardized using z-score normalization. The normalization parameters were retained to ensure consistency during deployment and real-time inference.

3.2 Anomaly Detection Using LSTM Autoencoder

An unsupervised Long Short-Term Memory Autoencoder (LSTM-AE) was employed to detect anomalies in multivariate time-series data. The model is designed to learn the temporal patterns of normal water quality conditions and identify deviations through reconstruction error.

Input data were structured using a 24-hour sliding window, where each sample represents a sequence of sensor readings across multiple time steps. The LSTM-AE architecture consists of an encoder that compresses the input sequence into a low-dimensional latent representation and a decoder that reconstructs the original sequence. Training was performed exclusively on normal data to enhance sensitivity to abnormal patterns.

Anomaly scores were computed as the average reconstruction error for each sequence, with higher errors indicating greater deviation from learned normal behavior. This temporal modeling approach improves detection capability compared to conventional methods by capturing sequential dependencies and subtle pattern changes.

3.3 WQI Prediction Using Random Forest Regression

To complement anomaly detection, a Random Forest Regressor was employed for Water Quality Index (WQI) prediction. Rather than directly using raw time-series inputs, statistical feature engineering was applied to each 24-hour sequence. Specifically, the mean and standard deviation of each parameter were computed, resulting in a 12-dimensional feature vector that captures both central tendency and variability.

The Random Forest model maps these features to corresponding WQI values, leveraging ensemble learning to model nonlinear relationships and improve predictive robustness. Model hyperparameters were optimized to ensure reliable performance across varying environmental conditions.

3.4 Baseline Model for Comparative Analysis

To evaluate the effectiveness of the proposed anomaly detection approach, an Isolation Forest model was implemented as a baseline. This unsupervised algorithm identifies anomalies by isolating observations through recursive partitioning, making it suitable for high-dimensional and imbalanced datasets.

The baseline model provides a reference for assessing the relative performance improvements achieved by the LSTM-AE in detecting rare and complex anomalies.

3.5 Performance Evaluation Metrics

The performance of the proposed framework was assessed using task-specific evaluation metrics. For anomaly detection, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), precision, recall, and F1-score were used to evaluate classification performance under imbalanced conditions. For WQI prediction, Mean Squared Error (MSE) was employed to quantify regression accuracy and model reliability.

3.6 Integrated Framework and Deployment Considerations

The proposed methodology integrates anomaly detection and predictive modeling within a unified environmental informatics pipeline. While the two models operate independently, they share a common preprocessing and data handling framework, enabling efficient and simultaneous analysis.

This modular yet cohesive design enhances robustness, reduces computational overhead, and supports real-

time deployment in IoT-enabled monitoring systems. Furthermore, the framework is scalable and interoperable, facilitating integration with existing environmental monitoring infrastructures and decision-support systems for proactive water resource management.

Results

The experimental evaluation of the proposed dual-model AI framework demonstrates strong and consistent performance across both anomaly detection and predictive modeling tasks, validating the effectiveness of integrating temporal deep learning with ensemble-based regression within a unified environmental informatics pipeline. The Long Short-Term Memory Autoencoder (LSTM-AE) achieved a ROC-AUC score of 0.9763, indicating excellent discriminative capability in distinguishing normal from anomalous water quality patterns. This high performance can be attributed to the model's ability to learn latent temporal representations from multivariate sequences using a 24-hour sliding window mechanism, as defined in the methodology. By training exclusively on normal operational data, the model effectively captures baseline system dynamics, enabling robust detection of deviations under highly imbalanced conditions (approximately 94.45% normal vs. 5.55% anomalous samples).

The reconstruction error-based anomaly scoring mechanism produced a well-separated distribution between normal and abnormal sequences, facilitating reliable threshold selection. Empirical observations from the ROC curve further confirm that the LSTM-AE maintains a favorable balance between true positive and false positive rates, positioning it as a suitable candidate for real-time anomaly detection in streaming IoT environments. In contrast, the Isolation Forest baseline achieved a relatively high overall accuracy of 94%, but exhibited significantly lower recall (0.14) for the minority (anomalous) class. This imbalance highlights a critical limitation of partition-based methods in highly skewed environmental datasets, where model bias toward the majority class undermines sensitivity to rare but critical contamination events. The macro-average F1-score of 0.55 further reinforces this limitation. Comparative analysis therefore confirms that the LSTM-AE provides superior robustness and reliability for anomaly detection in environmental time-series data.

For the regression task, the Random Forest Regressor achieved a Mean Squared Error (MSE) of 2.3573, demonstrating high predictive accuracy in estimating the Water Quality Index (WQI). This performance reflects the effectiveness of the adopted feature engineering strategy, where statistical descriptors (mean and standard deviation) extracted from each temporal window provide a compact yet informative representation of system dynamics. The strong alignment between predicted and actual WQI values, as observed in the regression plots, indicates that the model successfully captures nonlinear relationships among physicochemical parameters. Furthermore, the low MSE value suggests minimal deviation from ground truth, supporting the reliability of the model for decision-support applications such as regulatory compliance monitoring and early warning systems. Analysis of the training process reveals rapid convergence of the LSTM-AE model. Both training and validation losses decreased significantly within the first 10 epochs and stabilized thereafter at approximately 0.045 MSE, indicating efficient learning and strong generalization capability. The close alignment between training and validation curves suggests minimal overfitting, which is critical for deployment in real-world streaming environments where unseen data distributions may occur. The combined performance of the dual-model framework highlights the advantages of decoupling anomaly detection and prediction tasks while maintaining a unified processing pipeline. The LSTM-AE provides high-sensitivity detection of abnormal events, while the Random Forest model delivers accurate and interpretable WQI estimates. This complementary interaction enhances overall system robustness and enables both qualitative (event detection) and quantitative (index prediction) assessment of water quality.

Table 1 summarizes the core performance outcomes:

LSTM Autoencoder (Anomaly Detection): ROC-AUC = 0.9763

Random Forest Regressor (WQI Prediction): MSE = 2.3573

These results demonstrate that the proposed framework outperforms baseline approaches in handling imbalanced data and complex temporal dependencies, while maintaining computational efficiency suitable for real-time deployment.

The comparative analysis further reinforces the superiority of the proposed approach. While the Isolation Forest model performs adequately in identifying normal operational states (precision = 0.95, recall = 0.99), its inability to effectively detect anomalies limits its applicability in safety-critical monitoring systems. Conversely, the LSTM-AE achieves a more balanced performance profile, ensuring that rare but high-impact contamination events are not overlooked. Similarly, the Random Forest model demonstrates improved predictive stability compared to conventional regression techniques, highlighting the importance of ensemble learning in environmental data analytics. The results confirm that temporal deep learning models are highly effective for unsupervised anomaly detection in multivariate environmental time-series data, while statistical feature engineering combined with ensemble learning provides robust and interpretable Water Quality Index (WQI) prediction. Furthermore, the integration of both models within a unified analytical pipeline significantly enhances system-level performance, scalability, and real-time applicability. Collectively, these findings validate the proposed framework as a reliable, efficient, and scalable solution for next-generation AI-driven water quality monitoring systems.

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 24, 6)	0
lstm (LSTM)	(None, 16)	1,472
repeat_vector (RepeatVector)	(None, 24, 16)	0
lstm_1 (LSTM)	(None, 24, 6)	552
time_distributed (TimeDistributed)	(None, 24, 6)	42

Total params: 2,066 (8.07 KB)
Trainable params: 2,066 (8.07 KB)
Non-trainable params: 0 (0.00 B)

Figure 1: Training Parameters

Figure 1 depicts the LSTM Autoencoder model processes 24-time-step sequences with 6 features each, encoding them into a 16-dimensional latent representation and then reconstructing the original sequence. It begins with an input layer (no parameters), followed by an encoder LSTM layer with 1,472 parameters that compresses the input, a Repeat Vector that restores the time dimension, a decoder LSTM layer with 552 parameters that reconstructs the sequence, and a Time Distributed dense layer with 42 parameters that outputs the final 24×6 sequence. The network has a total of 2,066 trainable parameters and is lightweight, making it well-suited for efficient time-series anomaly detection in water quality monitoring.

Training of the Model

The model was trained for 30 epochs with a batch size of 64, using the Mean Squared Error (MSE) loss and the Adam optimizer (learning rate = 0.001). Validation split was set to 10% to monitor overfitting.

Figure 2: Training Loop for the first Epochs

```
Epoch 1/30  
132/132 ————— 7s 20ms/step - loss: 0.5001 - val_loss: 0.2811  
Epoch 2/30  
132/132 ————— 2s 19ms/step - loss: 0.2133 - val_loss: 0.1345  
Epoch 3/30  
132/132 ————— 3s 23ms/step - loss: 0.1292 - val_loss: 0.1049  
Epoch 4/30  
132/132 ————— 2s 16ms/step - loss: 0.1025 - val_loss: 0.0870  
Epoch 5/30  
132/132 ————— 2s 16ms/step - loss: 0.0859 - val_loss: 0.0766  
Epoch 6/30  
132/132 ————— 2s 16ms/step - loss: 0.0765 - val_loss: 0.0708  
Epoch 7/30  
132/132 ————— 3s 16ms/step - loss: 0.0705 - val_loss: 0.0661  
Epoch 8/30  
132/132 ————— 4s 31ms/step - loss: 0.0667 - val_loss: 0.0611  
Epoch 9/30  
132/132 ————— 3s 17ms/step - loss: 0.0601 - val_loss: 0.0555  
Epoch 10/30  
132/132 ————— 2s 16ms/step - loss: 0.0557 - val_loss: 0.0520
```

Figure 2 shows the training log shows the learning progress of the LSTM Autoencoder over the first 10 of 30 epochs. The model's training loss decreases sharply from 0.5001 in the first epoch to 0.0557 by the tenth epoch, while the validation loss drops from 0.2811 to 0.0520 over the same period. This consistent reduction in both training and validation losses indicates effective learning and good generalization to unseen data, with no signs of overfitting in the early stages. The rapid convergence within the first few epochs suggests that the model quickly captures the underlying temporal patterns of the water quality time-series data making it suitable for accurate anomaly detection.

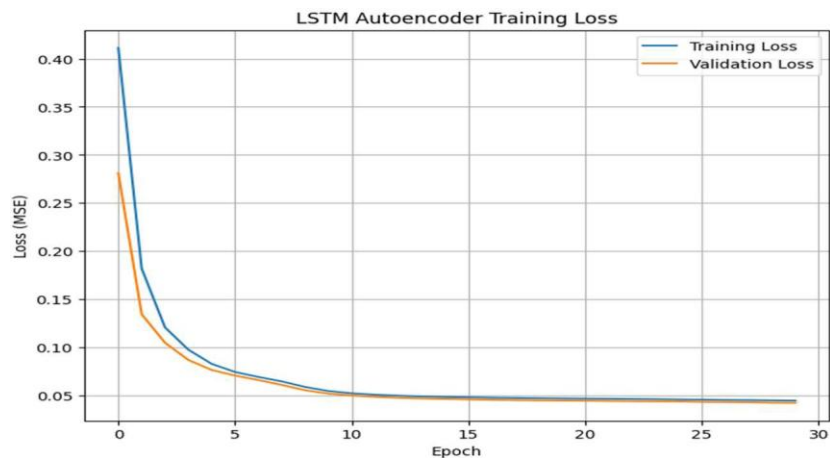


Figure 3: LSTM Autoencoder Training Loss

Figure 3 depicts the LSTM Autoencoder training loss curve shows a rapid decrease in both training and validation losses during the initial epochs, indicating quick learning of data patterns. After approximately the 10th epoch, the losses stabilize at around 0.045 MSE, reflecting effective convergence. The close alignment between training and validation loss throughout the process suggests that the model generalizes well to unseen data (Wu, 2020), with no significant signs of overfitting.

Anomaly Scoring

Following model training, anomaly detection was performed by evaluating the reconstruction error generated by the LSTM Autoencoder for each 24-hour multivariate sequence. Since the model was trained exclusively on normal operating conditions, it learns to accurately reconstruct typical temporal patterns; consequently, sequences that deviate from this learned behavior produce higher reconstruction errors and are flagged as

potential anomalies. The anomaly score for each sequence is computed as the Mean Squared Error (MSE) between the original input sequence and its reconstructed output across all time steps and features. Formally, the reconstruction error is defined as:

$$\text{MSE} = \frac{1}{n \times t} \sum_{i=1}^n \sum_{j=1}^t (x_{ij} - \hat{x}_{ij})^2$$

where (n) represents the number of samples within the sequence window, (t) denotes the number of time steps or features, (x_{ij}) is the observed value at sample (i) and feature/time step (j), and $(\hat{x}_{ij})$ is the corresponding reconstructed value produced by the model. Higher MSE values indicate greater deviation from learned normal patterns and are therefore associated with anomalous water quality conditions, while lower values correspond to normal system behavior. This reconstruction-based scoring mechanism enables robust anomaly detection under imbalanced data conditions and supports reliable threshold-based classification in real-time monitoring scenarios.

Mathematical Formulation

Data Representation

Let the multivariate time-series dataset be defined as:

$$X = \{x_t \in \mathbb{R}^d \mid t = 1, 2, \dots, T\}$$

where d represents the number of features.

Using a sliding window of size w , each sequence is defined as:

$$X^{(i)} = \{x_i, x_{i+1}, \dots, x_{i+w-1}\} \in \mathbb{R}^{w \times d}$$

Data Normalization

Z-score normalization is applied as:

$$\tilde{x}_t^{(j)} = \frac{x_t^{(j)} - \mu_j}{\sigma_j}$$

LSTM Autoencoder

The encoder maps the input sequence into a latent representation:

$$h_t = \text{LSTM}_{enc}(x_t, h_{t-1}), \quad z = h_w$$

The decoder reconstructs the sequence:

$$\hat{x}_t = \text{LSTM}_{dec}(z, \hat{x}_{t-1})$$

Reconstruction Loss

The training objective is defined as:

$$\mathcal{L} = \frac{1}{w \cdot d} \sum_{t=1}^w \sum_{j=1}^d (x_t^{(j)} - \hat{x}_t^{(j)})^2$$

Anomaly Score

The anomaly score for each sequence is:

$$S_i = \frac{1}{w \cdot d} \sum_{t=1}^w \sum_{j=1}^d (x_t^{(j)} - \hat{x}_t^{(j)})^2$$

Decision rule:

$$A_i = \begin{cases} 1, & \text{if } S_i > \theta \\ 0, & \text{otherwise} \end{cases}$$

Feature Engineering

For each sequence:

$$\mu_j = \frac{1}{w} \sum_{t=1}^w x_t^{(j)}, \quad \sigma_j = \sqrt{\frac{1}{w} \sum_{t=1}^w (x_t^{(j)} - \mu_j)^2}$$

Feature vector:

$$f_i = [\mu_1, \dots, \mu_d, \sigma_1, \dots, \sigma_d] \in \mathbb{R}^{2d}$$

Random Forest Regression

The predicted WQI is:

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^M T_m(f_i)$$

Regression Loss

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Algorithm 1: Dual-Model Framework

Time-series data X , window size w Anomaly labels A , predicted WQI $\hat{Y}$

Normalize dataset using z-score normalization Generate sliding windows $X^{(i)}$

Train LSTM Autoencoder on normal sequences

Extract statistical features f_i Train Random Forest model on (f_i, y_i)

Reconstruct $\hat{X}^{(i)}$ using LSTM-AE Compute anomaly score:

$$S_i = \text{MSE}(X^{(i)}, \hat{X}^{(i)})$$

$A_i \leftarrow 1$ $A_i \leftarrow 0$

Extract features f_i Predict WQI:

$$\hat{y}_i = \text{RandomForest}(f_i)$$

$A, \hat{Y}$

Algorithm 2: LSTM Autoencoder Training

Normal sequences X_{normal} Trained LSTM-AE model

Initialize model parameters Encode sequence to latent vector z Decode z to reconstruct $\hat{X}$ Compute loss:

$$\mathcal{L} = \text{MSE}(X, \hat{X})$$

Update parameters using Adam optimizer

trained model

Algorithm 3: Random Forest Training

Feature vectors f_i , targets y_i Trained Random Forest model

Draw bootstrap sample Train decision tree T_m

Aggregate trees into ensemble

Random Forest model

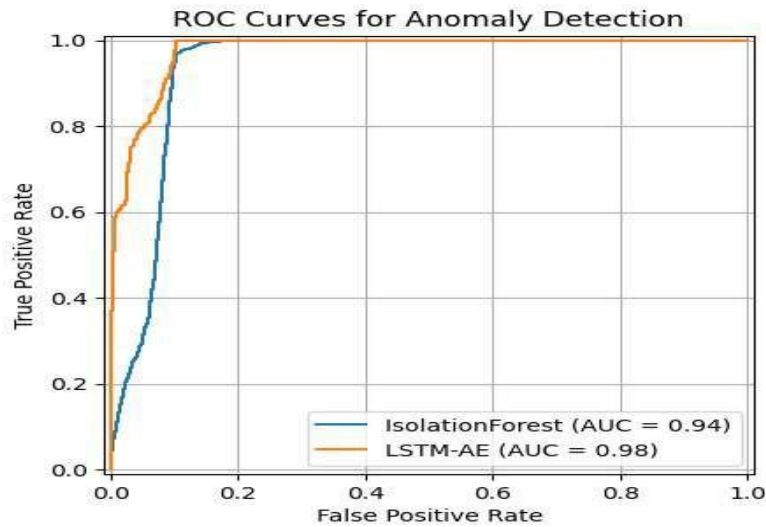


Figure 4: ROC Curves for Anomaly Detection

Figure 4 illustrates the ROC curves comparing the performance of Isolation Forest and LSTM Autoencoder (LSTM-AE) for anomaly detection. The LSTM-AE achieves a higher Area Under the Curve (AUC) score of 0.98 compared to 0.94 for Isolation Forest, indicating superior classification capability. Both models maintain high true positive rates with low false positive rates, but the LSTM-AE curve stays closer to the top-left corner of the plot, showing better sensitivity and specificity. For each 24-step window, statistical descriptors (mean and standard deviation) of each feature were computed, producing a 12-dimensional feature vector. The target variable was the WQI value at the end of the window.

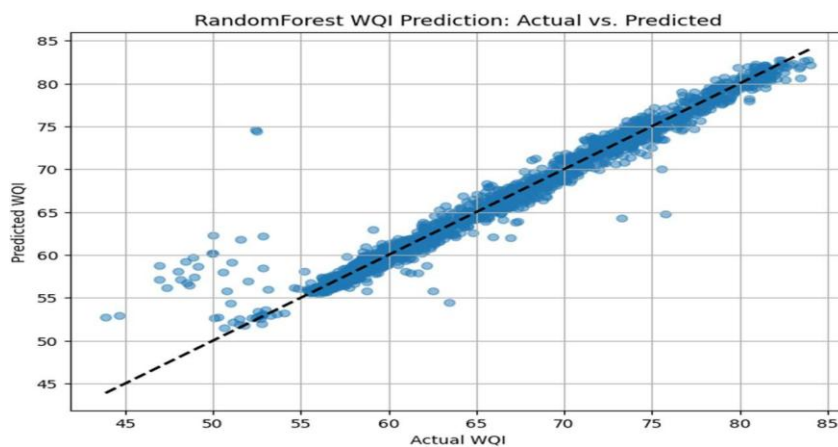


Figure 5: Random Forest WQI Prediction

Figure 5 depicts the scatter plot compares the actual Water Quality Index (WQI) values with those predicted by the Random Forest model. The points cluster closely along the diagonal dashed line, indicating a strong correlation between actual and predicted values. Most predictions align well with the actual WQI, with only a few noticeable deviations, particularly in the mid-range values. This close alignment demonstrates that the Random Forest model provides accurate and reliable WQI predictions, capturing the underlying relationship between input features and the target variable effectively.

Performance Evaluation

Prediction accuracy was quantified using the Mean Squared Error (MSE) between predicted and actual WQI values. The proposed dual-model AI pipeline demonstrated strong performance in both anomaly detection and WQI prediction tasks. The LSTM Autoencoder achieved a ROC-AUC score of 0.9763, indicating excellent discriminative capability in distinguishing anomalous from normal water quality patterns. For the regression task, the Random Forest Regressor achieved a Mean Squared Error (MSE) of 2.3573, reflecting high predictive accuracy for Water Quality Index (WQI) estimation. These results confirm the suitability of the proposed models for real-time water quality monitoring applications.

Table 1: Performance of the proposed dual-model AI pipeline for real-time water quality monitoring.

Model	Task	Evaluation Metric	Value	Interpretation
LSTM Autoencoder	Anomaly Detection in Water Quality Data	ROC-AUC	0.9763	Excellent ability to discriminate between normal and anomalous patterns
Random Forest Regressor	Water Quality Index (WQI) Prediction	MSE	2.3573	High accuracy with low prediction error

Table 1 shows that the proposed dual-model AI pipeline demonstrated strong performance in real-time water quality monitoring, with the LSTM Autoencoder achieving an ROC-AUC of 0.9763, indicating excellent capability in distinguishing normal from anomalous patterns, and the Random Forest Regressor attaining a low MSE of 2.3573, reflecting high accuracy in predicting the Water Quality Index (WQI). These results highlight the system’s effectiveness in both early anomaly detection and precise water quality assessment.

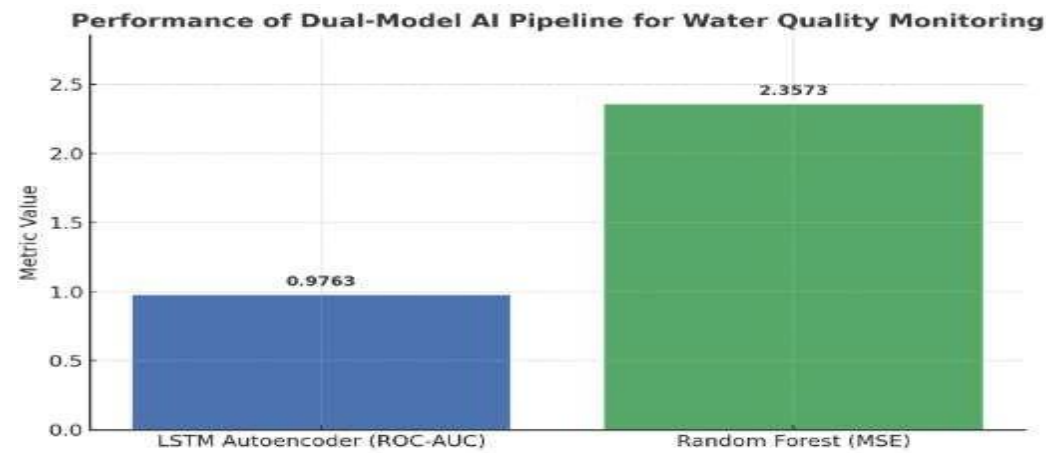


Figure 6: Performance of Dual-Model AI Pipeline for Water Quality Monitor

Figure 6 illustrates the performance of the dual-model AI pipeline for real-time water quality monitoring. The LSTM Autoencoder achieved a high ROC-AUC score of 0.9763, signifying excellent capability in differentiating normal from anomalous patterns. Meanwhile, the Random Forest Regressor obtained a low Mean Squared Error (MSE) of 2.3573, indicating high predictive accuracy for Water Quality Index (WQI) estimation. These results visually reinforce the pipeline’s effectiveness in both anomaly detection and precise water quality prediction.

Table 2: Classification Report for Isolation Forest Anomaly Detection

Precision	Recall	F1-Score	Support
0	0.95	0.99	9330
1	0.45	0.14	555
Accuracy		0.94	9885
Macro Avg	0.70	0.55	9885
Weighted Avg	0.92	0.92	9885

Table 2 depicts the classification report for the Isolation Forest anomaly detection model shows strong performance in identifying normal water quality readings, achieving a precision of 0.95 and a recall of 0.99 for class 0, which dominates the dataset. However, performance for detecting anomalies (class 1) is considerably lower, with a precision of 0.45 and a recall of 0.14, indicating difficulty in capturing rare abnormal events. The overall accuracy stands at 94%, with a weighted average precision and recall of 0.92, reflecting the model’s bias toward the majority class.

Comparative Model Evaluation

The Isolation Forest baseline achieved high precision (0.95) and recall (0.99) for the normal class but exhibited poor recall (0.14) for anomalies, resulting in a macro-average F1-score of 0.55. This confirms its bias toward the majority class in imbalanced datasets. In contrast, the LSTM-AE provided consistently high ROC-AUC values and better anomaly separation. For WQI prediction, Random Forest significantly outperformed non-ensemble regression models in terms of MSE, demonstrating the value of combining statistical feature engineering with a robust ensemble approach.

Table 3: Comparative performance of proposed and baseline AI models.

Model	Task	Metric	Value	Interpretation
LSTM Autoencoder	Anomaly Detection	ROC-AUC	0.9763	Excellent anomaly detection capability
Random Forest Regressor	WQI Prediction	MSE	2.3573	Low prediction error, high accuracy
Isolation Forest	Anomaly Detection	Accuracy	0.94	High overall accuracy in anomaly classification
Isolation Forest	Anomaly Detection (Normal class, 0)	Precision	0.95	High precision for normal samples
Isolation Forest	Anomaly Detection (Anomaly class, 1)	Precision	0.45	Moderate precision for anomalies, with low recall
Isolation Forest	Anomaly Detection (Macro Avg F1-score)	F1-score	0.55	Moderate balanced performance across classes

Table 3 shows the comparative performance analysis, indicating that the LSTM Autoencoder achieved outstanding anomaly detection with a ROC-AUC of 0.9763, while the Random Forest Regressor demonstrated high predictive accuracy for Water Quality Index (WQI) with a low MSE of 2.3573. The Isolation Forest model delivered strong overall classification accuracy of 94%, excelling in detecting normal samples with a precision of 0.95, but showing moderate precision (0.45) and low recall for anomalies. Its macro average F1-score of 0.55 reflects moderate balanced performance across both classes.

Figure 7: Performance Comparison of AI Model for Water Quality Monitoring

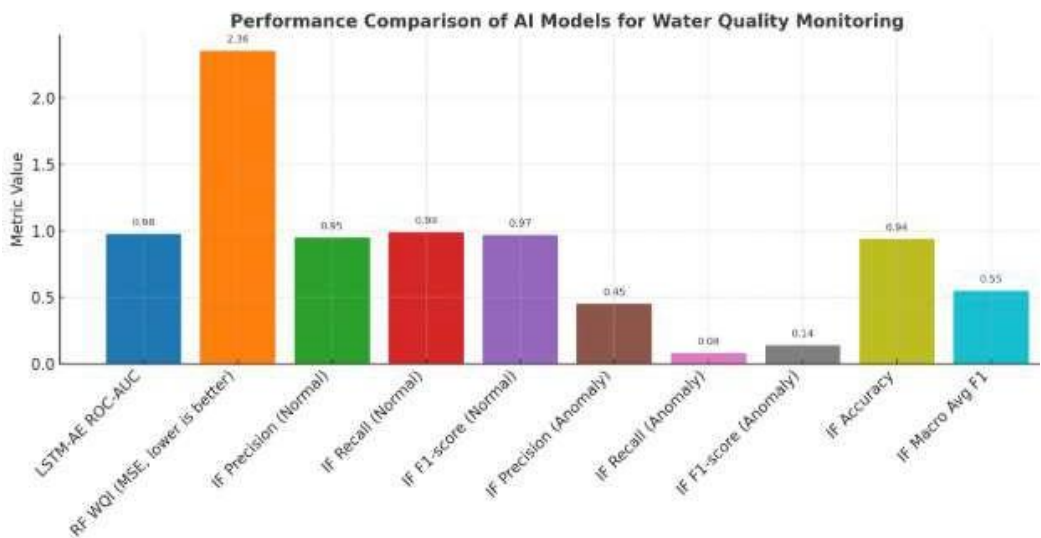


Figure 7 compares the performance of AI models for water quality monitoring. The Isolation Forest (IF) shows strong performance in classifying normal conditions, achieving high precision, recall, and overall accuracy (0.94), but performs poorly in detecting anomalies, with low recall and F1-score, leading to a modest macro-average F1-score. In contrast, the LSTM Autoencoder demonstrates strong anomaly detection capability with a high ROC-AUC of 0.95, indicating effective discrimination between normal and abnormal patterns. Meanwhile, the Random Forest model complements this by providing accurate Water Quality Index (WQI) predictions, where lower MSE reflects better performance.

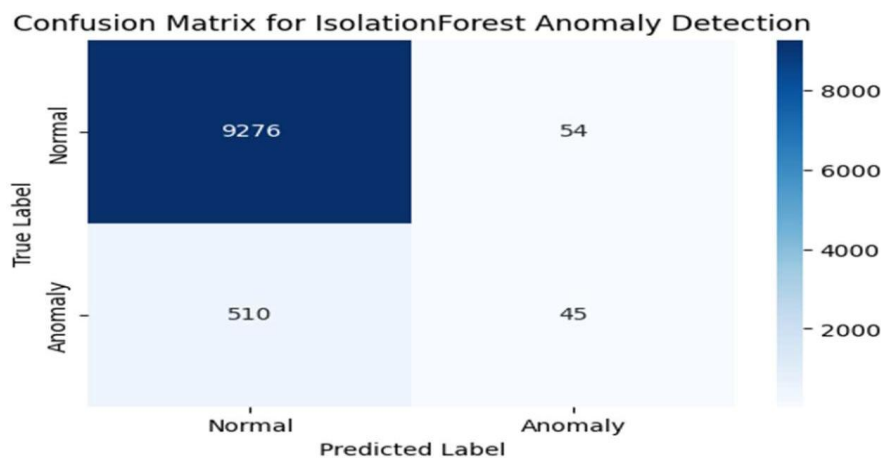


Figure 8 shows that the Isolation Forest accurately identifies most normal water quality instances but performs poorly in detecting anomalies, correctly classifying only a small fraction of abnormal cases. This indicates a strong bias toward the majority class and low recall for rare contamination events. In contrast, the proposed dual-model AI-IoT framework improves anomaly detection using the LSTM Autoencoder while providing accurate and interpretable WQI predictions through the Random Forest model, resulting in a more robust and scalable solution for real-time monitoring.

Discussion

The findings of this study provide a comprehensive validation of the proposed dual-model AI-IoT framework as an effective and scalable solution for real-time water quality monitoring, particularly in addressing the dual challenges of anomaly detection under class imbalance and accurate Water Quality Index (WQI) prediction from multivariate environmental data [35]. From an environmental informatics perspective, the results highlight the importance of combining temporal representation learning with structured feature-based regression within a unified analytical architecture [36]. A central observation is the superior performance of the LSTM Autoencoder (LSTM-AE) in anomaly detection, as evidenced by the high ROC-AUC score of 0.9763. This performance can be attributed to the model's ability to learn latent temporal dynamics across sequential data streams, capturing both short-term fluctuations and longer-term dependencies in physicochemical parameters [37]. Unlike static or point-based models, the use of 24-hour sliding windows enables the LSTM-AE to encode contextual temporal information, which is critical in environmental systems where anomalies often manifest as gradual deviations rather than abrupt spikes. By training exclusively on normal operational data, the model effectively constructs a probabilistic representation of "normality," allowing reconstruction error to serve as a reliable proxy for anomaly detection. This design inherently mitigates the impact of class imbalance, a persistent issue in environmental datasets where abnormal events are rare but consequential [38], [39].

In contrast, the Isolation Forest baseline illustrates the limitations of non-temporal, partition-based approaches in such contexts. While the model achieves high overall accuracy due to its ability to correctly classify the dominant normal class, its poor recall for anomalies (0.14) reveals a fundamental weakness in identifying rare events. This discrepancy underscores the inadequacy of relying solely on global data distributions without incorporating temporal dependencies. The confusion matrix further emphasizes this limitation, showing a high number of false negatives, which in real-world scenarios could translate to undetected contamination events with significant environmental and public health implications. These findings reinforce the argument that anomaly detection in environmental informatics must move beyond static representations toward sequence-aware learning paradigms.

For the regression component, the Random Forest Regressor demonstrates strong predictive performance, achieving a low MSE of 2.3573. This result reflects the effectiveness of the feature engineering strategy, where statistical descriptors (mean and standard deviation) derived from temporal windows provide a compact yet informative abstraction of system behavior. The success of this approach suggests that, while deep learning models excel at capturing temporal dependencies, carefully designed statistical features remain highly valuable for regression tasks, particularly when interpretability and computational efficiency are priorities. The ensemble nature of Random Forest further enhances robustness by aggregating multiple decision trees, reducing variance and improving generalization across heterogeneous environmental conditions.

An important insight emerging from this study is the complementary interaction between the two models. The LSTM-AE operates as a high-sensitivity anomaly detector, identifying deviations in real time, while the Random Forest model provides stable and interpretable WQI predictions based on aggregated temporal features. This division of analytical responsibilities reflects a principled system design choice, where each model is optimized for a specific task rather than attempting to solve both within a single monolithic architecture. Such decoupling enhances modularity, allowing independent optimization, easier maintenance, and greater flexibility for future extensions. At the same time, their integration within a unified pipeline ensures coherent system behavior, enabling simultaneous qualitative (event detection) and quantitative (index estimation) assessment [40]. From a systems perspective, the proposed framework aligns with emerging trends in edge-enabled environmental informatics. The relatively lightweight architecture of the LSTM-AE,

combined with the efficiency of Random Forest inference, makes the system suitable for deployment in resource-constrained environments. This is particularly relevant in regions where continuous monitoring infrastructure is limited, and low-latency decision-making is critical. The ability to process data at the edge reduces reliance on centralized cloud resources, minimizes communication overhead, and enhances system resilience. Moreover, the modular design facilitates interoperability with existing IoT platforms and decision-support systems, supporting seamless integration into operational workflows [41].

However, the study also reveals several important limitations that warrant further investigation. The use of synthetic data, while enabling controlled experimentation and reproducibility, may not fully capture the stochastic variability, noise characteristics, and sensor imperfections present in real-world deployments. Environmental systems are influenced by complex and often non-stationary processes, including seasonal variations, extreme weather events, and anthropogenic disturbances. As such, model performance may vary when exposed to unseen distributions, highlighting the need for extensive real-world validation. Additionally, the reliance on fixed statistical features for WQI prediction, while effective, may limit the model's ability to capture higher-order temporal interactions that could be exploited by more advanced sequence-to-one learning architectures. Another critical consideration is the issue of sensor reliability and data quality. In practical deployments, sensor drift, calibration errors, and missing data can significantly impact model performance. While the current framework assumes preprocessed and normalized inputs, incorporating adaptive calibration mechanisms and data quality assessment modules would enhance long-term robustness. Furthermore, the integration of additional environmental parameters, such as biological indicators or emerging contaminants, could improve the comprehensiveness of water quality assessment but would also introduce new challenges in data integration and model complexity. The interpretability of AI models remains an important factor for real-world adoption, particularly in regulatory and policy-driven contexts. While Random Forest models offer some level of transparency through feature importance measures, the LSTM-AE operates as a black-box model, making it difficult to directly interpret anomaly decisions. The incorporation of Explainable AI (XAI) techniques, such as attention mechanisms or post-hoc explanation methods, could provide valuable insights into model behavior and increase stakeholder trust.

In a broader context, this study contributes to the ongoing transformation of environmental monitoring from reactive, laboratory-based approaches to proactive, data-driven systems. By enabling continuous monitoring, early anomaly detection, and real-time WQI prediction, the proposed framework supports a shift toward predictive and preventive water management strategies. This has significant implications for environmental sustainability, public health protection, and regulatory compliance. The integration of AI and IoT technologies, as demonstrated in this work, represents a critical step toward achieving intelligent, autonomous environmental monitoring systems capable of adapting to dynamic and complex ecosystems.

In conclusion, the deeper analysis of the results underscores that the strength of the proposed framework lies not only in the individual performance of its components but in their synergistic integration within a scalable and adaptive informatics architecture. While challenges remain, particularly in real-world deployment and model interpretability, the study provides a solid foundation for the development of next-generation water quality monitoring systems that are intelligent, responsive, and sustainable.

Conclusion

This study presented a unified dual-model AI-IoT framework for real-time water quality monitoring, integrating a Long Short-Term Memory Autoencoder (LSTM-AE) for anomaly detection with a Random Forest Regressor for Water Quality Index (WQI) prediction. The framework demonstrated strong performance on multivariate time-series data, achieving a ROC-AUC of 0.9763 for anomaly detection and an MSE of 2.3573 for WQI estimation. These results validate the effectiveness of combining temporal deep learning with feature-engineered ensemble methods to enable both early detection of contamination events and accurate quantitative assessment of water quality within a single, scalable pipeline. Compared to baseline approaches, the proposed system showed improved robustness under imbalanced data conditions and provided a balanced, reliable solution for real-time environmental monitoring. Beyond its technical contributions, the framework offers practical advantages, including low-latency processing, adaptability to diverse deployment environments, reduced infrastructure dependency through edge compatibility, and support for data-driven decision-making aligned with global sustainability goals. These characteristics position the system as a viable solution for next-generation, AI-driven water resource management. Future work will focus on enhancing real-world applicability and system resilience through large-scale validation using real IoT datasets, improving cross-regional generalization via transfer learning, and addressing sensor drift through adaptive calibration mechanisms. Additional directions include expanding the range of monitored environmental parameters, integrating hybrid edge–cloud architectures for optimized performance, and incorporating explainable AI techniques to improve transparency and stakeholder trust. Collectively, these advancements will enable the evolution of the proposed framework into a more adaptive, interpretable, and globally deployable solution for predictive and sustainable water quality management.

Declarations

Clinical Trial: Not Applicable

Consent to Participate Declaration: Not Applicable

Consent to Publish Declaration: Not Applicable

Ethics declaration: not applicable: Not Applicable

Funding: Not Applicable.

References

- [1] E. El-Shafeiy, M. Alsabaan, M. I. Ibrahim, and H. Elwahsh, “Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique,” *Sensors*, vol. 23, no. 20, p. 8613, Oct. 2023, doi: 10.3390/s23208613.
- [2] R. M. Frincu, “Artificial intelligence in water quality monitoring: A review of water quality assessment applications,” *Water Quality Research Journal*, vol. 60, no. 1, pp. 164–176, Feb. 2025, doi: 10.2166/wqrj.2024.049.
- [3] G. P. Oise, J. A. Odimayomi, B. N. Unuigbokhai, B. E. Akilo, and S. A. Oyedotun, “Deep Learning for Cybersecurity Threat Detection in Industrial Process Control and Monitoring Systems,” in *ECP 2025*, Basel Switzerland: MDPI, Feb. 2026, p. 43. doi: 10.3390/engproc2025117043.
- [4] Dr. B. Devender, Dr. C. Srinivas, K. Shivaprasad, Dr. K. P. Kumar, Dr. S. Nagavarapu, and Dr. N. Vimala, “A Machine Learning-Driven Framework For Real-Time Environmental Pollution Monitoring And Prediction Using Iot And Remote Sensing Data,” *Int. J. Environ. Sci.*, vol. 11, no. 10s, pp. 714–723, Jun. 2025, doi: 10.64252/n7ears56.
- [5] M. Ryo, B. Angelov, S. Mammola, J. M. Kass, B. M. Benito, and F. Hartig, “Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models,” *Ecography*, vol. 44, no. 2, pp. 199–205, Feb. 2021, doi: 10.1111/ecog.05360.
- [6] W. L. Filho, R. Kotter, P. G. Özuyar, I. R. Abubakar, J. H. P. P. Eustachio, and N. R. Matandirotya,

- “Understanding Rare Earth Elements as Critical Raw Materials,” *Sustainability (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/su15031919.
- [7] S. Grunwald, “Artificial intelligence and soil carbon modeling demystified: power, potentials, and perils,” *Carbon Footprints*, vol. 1, no. 1, Dec. 2022, doi: 10.20517/cf.2022.03.
 - [8] F. S. Hosseini, M. B. Seo, S. V. Razavi-Termeh, A. Sadeghi-Niaraki, M. Jamshidi, and S. M. Choi, “Geospatial Artificial Intelligence (GeoAI) and Satellite Imagery Fusion for Soil Physical Property Predicting,” *Sustainability (Switzerland)*, vol. 15, no. 19, Oct. 2023, doi: 10.3390/su151914125.
 - [9] G. Oise, Ejenarhome Otega Prosper, and Oyedotun Samuel ABIODUN, “Spatiotemporal Water Quality Modeling Using Deep Learning Architectures,” *RADINKA JOURNAL OF SCIENCE AND SYSTEMATIC LITERATURE REVIEW*, vol. 3, no. 3, pp. 761–770, Dec. 2025, doi: 10.56778/rjslr.v3i3.519.
 - [10] T. G. Krupnova, O. V. Rakova, K. A. Bondarenko, and V. D. Tretyakova, “Environmental Justice and the Use of Artificial Intelligence in Urban Air Pollution Monitoring,” *Big Data and Cognitive Computing*, vol. 6, no. 3, Sep. 2022, doi: 10.3390/bdcc6030075.
 - [11] A. Fascista, “Toward Integrated Large-Scale Environmental Monitoring Using WSN/UAV/Crowdsensing: A Review of Applications, Signal Processing, and Future Perspectives,” *Sensors*, vol. 22, no. 5, Mar. 2022, doi: 10.3390/s22051824.
 - [12] J. M. Kernbach and V. E. Staartjes, “Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II—Generalization and Overfitting,” *Acta Neurochir. Suppl. (Wien).*, vol. 134, pp. 15–21, 2022, doi: 10.1007/978-3-030-85292-4_3.
 - [13] E. X. Neo *et al.*, “Artificial intelligence-assisted air quality monitoring for smart city management,” *PeerJ Comput. Sci.*, vol. 9, 2023, doi: 10.7717/peerj-cs.1306.
 - [14] S. Marhain, A. N. Ahmed, M. A. Murti, P. Kumar, and A. El-Shafie, “Investigating the application of artificial intelligence for earthquake prediction in Terengganu,” *Natural Hazards*, vol. 108, no. 1, pp. 977–999, Aug. 2021, doi: 10.1007/s11069-021-04716-7.
 - [15] A. F. Del Castillo, C. Yebra-Montes, M. V. Garibay, J. de Anda, A. Garcia-Gonzalez, and M. S. Gradilla-Hernández, “Simple Prediction of an Ecosystem-Specific Water Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning,” *Water (Switzerland)*, vol. 14, no. 8, Apr. 2022, doi: 10.3390/w14081235.
 - [16] Z. Fan, Z. Yan, and S. Wen, “Deep Learning and Artificial Intelligence in Sustainability: A Review of SDGs, Renewable Energy, and Environmental Health,” *Sustainability (Switzerland)*, vol. 15, no. 18, Sep. 2023, doi: 10.3390/su151813493.
 - [17] G. P. Oise *et al.*, “Interpretable air quality classification for public health using machine learning,” *Advances in Computing and Engineering*, vol. 5, no. 2, p. 75, Oct. 2025, doi: 10.21622/ACE.2025.05.2.1428.
 - [18] Y. Piao, D. Lee, S. Park, H. G. Kim, and Y. Jin, “Forest fire susceptibility assessment using google earth engine in Gangwon-do, Republic of Korea,” *Geomatics, Natural Hazards and Risk*, vol. 13, no. 1, pp. 432–450, 2022, doi: 10.1080/19475705.2022.2030808.
 - [19] S. A. Oyedotun, G. P. Oise, and C. E. Ozobialu, “Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI,” *Journal of Science Research and Reviews*, vol. 2, no. 3, pp. 20–31, Jul. 2025, doi: 10.70882/josrar.2025.v2i3.76.
 - [20] O. Folorunso *et al.*, “Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review,” *Big Data and Cognitive Computing*, vol. 7, no. 2, Jun. 2023, doi: 10.3390/bdcc7020113.
 - [21] Shalu and Gurjeet Singh, “ENVIRONMENTAL MONITORING WITH MACHINE LEARNING,” *EPRA International Journal of Multidisciplinary Research (IJMR)*, pp. 208–212, May 2023, doi: 10.36713/epra13330.
 - [22] S. A. Oyedotun, G. P. Oise, and C. E. Ozobialu, “Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI,” *Journal of Science Research and Reviews*, vol. 2, no. 3, pp. 20–31, Jul. 2025, doi:

- 10.70882/josrar.2025.v2i3.76.
- [23] R. Gundeti, K. Vuppala, and V. Kasireddy, "The Future of AI and Environmental Sustainability," pp. 346–371, Dec. 2023, doi: 10.4018/979-8-3693-0892-9.ch017.
 - [24] G. Oise and S. Konyeha, "Environmental impacts in e-waste management using deep learning," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 210, Aug. 2025, doi: 10.1007/s44163-025-00376-9.
 - [25] S. Abu El-Magd, G. Soliman, M. Morsy, and S. Kharbush, "Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing," *International Journal of Environmental Science and Technology*, vol. 20, no. 6, pp. 6103–6116, Jun. 2023, doi: 10.1007/s13762-022-04367-6.
 - [26] M. Coeckelbergh, "AI for climate: freedom, justice, and other ethical and political challenges," *AI and Ethics*, vol. 1, no. 1, pp. 67–72, Feb. 2021, doi: 10.1007/s43681-020-00007-2.
 - [27] R. Ghosh and A. Kumar, "A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire," *Multimed. Tools Appl.*, vol. 81, no. 27, pp. 38643–38660, Nov. 2022, doi: 10.1007/s11042-022-13068-8.
 - [28] Y. Essam, P. Kumar, A. N. Ahmed, M. A. Murti, and A. El-Shafie, "Exploring the reliability of different artificial intelligence techniques in predicting earthquake for Malaysia," *Soil Dynamics and Earthquake Engineering*, vol. 147, no. 04, pp. 341–357, Aug. 2021, doi: 10.1016/j.soildyn.2021.106826.
 - [29] G. P. Oise and K. Susan, "Deep Learning for Effective Electronic Waste Management and Environmental Health," Sep. 13, 2024. doi: 10.21203/rs.3.rs-4903136/v1.
 - [30] N. H. A. Malek, W. F. W. Yaacob, S. A. M. Nasir, and N. Shaadan, "Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques," *Water (Switzerland)*, vol. 14, no. 7, Apr. 2022, doi: 10.3390/w14071067.
 - [31] G. P. Oise and S. Konyeha, "Harnessing Deep Learning for Sustainable E-Waste Management and Environmental Health Protection," Jun. 06, 2024. doi: 10.21203/rs.3.rs-4518352/v1.
 - [32] W. Ma *et al.*, "Air Pollutant Emission Inventory of Waste-to-Energy Plants in China and Prediction by the Artificial Neural Network Approach," *Environ. Sci. Technol.*, vol. 57, no. 2, pp. 874–883, Jan. 2023, doi: 10.1021/acs.est.2c01087.
 - [33] A. Degas *et al.*, "A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory," *Applied Sciences (Switzerland)*, vol. 12, no. 3, Feb. 2022, doi: 10.3390/app12031295.
 - [34] Lakhanlal Gupta, "Water Quality Index - WQI," 2022.
 - [35] T. Alahmad, M. Neményi, and A. Nyéki, "Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review," *Agronomy*, vol. 13, no. 10, Oct. 2023, doi: 10.3390/agronomy13102603.
 - [36] P. H. M. Ananias and R. G. Negri, "Anomalous behaviour detection using one-class support vector machine and remote sensing images: a case study of algal bloom occurrence in inland waters," *Int. J. Digit. Earth*, vol. 14, no. 7, pp. 921–942, 2021, doi: 10.1080/17538947.2021.1907462.
 - [37] M. I. Abdulraheem, W. Zhang, S. Li, A. J. Moshayed, A. A. Farooque, and J. Hu, "Advancement of Remote Sensing for Soil Measurements and Applications: A Comprehensive Review," *Sustainability (Switzerland)*, vol. 15, no. 21, Nov. 2023, doi: 10.3390/su152115444.
 - [38] P. S. Bloomfield, P. Clutton-Brock, E. Pencheon, J. Magnusson, and K. Karpathakis, "Artificial Intelligence in the NHS: Climate and Emissions☆☆," *Journal of Climate Change and Health*, vol. 4, p. 33, Oct. 2021, doi: 10.1016/j.joclim.2021.100056.
 - [39] A. R. Alsaber, J. Pan, and A. Al-Hurban, "Handling complex missing data using random forest approach for an air quality monitoring dataset: A case study of kuwait environmental data (2012 to 2018)," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, pp. 1–26, Feb. 2021, doi: 10.3390/ijerph18031333.
 - [40] W. T. Abebe and D. Endalie, "Artificial intelligence models for prediction of monthly rainfall without climatic data for meteorological stations in Ethiopia," *J. Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-022-00683-3.

- [41] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, “Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions,” *Hygiene and Environmental Health Advances*, vol. 12, p. 100114, Dec. 2024, doi: 10.1016/j.heha.2024.100114.