engineering proceedings

MDPI

Proceeding Paper

Real-Time Pollutant Forecasting Using Edge–AI Fusion in Wastewater Treatment Facilities †

Siva Shankar Ramasamy [1,*], Vijayalakshmi Subramanian [2], Leelambika Varadarajan [3] and Alwin Joseph [2]

[1] International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand
[2] School of Sciences, CHRIST University, Bengaluru 560029, Karnataka, India; s.vijayalakshmi@christuniversity.in (V.S.)
[3] Presidency School of Computer Science and Engineering, Presidency University, Bengaluru 560119, Karnataka, India; vleelambika@gmail.com
* Correspondence: sivashankar.r@cmu.ac.th
† Presented at the 4th International Electronic Conference on Processes, 20–22 October 2025; Available online: https://sciforum.net/event/ECP2025.

Abstract

Wastewater treatment is one of the major challenges in the reuse of water as a natural resource. Cleaning of water depends on analyzing and treating the water for the pollutants that have a significant impact on the quality of the water. Detecting and analyzing the surges of these pollutants well before the recycling process is needed to make intelligent decisions for water cleaning. The dynamic changes in pollutants need constant monitoring and effective planning with appropriate treatment strategies. We propose an edge-computing-based smart framework that captures data from sensors, including ultraviolet, electrochemical, and microfluidic, along with other significant sensor streams. The edge devices send the data from the cluster of sensors to a centralized server that segments anomalies, analyzes the data and suggests the treatment plan that is required, which includes aeration, dosing adjustments, and other treatment plans. A logic layer is designed at the server level to process the real-time data from the sensor clusters and identify the discharge of nutrients, metals, and emerging contaminants in the water that affect the quality. The platform can make decisions on water treatments using its monitoring, prediction, diagnosis, and mitigation measures in a feedback loop. A rule-based Large Language Model (LLM) agent is attached to the server to evaluate data and trigger required actions. A streamlined data pipeline is used to harmonize sensor intervals, flag calibration drift, and store curated features in a local time-series database to run ad hoc analyses even during critical conditions. A user dashboard has also been designed as part of the system to show the recommendations and actions taken. The proposed system acts as an AI-enabled system that makes smart decisions on water treatment, providing an effective cleaning process to improve sustainability.

Keywords: wastewater treatment; Edge–AI; sensor fusion; real-time monitoring; pollutant forecasting; anomaly detection; Large Language Models (LLMs); smart water management; sustainability; decision support systems

check for updates

Academic Editor: Omprakash Sarkar

Published: 22 January 2026

Citation: Ramasamy, S.S.; Subramanian, V.; Varadarajan, L.; Joseph, A. Real-Time Pollutant Forecasting Using Edge–AI Fusion in Wastewater Treatment Facilities. *Eng. Proc.* **2025**, *117*, 31. https://doi.org/10.3390/engproc2025117031

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Water scarcity and water pollution are two of the most serious and globally pervasive environmental challenges [1,2]. Drinking safe water is a basic requirement for living beings. The huge change in the dynamics of urbanization, industrialization, and population growth

has increased the need for more water; the rise in water demand has exhausted the current freshwater resources. One solution to meet the rise in demand for water is wastewater treatment [3,4]. Conventional wastewater treatment systems use periodic sampling and centralize their control and data acquisition. Centralized processing and monitoring are effective for compliance monitoring, but these methods are reactive and lack the real-time adaptability where there is a requirement to respond to the dynamic fluctuations in pollutant concentrations [5]. The current system faces a delay in data acquisition, evaluation and treatment adjustment, resulting in a lack of proper purification, and the resources are not used to their full efficiency [6].

Advances in Internet of Things (IoT) and Artificial Intelligence (AI), when applied to the design and building of systems for wastewater management, can perform automatic water-quality monitoring with intelligent decision making [3,4,7]. The sensor-driven systems continuously measure various physical, chemical, and biological parameters. Some of the parameters that are measured constantly from water sources using sensors include pH, turbidity, dissolved oxygen, and metal ions and other metallic content [5,8]. There are certain AI-based approaches that are designed and depend on cloud-centric computation, where issues like network latency, bandwidth consumption, vulnerability to connectivity failures, etc., cause issues that lead to the compromise of critical treatment processes [7]. AI models sometimes also operate as "black boxes," where limited interpretability and trust cause issues in making sudden decisions [8,9].

Edge computing is an approach that enables system design by processing the required data at the data source [7,10]. Designing and deploying lightweight AI models at the network edge also reduces the latency, improves the analysis, accelerates the anomaly detection, and empowers the nodes to make preliminary decisions rather than waiting for control from the central server [7,8]. Combining the capability of data processing at different layers also enables predictive models capable of processing the continuous forecasting of pollutants and their trends with the aim of providing early intervention. Integrating a rule-based Large Language Model (LLM) agent further enhances interpretability and automation by evaluating model outputs and recommending treatment actions in natural language that operators can validate [9].

The issues with water scarcity and pollution have driven the need for innovative wastewater treatment solutions that move beyond the limitations of traditional, centralized, and reactive systems. Recent advancements in IoT and AI have enabled real-time monitoring of water quality through sensor-driven platforms, providing continuous measurements of key parameters such as pH, turbidity, dissolved oxygen, and metal ion concentrations. Existing AI-based approaches rely heavily on cloud-centric computation, which introduces latency, bandwidth constraints, and dependence on stable network connectivity factors that can hinder timely intervention in critical treatment scenarios. Edge computing has emerged as a promising paradigm by bringing lightweight analytics and decision-making closer to the data source, facilitating rapid anomaly detection and enabling localized, preliminary responses without waiting for centralized control. The integration of explainable, rule-based LLM agents further enhances the interpretability and automation of these systems, empowering operators to make informed decisions and optimize water treatment processes for sustainability and resilience. The development of an edge computing-based smart framework makes use of multimodal sensor fusion, hybrid AI models, and LLM-based decision support to address the dynamic needs of modern wastewater management.

There are researchers who see the scope of an Edge–AI fusion framework for the real-time forecasting of pollutants and making intelligent decisions for wastewater treatment facilities [1,4,8]. These systems integrate multimodal sensor data, lightweight edge-level analytics, a hybrid AI model for pollutant prediction, and an LLM-based decision layer for

intelligent treatment control. The primary objective of the current research is to design and implement a system that is capable of the following:

- Enabling real-time pollutant detection and forecasting using edge computing,
- An architecture of a system that reduces data-processing latency compared to cloud-only systems,
- Incorporate explainable, rule-driven automation via an LLM agent, and
- Enhance operational efficiency and sustainability in wastewater management.

Unlike existing works that independently explore IoT-based monitoring, cloud-centric AI models, or edge-enabled anomaly detection, the novelty of the proposed study lies in the system-level integration of three components: (i) edge-level anomaly detection using lightweight ensemble models, (ii) hybrid time-series forecasting that combines anomaly likelihood with long-term temporal prediction, and (iii) an explainable, rule-based LLM decision layer for treatment recommendation and operator interaction. The contribution is architectural and operational, demonstrating how existing AI models are turned into a real-time, low-latency, and interpretable wastewater treatment decision-support pipeline. The integration and feedback-driven control loop distinguishes the proposed framework from prior Edge–AI and IoT-based water quality systems.

2. Related Work

Wastewater management is highly dependent on continuous monitoring and predictive analysis, which helps to ensure that the water quality is maintained and the water can be reused to ensure sustainable development. Traditional wastewater treatment systems are based on a centralized supervisory control and data acquisition approach and are constrained by a lack of real-time data and decisions. Traditional approaches use offline sampling and laboratory analysis, which results in delayed responses, and pollutant fluctuations and inefficient resource usage are some of the common issues with these systems [1–3]. Recent systems that have been designed for ensuring quality and monitoring water treatment in real time have emphasized the need for intelligent, automated frameworks that combine sensors and computational intelligence, which enable dynamic control and adaptation [2,5,8].

The integration of the Internet of Things (IoT) with machine learning (ML) has enabled the development of sensors that are capable of capturing and processing real-time data. IoT-based models are used to capture the parameters using distributed sensors and to capture key water quality parameters such as pH, temperature, turbidity, dissolved oxygen, and conductivity [5,6,9–11]. Some recent researchers have demonstrated that IoT-driven systems connect sensors to cloud computing infrastructure to attain centralized resource monitoring and anomaly detection [7,9]. When connecting to the internet and to the cloud computing infrastructure, there are high chances that the entire system suffers from high latency, and this is heavily dependent on the network and commonly affected by data bottlenecks when processing large sensor streams. This limits the deployment of such systems in resource-constrained environments.

AI-driven predictive models could enhance wastewater treatment by giving directions based on the performance measures monitored by the sensors in real time. Researchers have also trialed the application of solutions designed with deep learning approaches such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks for pollutant concentration forecasting, treatment optimization, and effluent quality prediction [12–15]. Researchers also confirm that the design of hybrid models helps to combine neural networks and ensemble methods and aims to improve the forecasting accuracy for chemical oxygen demand, total suspended solids, and heavy metal detection [12,14,16]. Similarly, application of explainable AI (XAI) techniques help to provide interpretable

predictions, addressing the issues of black-box nature and limitations of conventional AI models [12,17,18]. Most of the AI-based implementations focus on remaining in the cloud-centric and lack local processing and decision capabilities, which gives them a limitation with real-time control.

Edge computing helps the processing be offloaded into the edge devices and then sends the command to the central node, and it has emerged as a viable solution to bridge this gap. By moving computation closer to data sources, edge nodes execute lightweight models and process data for real-time analysis, which helps to reduce network load and dependence. Researchers have demonstrated edge-enabled predictive control in water infrastructure, achieving faster anomaly detection and lower decision latency compared to purely cloud-based frameworks [7,16,19,20]. The combination of Edge–AI with rule-based decision systems offers enhanced interpretability and scalability for smart water treatment facilities [21,22]. However, there is still limited research integrating Edge–AI, multimodal sensor fusion, and LLM-based decision layers for real-time pollutant forecasting.

The current research aims to address the research gap by proposing a unified Edge–AI fusion framework capable of continuous pollutant monitoring, anomaly detection, and intelligent decision making. Unlike previous cloud-dependent systems, the proposed system uses the capacity of edge system deployments and uses data to streamline analysis and oversee the control dynamics. Hybrid AI models help in the decision-making process through the help of a rule-based LLM agent that can interpret the observations, generate real-time responsiveness and explainability on the observations, and plan for actions that are sustainability suitable for the wastewater treatment operations.

3. Methodology

The proposed system is designed as an Edge–AI fusion framework that can capture water-quality signals with the help of sensors that are able to capture multiple data points. The data are then processed and analyzed in real-time for anomalies in the device itself. Along with capturing and verifying the data, a forecasting agent forecasts the pollutant levels using hybrid AI models, which helps in generating automated treatment recommendations through a rule-based controlled LLM engine. The overall architecture of the proposed system and its workflow are shown in Figure 1.

3.1. System Architecture

The proposed system consists of five primary layers: *(i) Sensor Layer, (ii) Edge Processing Layer, (iii) Central Analytics Server, (iv) LLM Decision Layer, and (v) User Dashboard*.

3.1.1. Sensor Layer

Multiple layers of water-quality sensors, along with ultraviolet photometric probes, electrochemical electrodes, and microfluidic sensors, constitute the primary sensor layer. The sensors continuously measure the water parameters such as pH, turbidity, dissolved oxygen, nitrate concentration, conductivity, and metal traces. The sensors that measure data in real-time push the time-stamped readings to the nearest edge node through a lightweight MQTT/HTTP protocol.

3.1.2. Edge Processing Layer

Each edge device executes a compact model for initial screening and anomaly flagging. Noise filtering is performed using a moving-window median filter. Low-latency classification is achieved through a Random Forest-based anomaly classifier trained on short-term signal features, including the mean, variance, and slope change. The compressed feature vectors and abnormal events are notified to the server using a sync mechanism, which helps in reducing bandwidth and response time.

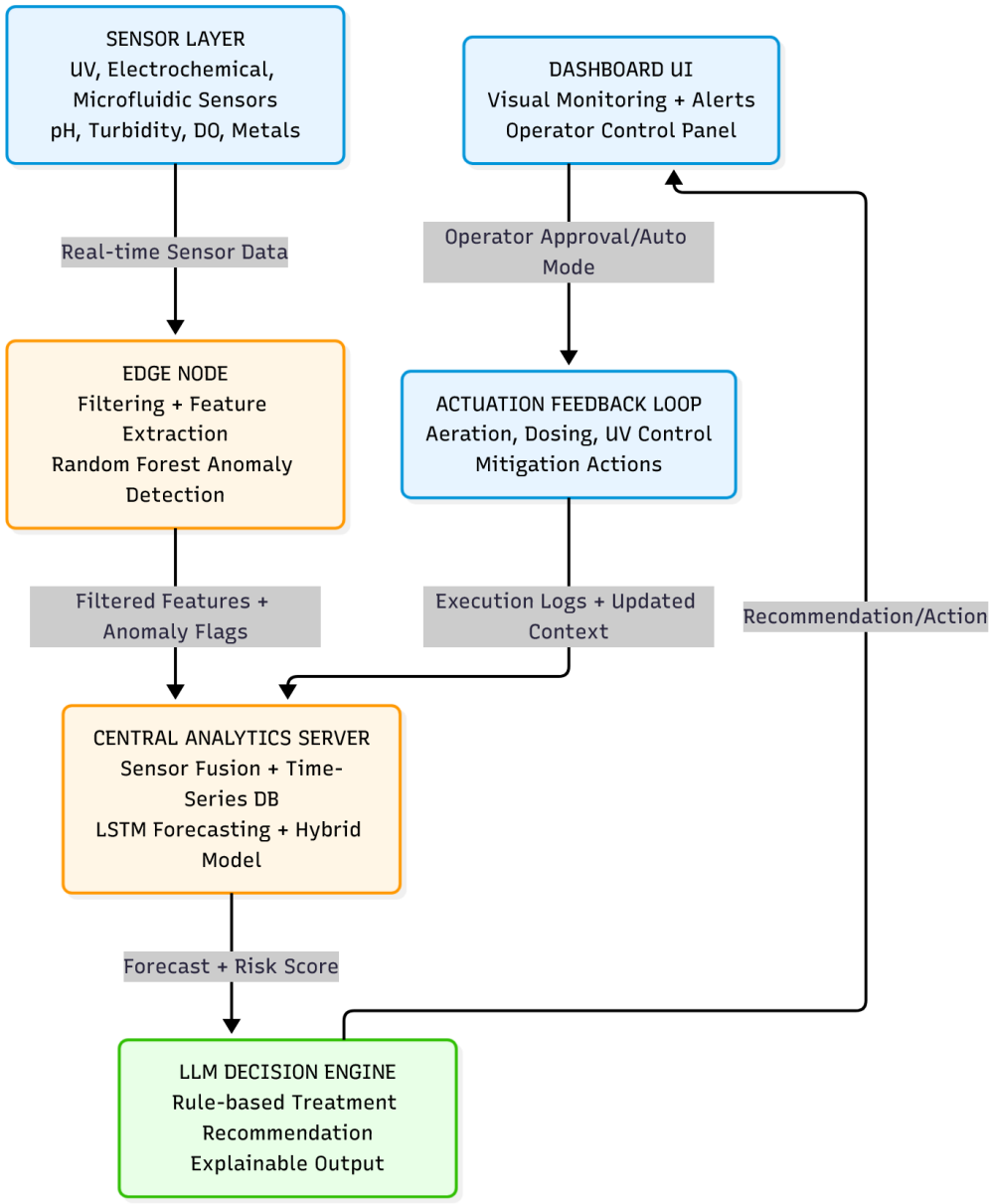


Figure 1. Workflow of the proposed Edge–AI fusion framework.

3.1.3. Central Analytics Layer

The server processes data from the edge devices, performs multimodal sensor data fusion, and runs the forecasting module. An LSTM-based predictor is designed to capture long-term pollutant trends and temporal dependencies. The fusion model layer is used to combine LSTM prediction scores with anomaly likelihood from the Random Forest classifier with the help of a weighted ensemble approach. When the pollutant concentration is predicted to cross safe thresholds, an alert packet is generated and passed to the decision layer and signals the required actions to mitigate the issue.

3.1.4. LLM Decision Layer

A rule-based Large Language Model (LLM) evaluates the model outputs from the central analytics layer against various thresholds and policies. The LLM engine, tuned to make decisions based on the interpretations from the data points, including pollutant trends, compares the results with the pre-configured threshold matrices. The engine is then able to generate actionable recommendations to the system, and this includes actions such as aeration increase, dosing adjustment, Ultraviolet (UV) activation, or valve shutoff.

Safety-critical decisions, when suggested by the system, are alerted to the operator to verify and confirm human-in-loop governance to ensure trust and security.

3.1.5. User Dashboard

A web interface has been aimed at visualizing pollutant levels, anomaly timelines, model predictions, and recommended actions. Treatment operations executed through the feedback loop are logged for traceability and future reinforcement-based optimization.

3.2. Data Pipeline

A unified data pipeline is used to harmonize the raw sensor streams and prepare them for learning-based inference. The pipeline has different elements, which include data acquisition, preprocessing, feature engineering and data storage, and the detailed role of these different processes is as follows:

- *Data Acquisition*: The readings from the sensors, which are collected through the edge devices, are normalized to standard operating ranges and synchronized using a uniform time base.
- *Pre-processing*: Outliers in the data due to various factors are carefully detected and processed by using interquartile filtering and z-score clipping. Calibration drift is also handled through periodic reference checks and linear gain correction.
- *Feature Engineering*: Statistical, frequency, and derivative-based features are extracted with the help of a sliding window approach of (5–60 s).
- *Storage*: A local time-series database stores curated signals and model outputs. Edge-generated summaries ensure minimal storage overhead.

The unified data pipeline processes approximately 150,000–200,000 sensor records per day across all deployed sensor types. The experimental data stream spans a simulated duration of 30 days, resulting in over 4 million time-stamped observations. Data sources include synthetic sensor streams modeled on published wastewater benchmarks and noise-injected perturbations to emulate sensor drift and environmental variability. Feature extraction is performed on sliding windows of 5–60 s, generating both short-term anomaly indicators and long-term forecasting inputs. Table 1 covers the various characteristics of deployed sensors, captured parameters, sampling rate, extracted features, and respective processing layer within the proposed Edge–AI framework.

Table 1. Characteristics of deployed sensors, captured parameters, sampling rate, extracted features, and respective processing layer within the proposed Edge-AI framework.

Sensor Type/Parameter	Measured Attribute	Sampling Frequency	Extracted Features	Processing Location	Notes
UV Photometric Sensor	Nitrates, organics	1–5 s interval	Absorbance intensity, spectral curve, derivative peaks	Server (Forecasting) + Edge (Threshold-based precheck)	High accuracy for nitrate monitoring
Electrochemical Sensor	pH, conductivity, ORP	2–10 s interval	Mean, variance, rate-of-change, spike detection	Edge (Anomaly Detection)	Low-latency, lightweight computation
Microfluidic Sensor	Heavy metals (Zn, Pb, Cu)	5–30 s interval	Concentration profile, detection peaks, drift compensated values	Server + Edge (compressed features)	Used for early contamination alerting
Dissolved Oxygen Probe	DO concentration	1–3 s interval	Smoothed curves, rolling averages	Edge	Critical for aeration-based treatment
Turbidity Sensor	Suspended solids	2–4 s interval	NTU variation, slope, gradient	Edge → Server	Helps detect sludge imbalance
Temperature Sensor	Water temperature	10 s interval	Rolling mean, seasonal trend	Edge	Supports biological treatment decay models
Flow Rate Sensor	Inflow/outflow velocity	1–2 s interval	Flow variance, shock load indicator	Server	Influences treatment dosing logic

3.3. Forecasting and Anomaly Models

To balance the accuracy and computational efficiency in the proposed system, a hybrid model is designed that works as follows:

- *LSTM Prediction Model:* This model is used and best suited for pollutant concentration forecasting due to the model's efficiency in handling the temporal dimension in the data. The model uses two LSTM layers with dropout regularization followed by dense regression output. The model is effective in predicting the pollutant values 6–24 h ahead.
- *Random Forest Anomaly Classifier:* Trained on temporal signatures of contamination events, the model is designed to detect abnormal readings that are rapidly captured at the edge. Metrics used to evaluate the performance of the system are accurate in capturing the variance in spikes, sudden pH shifts, conductivity jumps, and sensor deviation patterns.
- *Hybrid Fusion Strategy:* The final pollutant state is inferred using the following equation that defines the hybrid fusion approach that is used in the system to handle the anomalies:

$$F_{hybrid} = \propto \times Pred_{LSTM} + (1 - \propto) \times Score_{RF}$$

where α is determined empirically. This improves stability over single-model dependence.

The hybrid model is trained using a 70/20/10 dataset split for training, testing, and validation, respectively. The Adam optimizer was used with a learning rate 1×10^{-3} for 100 epochs. The evaluation metrics are used to ensure that the performance of the designed model helps in the prediction and ensures that the predictions are always monitored and evaluated using a custom-designed algorithm.

3.4. Decision and Feedback Mechanism

Model outputs are then routed to a specially designed LLM-based control decision engine. The module compares the predicted pollutant levels with treatment thresholds and executes action rules using a semantic-logic checklist. The specially designed algorithm analyses various factors and can explain the cause of the issue by gathering the required information from multiple sources. An example of the sample decision rule made for the LLM engine is as follows:

IF (Nitrate_future > threshold) AND (Rise_rate > critical_slope)
THEN Recommend "Increase dosing by X%"
ELSE "Continue monitoring"

The operator's presence and actions are designed to approve and override the recommendations if required. Final decisions are executed and logged back into the system for feedback-driven refinement. Completed actions update the historical database, enabling adaptive learning in future iterations.

The evaluation was conducted using a controlled simulation environment designed to emulate wastewater sensor streams under normal and abnormal operating conditions. Sensor data were generated using statistically grounded pollutant profiles derived from ranges reported in prior wastewater monitoring studies [5,9,12]. The simulation covered a continuous period of 30 days with a sampling interval ranging from 1 to 30 s, depending on sensor type. Controlled pollutant surge events were injected synthetically by increasing nitrate and metal concentration values beyond regulatory thresholds to evaluate anomaly detection, forecasting response, and decision latency. Model training and evaluation followed a fixed 70/20/10 split for training, testing, and validation, ensuring repeatability of results.

4. Results and Discussion

The proposed Edge–AI fusion framework was evaluated through simulation and the controlled testing of pollutant-monitoring workflows. The system performance was assessed in terms of forecasting accuracy, anomaly detection capability, and latency improvements achieved by edge deployment. Three AI models, LSTM, Random Forest (RF), and a Hybrid Fusion model, are also used to benchmark the proposed system and study the predictive capability of the system for factors like nitrate, turbidity, and heavy metal concentration.

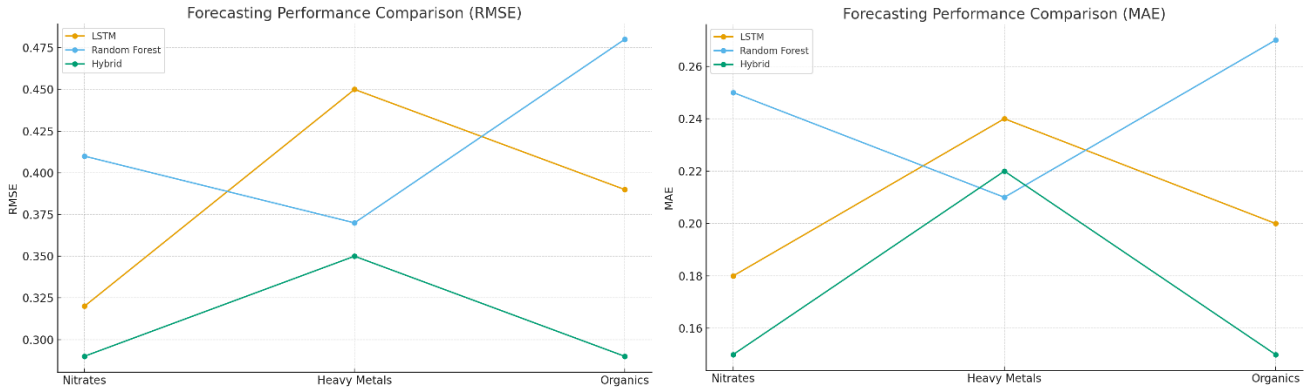
4.1. Forecasting Performance

The forecasting model that is used in the proposed system has demonstrated stable pollutant trend prediction with minimal deviation between actual and predicted curves. The hybrid architecture has achieved better generalization than standalone models, particularly during sudden fluctuations in pollutant levels. Table 2 shows the observed model performance using standard evaluation metrics. As evaluation metrics, we have computed the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). In addition to error-based metrics, Pearson correlation analysis between predicted and observed pollutant concentrations yielded strong positive correlations (r = 0.91–0.94) with statistical significance at $p < 0.01$ across all evaluated pollutants, confirming the robustness of the forecasting model.

Table 2. Forecasting model accuracy using test dataset.

Pollutant Type	LSTM (MAE/RMSE)	RF (MAE/RMSE)	Hybrid (MAE/RMSE)	Best Performing Model
Nitrates	0.18/0.32	0.25/0.41	0.15/0.29	Hybrid
Heavy Metals	0.24/0.45	0.21/0.37	0.22/0.35	Random Forest
Organics	0.20/0.39	0.27/0.48	0.15/0.29	Hybrid

The results indicate that the hybrid ensemble approach gives consistently reliable and accurate predictions when pollutant characteristics exhibit non-linear temporal behavior. The time-series plot shows a high correlation between predicted and actual values, confirming model stability. Figure 2 shows the predicted vs. actual pollutant concentration curves using Hybrid model. Figure 2a shows the forecasting performance comparison using the RMSE values, and Figure 2b shows the forecasting performance comparison using the MAE values.



(**a**) Forecasting performance comparison (RMSE) values (**b**) Forecasting performance comparison (MAE) values

Figure 2. Forecasting performance comparison of LSTM, Random Forest, and Hybrid models: (**a**) RMSE (mg/L) across pollutant types; (**b**) MAE (mg/L) across pollutant types.

4.2. Latency and System Efficiency

In order to test, validate, and evaluate the edge deployment benefits, the inference time of cloud-only and Edge–AI configurations was compared with a sample experiment configuration. Table 3 shows the latency comparison of the computational pipeline deployments in the two different configurations.

Table 3. Latency comparison of computational pipelines.

Mode	Avg Inference Delay	Data Process Location	Observation
Cloud-only	2.3 s	Remote server	Network-dependent; slower response
Edge–AI Fusion	0.8 s	Local edge node	~65% faster anomaly handling

The simulations and the experiments could show that there is a reduction in the latency and demonstrate the suitability of Edge–AI for time-critical water treatment decisions. Initial anomaly detection and data processing are offloaded to the local edge deployments; the summarized signals are then evaluated and transmitted to the central node, which helps in reducing the bandwidth and usage to improve responsiveness and action. Figure 3 shows the latency comparison using the cloud-only and Edge–AI approaches.

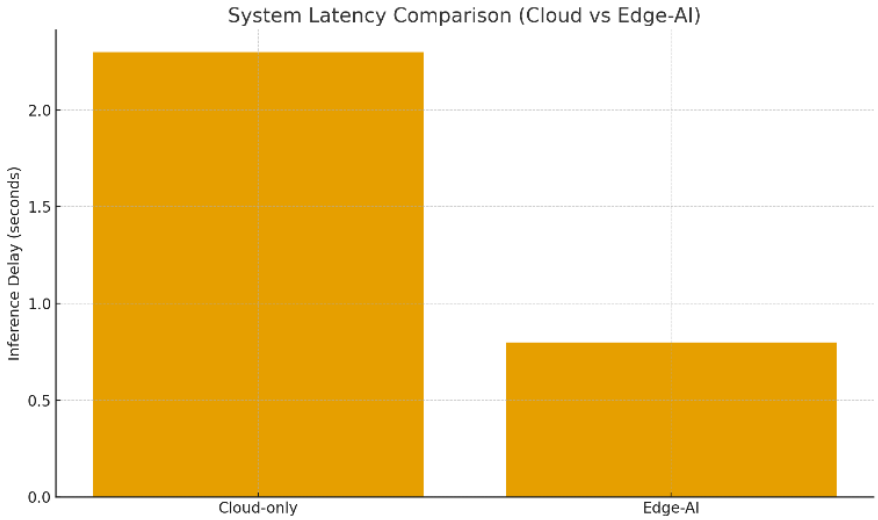


Figure 3. Comparison of inference latency (seconds) between cloud-only and Edge–AI deployments.

4.3. Case Demonstration

To demonstrate the working of the proposed system, a controlled nitrate spike scenario was simulated to evaluate closed-loop decision control. The various steps and actions that happened in the system can be represented as follows:

- The edge node detected abnormal deviation within 1 s.
- Forecast module predicted concentration rise for the next 12 h window.
- LLM engine recommended 15% dosing increase with justification text.
- Dashboard displayed alert → operator confirmed action.
- Post-treatment levels normalized within an acceptable threshold range.

This experiment and simulation flow confirms that combining the forecasting prediction and decision automation reduces the need for manual interaction and thereby reduces the reaction time and supports proactive mitigation.

5. Conclusions

The proposed research presented an Edge–AI fusion framework that is capable of real-time pollutant forecasting and intelligent decision support in wastewater treatment systems. The proposed architecture works by combining multimodal sensor inputs, edge-level anomaly detection, hybrid AI forecasting using AI models like LSTM and Random Forest, with the design of a rule-based LLM decision engine that generates actionable treatment recommendations. The system underwent an experimental evaluation, which demonstrated that the hybrid model improved pollutant prediction accuracy across multiple variables, while the adoption of edge processing significantly reduced system latency compared to cloud-only execution. The results also indicate that the system can transition wastewater management from a reactive approach to a predictive and proactive model, supporting timely dosing decisions, anomaly mitigation, and sustainable resource utilization. The current research helps to establish a real-time water quality monitoring pipeline that is capable of continuous diagnosis, forecasting and automated treatment for cleaning up the water with appropriate action generation, supported by explainable decision-making logic through an integrated LLM layer. A manual oversee dashboard is also designed to verify and validate the observations of the system to ensure smooth predictions and actions minimizing the risk.

The current research presents an Edge–AI fusion framework for real-time pollutant forecasting and decision support in wastewater treatment. Integrating multimodal sensor data with AI models and a rule-based LLM engine, the system improves prediction accuracy and reduces latency, encouraging proactive management. There are certain imitations that exist, such as resource constraints of edge devices, increased power usage during high-frequency processing, and reliance on high-quality historical data, especially for rare contaminants. The framework has only been tested experimentally, not in full-scale plants, and requires policy updates for rule changes. Sensor drift and hardware degradation may also affect performance unless regular recalibration is performed.

Future work needs to focus on extending the scalability and autonomy of the platform. Reinforcement learning-based control loops can be incorporated to give more autonomy for decisions, which can dynamically adapt to the changing water profiles. The integration of a digital-twin simulation environment will also help in testing and validating the performance, load forecasting, and predictive planning before actions are executed in the physical plant. The sensor system needs to be expanded to capture pollutant categories to pharmaceuticals, PFAS, microplastics and other emerging contaminants, to improve the coverage of the proposed system in various other use cases. Automated sensor self-calibration and fault diagnosis are also required to enhance reliability. The proposed system could evolve toward a scalable and autonomous AI-driven wastewater management platform suitable for deployment in large treatment facilities.

Author Contributions: Conceptualization, S.S.R. and V.S.; methodology, S.S.R.; software, A.J.; validation, S.S.R., V.S. and L.V.; formal analysis, S.S.R.; investigation, A.J.; resources, L.V.; data curation, A.J.; writing—original draft preparation, S.S.R. and A.J.; writing—review and editing, V.S. and L.V.; visualization, A.J.; supervision, S.S.R.; project administration, V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data generated as part of the study will be shared on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
ML	Machine Learning
AI	Artificial Intelligence
LLM	Large Language Models
CNN	Convolutional Neural Network
XAI	Explainable Artificial Intelligence
LSTM	Long Short-Term Memory

References

1. Baarimah, A.O.; Bazel, M.A.; Alaloul, W.S.; Alazaiza, M.Y.; Al-Zghoul, T.M.; Almuhaya, B.; Khan, A.; Mushtaha, A.W. Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis. *Case Stud. Chem. Environ. Eng.* **2024**, *10*, 100926. [CrossRef]
2. Maqsood, Q.; Fatima, R.; Rafaqat, F.; Mehmood, T.; Ali, S.W.; Hussain, M. Revolutionizing water and wastewater treatment: Data-driven approaches for advanced solutions. *Desalination Water Treat.* **2025**, *324*, 101432. [CrossRef]
3. Frincu, R.M. Artificial intelligence in water quality monitoring: A review of water quality assessment applications. *Water Qual. Res. J.* **2025**, *60*, 164–176. [CrossRef]
4. Zamfir, F.-S.; Carbureanu, M.; Mihalache, S.F. Application of Machine Learning Models in Optimizing Wastewater Treatment Processes: A Review. *Appl. Sci.* **2025**, *15*, 8360. [CrossRef]
5. Forhad, H.; Uddin, R.; Chakrovorty, R.; Ruhul, A.; Faruk, H.; Kamruzzaman, S.; Sharmin, N.; Jamal, A.S.I.M.; Haque, M.-U.; Morshed, A.M. IoT based real-time water quality monitoring system in water treatment plants (WTPs). *Heliyon* **2024**, *10*, e40746. [CrossRef]
6. Essamlali, I.; Nhaila, H.; El Khaili, M. Advances in machine learning and IoT for water quality monitoring: A comprehensive review. *Heliyon* **2024**, *10*, e27920. [CrossRef] [PubMed]
7. Ren, J.; Zhu, Q.; Wang, C. Edge Computing for Water Quality Monitoring Systems. *Mob. Inf. Syst.* **2022**, *2022*, 5056606. [CrossRef]
8. Zou, S.; Ju, H.; Zhang, J. Water Quality Management in the Age of AI: Applications, Challenges, and Prospects. *Water* **2025**, *17*, 1641. [CrossRef]
9. Nallakaruppan, M.K.; Gangadevi, E.; Shri, M.L.; Balusamy, B.; Bhattacharya, S.; Selvarajan, S. Reliable water quality prediction and parametric analysis using explainable AI models. *Sci. Rep.* **2024**, *14*, 7520. [CrossRef] [PubMed]
10. Durgun, Y. Real-time water quality monitoring using AI-enabled sensors: Detection of contaminants and UV disinfection analysis in smart urban water systems. *J. King Saud Univ. Sci.* **2024**, *36*, 103409. [CrossRef]
11. Xiong, Z.; Liu, X.; Igou, T.; Li, Z.; Chen, Y. Using Hybrid Machine Learning to Predict Wastewater Effluent Quality and Ensure Treatment Plant Stability. *Water* **2025**, *17*, 1851. [CrossRef]
12. Zhi, W.; Appling, A.P.; Golden, H.E.; Podgorski, J.; Li, L. Deep learning for water quality. *Nat. Water* **2024**, *2*, 228–241. [CrossRef] [PubMed]
13. Xu, W. Sustainable Innovations in Shrimp Aquaculture: Current Advances and Future Horizons. *Fishes* **2025**, *10*, 498. [CrossRef]
14. Banjara, I.P.; Poudel, S.; Pariyar, K.; Upreti, D.; Zafeirakou, A.; Paudel, S.R. AI for enhanced water quality data imputation: A deep learning perspective. *Environ. Sci. Adv.* **2025**, *4*, 817–823. [CrossRef]
15. Ling, T.; Singh, V.; Qian, C.; Biessmann, F.; Schiele, G. Automated Energy-Aware Time-Series Model Deployment on Embedded FPGAs for Resilient Combined Sewer Overflow Management. *arXiv* **2025**, arXiv:2508.13905v1.
16. Han, M.; Yao, J.; Law, A.W.-K.; Yin, X. Efficient Economic Model Predictive Control of Water Treatment Process with Learning-Based Koopman Operator. *Control. Eng. Pract.* **2024**, *149*, 105975. [CrossRef]
17. Nagpal, M.; Siddique, M.A.; Sharma, K.; Sharma, N.; Mittal, A. Optimizing wastewater treatment through artificial intelligence: Recent advances and future prospects. *Water Sci. Technol.* **2024**, *90*, 731–757. [CrossRef] [PubMed]
18. Shaban, W.M.; Xie, D.; Elbaz, K.; Shen, S.-L. Real-time water quality prediction of wastewater treatment plants using advanced deep learning networks. *J. Water Process. Eng.* **2024**, *65*, 105775. [CrossRef]
19. Sharma, S.; Sharma, K.; Grover, S. Real-Time Data Analysis with Smart Sensors. In *Application of Artificial Intelligence in Wastewater Treatment*; Springer: Cham, Switzerland, 2024; pp. 127–153. [CrossRef]
20. Dikmen, F.; Demir, A.; Özkaya, B.; Raza, M.O.; Rasheed, J.; Asuroglu, T.; Alsubai, S. AI-driven wastewater management through comparative analysis of feature selection techniques and predictive models. *Sci. Rep.* **2025**, *15*, 25347. [CrossRef] [PubMed]

21. Sakkaravarthy, S.; Jano, N.A.; Vijayakumar, A. Overcoming Challenges in Traditional Waste Water Treatment Through AI-Driven Innovation. In *The AI Cleanse: Transforming Wastewater Treatment Through Artificial Intelligence*; Springer Water; Springer: Cham, Switzerland, 2024; pp. 53–81. [CrossRef]
22. Cheng, T.; Harrou, F.; Kadri, F.; Sun, Y.; Leiknes, T. Forecasting of wastewater treatment plant key features using deep learning-based models: A case study. *IEEE Access* **2020**, *8*, 184475–184485. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.