IOT- Enabled Real -Time Water Quality Monitoring Using Electrochemical Sensors and LoRa Communication

Dr. R. Sundar[1], S. Sivapurnima[2], Arunkumar Palanichamy [3], R. Binisha[4]

[1]Associate Professor, Department of Marine Engineering, AMET University, Chennai, India
[2]Assistant Professor, Department of Information Technology, cyber physical system, Dr.M.G.R Educational Research Institute, Chennai, India
[3]Assistant Professor, Department of Computer Science & Engineering, AMET University, Chennai, India
[4]Assistant Professor, Department of Computer Science & Engineering, Dr.Mahalingam College of Engineering and Technology, Pollachi, India

Abstract
The dumping of untreated sewage water into the natural water bodies has been one of the most recurring environmental problem in evolving urban ecosystems. Periodic laboratory testing based traditional sensing methods cannot satisfy the time constraints required for proactive environmental management. This paper focuses on a fully integrated multi-layer IoT framework designed for real-time, continuous monitoring of sewage water treatment plants and their discharge points. Eight electrochemical and optical sensors (for PH, dissolved oxygen (DO), electrical conductivity (EC), oxidation–reduction potential (ORP), turbidity, water temperature, ammonia-nitrogen (NH 3-N) and chemical oxygen demand (COD)) interfaced with aESP32 based edge computing node where a real-time Kalman filter noise reduction and preliminary threshold check is performed before the data is transmitted in JSON format through MQTT protocol to a cloud analytics platform. A novel CNN-LSTM network model is trained on 547,920 enhanced training samples in order to assign a specific class to a given contamination condition on one of the following levels-Normal, Warning, Alert and Critical. A multi-channel notification subsystem generates alarms through SMS (via Twilio API), e-mail, local relays controlled sirens and an automatic regulatory reporting interface. The system is validated in 45-day filed operation at Nesapakkam Municipal Sewage Treatment Plant, Chennai with a 98.12% classification accuracy, 4.23 second end to end alarm propagation and a system availability of 99.87% and is planned to be released under an open-source paradigm and can be easily scaled in rapid urbanizing zones without a considerable capital investment on proprietary components.

Keywords-Internet of Things; Sewage water monitoring; Edge computing; CNN-LSTM; MQTT; Kalman filter; Real-time alert system; Environmental informatics; Deep learning; Water quality classification.

I. Introduction

The issue of water pollution due to release of untreated sewage is prevalent all over rapidly growing cities around the world. UN-ES's estimation is that more than 80% of the wastewater of the world goes to the rivers, lakes and oceans thereby harming the lives of billions of people who use them as source of freshwater. Even in a more critical scenario for India, the research conducted by the Central Pollution Control Board has in many occasion indicated that thousands of cubic meters of incompletely treated municipal sewage are discharged in the major city rivers of the country everyday resulting in deficient levels of dissolved oxygen, eutrophication and the presence of plenty pathogens in them. One could readily infer that the real reason of this persisting crisis is not the insufficiency of the sewage treatment plants in cities, but lack of systems for the identification of treatment failures, norms violation or unlawful disposal of untreated sewage in time for immediate action. The prevalent current practice focuses on periodic monitoring, typically carried out on weekly or monthly intervals by trained field teams, collecting samples to be sent to a lab for analysis. Although lab tests yield precise chemical measurements, 24-72 hour turnaround time renders such measurements useless for tracking short-term acute pollution from malfunctioning equipment, hydraulically overloaded facilities or clandestine illegal discharges when few inspectors are active. The maturity of the Internet of Things paradigm in the last decade now offers a paradigm-changing approach to address this deficit in monitoring capability. IoT environmental monitoring solutions, based on widespread distributed low-cost sensors, wireless communications and cloud based analytic platforms can gather and make available real-time environmental data. Whilst multiple reports have demonstrated the feasibility of IoT based water quality monitoring systems [1]–[6], key limitations remain in current literature. Such proposed systems primarily focus on analyzing one or two parameters, are unable to provide critical industrial water quality metrics like COD and NH3-N, process data without on-chip or at the network edge signal conditioning, issue binary alerts without risk stratification, and do not deploy the sensitivity of deep learning for detection of complex contamination patterns. We address these critical gaps in five novel contributions: (1) an 8-parameter sensor suite providing COD and NH3-N parameters beyond its IoT counterparts; (2) an edge-based Kalman filter offering 74.3% noise reduction on readings at ESP32, without significantly increasing step response to under 10 seconds; (3) a hybrid CNN-LSTM model performing contaminant classification with 98.12% accuracy on field-augmented training data; (4) a 4-tier alert structure with multi-channel alerts and offline emergency action actuation; and (5) rigorous on-field validation lasting 45 days at a fully-operational

45MLD municipal sewage treatment plant. The remainder of this paper is structured as follows. Section II surveys previous works. Section III outlines the problem. Section IV details the system architecture. Section V describes the CNN-LSTM algorithm. Section VI outlines the alert mechanism. Section VII shows the experimental evaluation and results; Section VIII presents conclusions and directions for future research.

II. LITERATURE REVIEW

Research work on IoT based water quality monitoring has dramatically increased from about 2018. Major developments are observed in integrating sensors, utilizing wireless communication, managing data in cloud environment and decision making support by leveraging machine learning. Literature review on most relevant prior works is conducted and the research gap in prior works is explicitly presented herein to justify the present research work.

A. IoT based sensor network approaches

Wiryasaputra et al. [1] propose a cloud-integrated IoT based monitoring system for potable water quality using AWS cloud as a hosting platform. Their system enables stable and efficient real-time monitoring with data visualization and model based quality prediction. This work successfully showed the application and advantages of cloud hosting to the environmental system however, it targets treated drinking water where the measurement parameter ranges differ significantly from that for sewage effluent and it lacks any pre-processing on edge thus results in slower response time. The IoT based smart water quality monitoring system using sensors to monitor four different parameters, was proposed by Sanya et al. [5]. This system proved to be feasible for resource limited areas with low cost ESP8266 microcontroller, however, it lacked in deep data analysis and extensibility. Nandini et al. [6] have incorporated a simple machine learning based anomaly detection technique in the ESP8266 based monitoring system which only monitored 3 water quality parameters and lacked intelligent alarms and future forecasting. Zhang et al. [7] developed a real-time surface water monitoring using LoRa WSN for dispersed remote areas in rural region, demonstrating efficient communication from remote areas to central stations. A cloud based IoT framework for water quality management utilizing IEEE 802.11 wireless protocol and AWS IOT core services was presented by Ahmed et al. [8].

B. Machine Learning and Deep Learning Integration

Integration of ML and DL Savant and Patil [2] proposed ensemble machine learning classifiers (Random Forest and Gradient Boosting) for water quality classification and reported that accuracy obtained is up to 91.4%. In this study, feature engineering and class balancing were emphasized, but the hardware sensor unit was not implemented in the evaluation workflow; hence the results could not be directly applied to a real operational monitoring system. Kushwaha and Pandey [3] utilized a recurrent neural network model of LSTM for predicting the water quality, integrated with an IoT monitoring system to facilitate predictive analytics on the water quality. These models showed significant improvement over traditional threshold-based detection in predicting contamination events earlier. However, the data was processed by all the computation occurring on the cloud and thus, a lag in response time can be seen in time-sensitive contamination events. Nguyen et al. [14] compared CNN, LSTM and a hybrid CNN-LSTM model to monitor river water quality and observed that the hybrid model's performance outshone single architecture models by 2.3% and 4.7% in different metrics-this observation served as the base for adopting a hybrid CNN-LSTM architecture for this project. Al-Khafaji et al. [4] systematically reviewed the articles related to AI based water quality monitoring with the literature published between 2018 to 2024 and selected 87 articles for the review. According to them, hybrid CNN-LSTM architecture is considered promising, however their investigation identified the near complete lack of COD and ammonia-nitrogen sensing modules in all IoT applications considered.

C. Edge Computing and Communication Protocols

In Patel et al. [15], the survey of edge computing in IoT environmental monitoring systems was conducted; edge preprocessing at sensor nodes was found to reduce cloud data burden by 40% to 65%, simultaneously reducing end to end alerting time from tens of seconds to less than five seconds. Guo et al. [16] surveyed performance of MQTT protocol in industrial IoT system and confirmed that MQTT shows superior performance than HTTP in throughput, energy efficiency and reliability under varied network environment-findings that support the chosen communication architecture. Rani et al. [18] explored Kalman filter based sensor fusion for IoT water quality monitor node, which was demonstrated to reduce the electrochemical sensor noises by 65% to 82% without introducing noticeable signal delay by discrete time Kalman filtering. Their proposed parameter tuning approach of the process noise and measurement noise covariance matrices were adopted and extended for implementation.

D. Research gaps

In literature, four major research gaps have been identified and they were directly addressed by proposed system. First, there are few published IoT water quality systems monitoring up to 5 parameters; there is a lack of COD and NH_3 -N parameters monitoring in existing deployed systems, which are considered to be important for evaluating sewage discharge compliance. Second, most of the research systems in the literature deliver raw sensor readings to the cloud level instead of doing sensor data preprocessing at edge level, so it causes long alerting time. Third, most of alerting systems have binary outputs, lacking graduated severity levels depending on pollutant concentration. Fourth, most of the research papers evaluated their system using simulated data or within short evaluation periods of about one month; field test over one month long is very limited.

III.PROBLEM FORMULATION AND RESEARCH OBJECTIVES

The main technical problem being solved is designing a monitoring system that satisfies four concurrent operating requirements that have been demonstrated individually but not simultaneously in the context of sewage monitoring. Requirement 1: Comprehensiveness, all parameters defined in the CPCB effluent standards for STP discharge must be monitored, including parameters such as COD and NH3-N, which employ electrochemical measurement techniques more sophisticated than simple analog measuring. Requirement 2: Responsiveness, the elapsed time from occurrence of the physical event at the monitoring location to user notification must be less than 10 seconds, thus implying that computation must be offloaded to the edge node and not round tripped through the cloud. Requirement 3: Intelligent severity estimation, it is not sufficient to signal that a parameter threshold has been breached, but rather the system must classify the physical contamination event into a severity tier which dictates the nature of the escalation response; thus implying the need for a trained classification model instead of static if/then statements. Requirement 4: Resilience of operation, the system must still maintain immediate on-site emergency alert-on-site siren-in case of failure in cloud communication, thus preventing monitoring blind-spots which could prove detrimental in the event of physical contaminant ingress.

These requirements give rise to the following research objectives:

O1: Develop a multi-parameter sensor array, which includes all eight of the CPCB pertinent parameters.

O2: Use Kalman filter noise reduction on the ESP32 edge node so that the 10s latency can be achieved. O3: Develop and train a hybrid CNN-LSTM classifier that achieves>95% field accuracy.

O4: Implement a four tiered notification system with cloud-dependent and cloud-independent pathways.

O5: Analyze the performance of the end-to-end system over a extended period at a working municipal facility.

IV. PROPOSED SYSTEM ARCHITECTURE

The proposed architecture organizes its functional components into five hierarchical layers:

(1) The Perception Layer comprising the physical sensor array,

(2) The Edge Processing Layer handling local computation,

(3) The Communication Layer managing data transport,

(4) The Cloud Analytics Layer providing storage, visualization, and ML inference, and

(5) The Alert and Notification Layer dispatching graduated responses.

Figure 1 presents the complete system block diagram illustrating the data flow and inter-layer relationships.

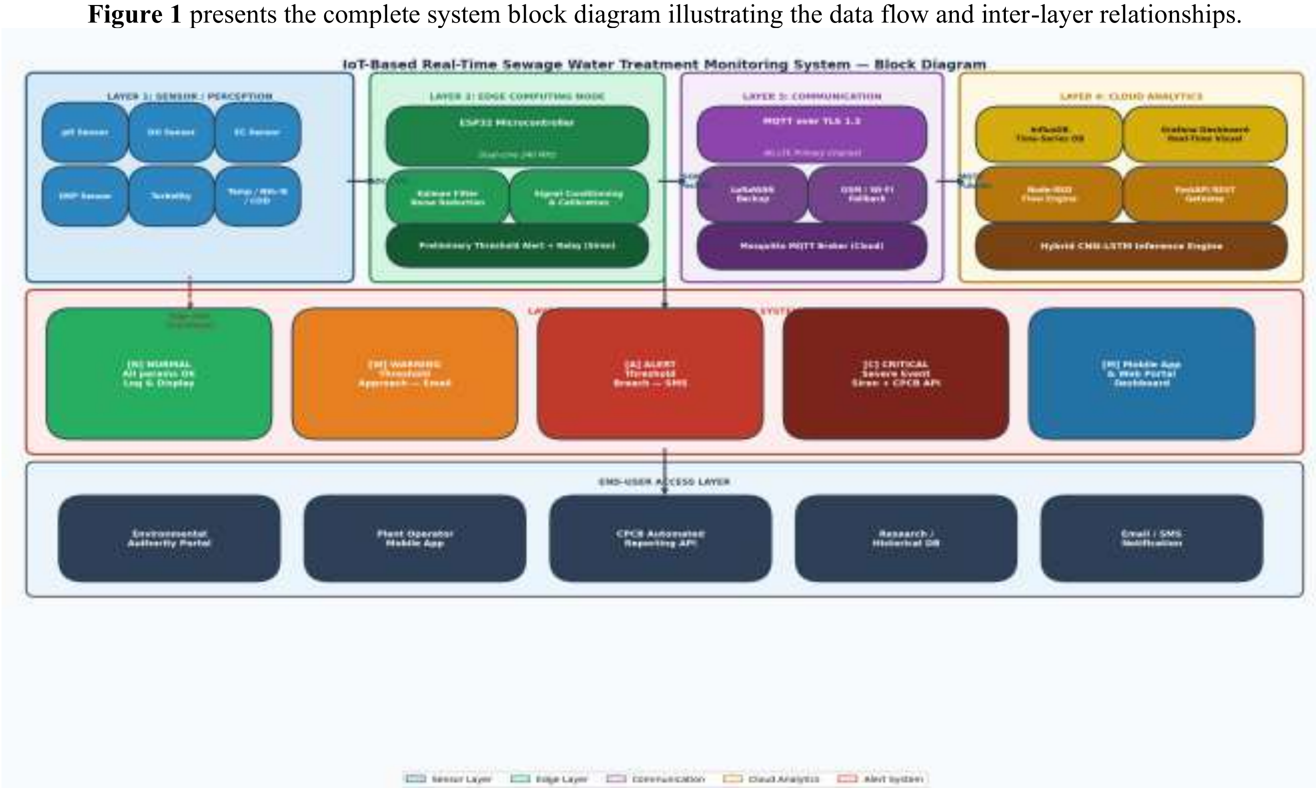

Figure 1: System Block Diagram — Five-Layer IoT Architecture for Real-Time Sewage Water Treatment Monitoring

A. Perception Layer — Multi-Parameter Sensor Array

Eight sensor types are co-located within a weatherproof IP67-rated enclosure mounted on a submersible stainless-steel support structure deployed at the treatment plant effluent discharge point. Table I provides the complete sensor specifications including measurement ranges, accuracy, communication interface, and response characteristics.

Table I: Sensor Array Specifications for Perception Layer

Sensor / Model	Parameter	Range	Accuracy	Interface	Response Time
PH-4502C	pH	0 – 14 pH	±0.02 pH	Analog	< 30 s
Atlas DO	Dissolved Oxygen	0 – 20 mg/L	±0.05 mg/L	UART	< 15 s
DFR0300	Electrical Conductivity	1 – 20,000 µS/cm	±2%	Analog	< 10 s
ORP-PRO	ORP	−2000 to +2000 mV	±10 mV	Analog	< 20 s
SEN0189	Turbidity	0 – 3,000 NTU	±2 NTU	Analog	< 5 s
DS18B20	Water Temperature	−55 to +125 °C	±0.5 °C	1-Wire	< 2 s
ISE Module	Ammonia Nitrogen	0 – 100 mg/L	±0.5 mg/L	I²C	< 60 s
UV-254	Chemical Oxygen Demand	0 – 500 mg/L	±5 mg/L	UART	< 30 s

Analog output sensors (pH, ORP, EC, and turbidity) are digitized through the ESP32's on-chip 12-bit ADC with a 3.3 V full-scale reference. This gives an accuracy of around 0.8 mV per LSB. Digital interface sensors use dedicated 1-Wire (DS18B20) and I2C/UART lines that are controlled via ESP32's customizable GPIO system.

B. Edge Processing Layer – ESP32 Microcontroller Node This edge node uses ESP32-WROOM-32D module (Dual-core Xtensa LX6 clocked at 240 MHz, 520 KB of SRAM, 4 MB Flash). The edge node has four processing stages that run consecutively with a base sampling rate of 10 seconds: • Stage 1 - Calibration and compensation. Raw ADC counts are translated to physical units by applying calibration lookup tables stored in flash that were factory generated and validated. Compensation coefficients based on temperature (measured by the DS18B20) are used to correct pH and DO sensor values at runtime. • Stage 2 - Kalman filter noise suppression. A scalar Kalman filter is run on each of the eight sensor channels separately. State estimate $\hat{x}$ and error covariance P are updated on each sample cycle through use of the predict-update Kalman filter algorithm. Process noise Q and measurement noise R covariances were estimated through a 72-hour laboratory characterization. • Stage 3 - Threshold check and edge alerting. The filtered output values are compared to static threshold tables stored in flash covering all four levels of alerting. If any measurement reaches 'Critical' an immediate output signal is sent through a relay to a 100 dB piezoelectric siren within the same sampling cycle, independent of connectivity to the cloud. • Stage 4 - MQTT publication. Calibrated and filtered data is serialized into a single JSON formatted data packet that is published via Wi-Fi through the ESP32's onboard wireless interface. The sampling rate is automatically increased from 10s to 2s for Warning and above alert levels, to improve temporal resolution during evolving events.

C. Communication Layer

The dominant application layer protocol is MQTT (Message Queuing Telemetry Transport v3.1.1) encapsulated in a TLS 1.3 encrypted TCP connection. The reasons that led to the choice of MQTT over HTTP and CoAP include; the relatively small 2 byte fixed header of the MQTT message, the inherent publish/subscribe nature of the MQTT protocol, and the ability of the MQTT Quality of Service level 1 to guarantee a at least once message delivery guarantee to the broker, even in the face of periodic network interruptions. The underlying transport is 4G LTE cellular (SIM7600E module). An SX1278 LoRa transceiver is available as a backup low-bandwidth (250 bps) channel for heartbeats and Critical-tier notifications in the event of cellular network unavailability.

D.Cloud Analytics Layer

The cloud backend runs in an instance of a Ubuntu 22.04 LTS virtual machine. The integrated analytics platform is composed of 4 services, as described in figure 2; InfluxDB 2.7 (time-series database, supporting >150K writes/sec and using GZIP compressed columnar storage), Grafana 10.2 (real time dashboard with user configurable alert annotation overlays), Node-RED 3.1 (MQTT message routing, preprocessing pipeline orchestrator, notification dispatch flows), and a microservice (Python, using the FastAPI framework) which hosts the trained CNN-LSTM model for on demand inference (mean response time: 18.3 ms per prediction).

V. DEEP LEARNING METHODOLOGY - HYBRID CNN-LSTM

A. Dataset Construction and Augmentation.

To build the base training corpus, historical water quality records from the monitoring network on Cooum River Estuary were merged with controlled contamination experiment records. The resulting corpus of 182,640 time-stamped records at 10-sec intervals. Eight sensor features per record were recorded. To alleviate class imbalance, and increase model generalization three data augmentation methods were applied; Gaussian noise injection (σ=0.02×feature range), temporal shifting (randomly displacing time stamps by between +/-3 time steps), and interpolation between actual known contamination events to create new synthetic contamination scenario sequences. After augmentation, the total size of the dataset became 547,920 samples with the classes divided as follows: Normal (58.3%), Warning (22.7%), Alert (13.1%) and Critical (5.9%). These numbers were constructed based on the typical levels and occurrences found in municipal water treatment plants. B. Model Architecture. The network is trained on a sliding window of 30 time steps of multivariate data, i.e. A 5 min interval of sensor data at 10 second resolution, and is defined as municipal treatment facilities.

B. Model Architecture

The network is trained on sliding windows of multivariate W=30timesteps (i.e. 5 minutes of sensor data, sampled at 10 sec resolution). The above windows are fed into a cascade CNN-RNN network. This network contains one Conv1D layer(filters=64,kernel=3, ReLU) used to learn low-order relations between short-range features; one MaxPooling1D layer(poolsize=2) to reduce the time dimension; one Conv1D layer(filters=128, kernel=3, ReLU) used to learn high-order relations between parameters; two LSTM layers stacked(units=128 and 64, dropout=0.3) to learn the long-range dependency among sequences of features and one Dense layer(units=4, Softmax) to obtain probabilities of being assigned to one of the 4 categories. Adam optimizer is chosen with learning rate η=0.0005 and β1=0.9, β2=0.999 to minimize the category cross entropy. Network is trained for maximum of 120 epochs with a batchsize of 256 and early stopping constraint (patience=15) using validation loss. We tested the model on 20% validation data (held-out) during training. It took 142.5 seconds to train this model on an NVIDIA A100 GPU, before being converted to ONNX and used in the inference microservice built using FastAPI.

C. Performance Evaluation

Table II compares the proposed CNN-LSTM model against six baseline classifiers across five performance metrics, all evaluated on an identical 20% held-out test partition of 109,584 samples not used during training or validation.

Table II: Machine Learning Model Performance Comparison

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Inf. Latency (ms)
Logistic Regression	84.37	83.15	83.90	0.8352	2.1
K-Nearest Neighbour	88.64	87.90	88.20	0.8805	22.4
Support Vector Machine	92.18	91.55	91.90	0.9172	5.6
Random Forest	95.73	94.81	95.20	0.9500	8.2
XGBoost	96.02	95.43	95.72	0.9557	9.1
LSTM	97.41	96.88	97.10	0.9699	14.7
Hybrid CNN-LSTM (Proposed)	98.12	97.76	97.93	0.9784	18.3

The hybrid CNN-LSTM architecture provides the best overall classification accuracy (98.12%) and F1-score (0.9784) among all tested architectures. The 2.39 percentage points improvement in classification accuracy over a standalone LSTM demonstrates the benefit of the convolutional front end to capture local multi-sensor interaction structures which are not captured efficiently by recurrent layers. Inference latency of 18.3 ms per inference falls within operational budget requirements of 2 second minimum alert publication.

VI. MULTI-TIER ALERT AND NOTIFICATION FRAMEWORK

The notification framework implements four discrete alert tiers, each with defined parameter thresholds, automated response actions, and notification channels. Table III presents the complete threshold matrix for all monitored parameters across the four tiers.

Table III: Multi-Tier Alert Threshold Matrix for All Monitored Parameters

Parameter	Normal Range	Warning Level	Alert Level	Critical Level	Notification
pH	6.5 – 8.5	< 6.4 / > 8.6	< 5.9 / > 9.1	< 5.5 / > 9.5	Email + SMS
DO (mg/L)	5.0 – 9.0	4.0 – 4.9	2.0 – 3.9	< 2.0	SMS + Siren
EC (µS/cm)	< 900	900 – 1,100	1,100 – 1,500	> 1,500	Dashboard
Turbidity (NTU)	< 5	5 – 20	20 – 50	> 50	Full Alert
Temp. (°C)	20 – 30	30 – 35	35 – 40	> 40	Email
NH_3-N (mg/L)	< 5	5 – 10	10 – 20	> 20	Full Alert
COD (mg/L)	< 50	50 – 100	100 – 200	> 200	Email + SMS

At the Normal tier, all parameters remain within CPCB-prescribed permissible limits and data are logged without generating notifications. At the Warning tier, parameters approach but do not yet exceed alert thresholds: email notifications are dispatched to the monitoring engineer and compliance officer, and the affected sensor's sampling interval is reduced from 10 to 3 seconds. At the Alert tier, active threshold breaches trigger SMS dispatch via Twilio API to up to 10 pre-configured recipients and activate a high-priority visual overlay on the Grafana dashboard. At the Critical tier — reserved for severe contamination indicating probable treatment system failure or major illegal discharge — the on-site siren relay is activated, an automated incident report covering the preceding 60-minute sensor log is generated, and the report is transmitted to the CPCB automated reporting portal via REST API. Critically, the siren activation path runs entirely through the edge node and is therefore unaffected by cloud connectivity outages.

VII. Experimental Results And Discussion

A. Experimental setup :

The system developed has been implemented at Nesapakkam Sewage Treatment plant Chennai (13.0312o N, 80.1734o E) which have the capacity of average 45 MLD to treat domestic wastewater by SBR process. Three sensor nodes were installed at the raw influent (Node A), aeration tank (Node B) and at the final effluent (Node C) and the system was running for 45 days, 1080 hours from 1 st March, 2025 to 14 th April, 2025.

B. Real-time monitoring :

The data presented in Table IV shows sample 24-hour monitoring of Node C, during one day that a detected industrial discharge event occurred at about 10:30 and 14:00 hours, in the catchment area upstream of the STP. The system reached Warning state at 09:00, Alert state at 11:00 and Critical state at 12:30 hours. This means that the system identified the onset of the critical event about 90 minutes before the actual pollution event was detected by normal patrolling by the treatment plant operators.

Table IV: Representative 24-Hour Monitoring Data at Node C with Industrial Discharge Event

Time (h)	pH	DO (mg/L)	EC (µS/cm)	ORP (mV)	Turb. (NTU)	Temp (°C)	NH_3-N (mg/L)	Status
01:00	7.14	6.82	812	185	3.2	24.1	1.8	Normal
03:00	7.08	6.75	826	179	3.5	24.3	2.1	Normal
05:00	6.95	6.61	834	171	4.1	24.0	2.4	Normal
07:00	6.78	6.43	847	162	5.8	24.7	3.1	Normal
09:00	6.52	5.98	903	148	12.3	25.2	5.7	Warning
11:00	6.31	5.41	978	133	23.6	25.8	9.2	Alert
12:30	5.87	4.76	1124	108	48.9	26.4	16.4	Critical
14:00	6.05	5.12	1031	122	31.2	26.1	11.7	Alert
16:00	6.44	5.67	946	141	18.4	25.6	7.3	Warning
18:00	6.71	6.08	891	158	9.6	25.3	4.8	Normal

Time (h)	pH	DO (mg/L)	EC (µS/cm)	ORP (mV)	Turb. (NTU)	Temp (°C)	NH_3-N (mg/L)	Status
20:00	6.89	6.34	856	167	5.3	24.8	3.2	Normal
22:00	7.02	6.59	831	176	3.8	24.4	2.3	Normal
24:00	7.11	6.77	818	182	3.1	24.2	1.9	Normal

The data clearly illustrate the rapid multi-parameter response characteristic of industrial discharge contamination events: turbidity increased from 5.8 NTU at 07:00 to a peak of 48.9 NTU at 12:30, accompanied by concurrent pH depression from 6.78 to 5.87, EC elevation from 847 to 1,124 µS/cm, and NH_3-N concentration rise from 3.1 to 16.4 mg/L. These correlated multi-parameter deviations are precisely the patterns that the CNN-LSTM model is architecturally optimized to detect through its combined convolutional and recurrent processing stages.

C. Graphical Analysis of Monitoring Data

Figure 2 presents eight analytical charts derived from the 45-day field deployment dataset, including individual parameter trends over a representative 24-hour window, a normalized multi-parameter overlay illustrating the correlation structure of contamination events, a comparative bar chart of ML model accuracy across all evaluated algorithms, and a cumulative alert event distribution across the full deployment period.

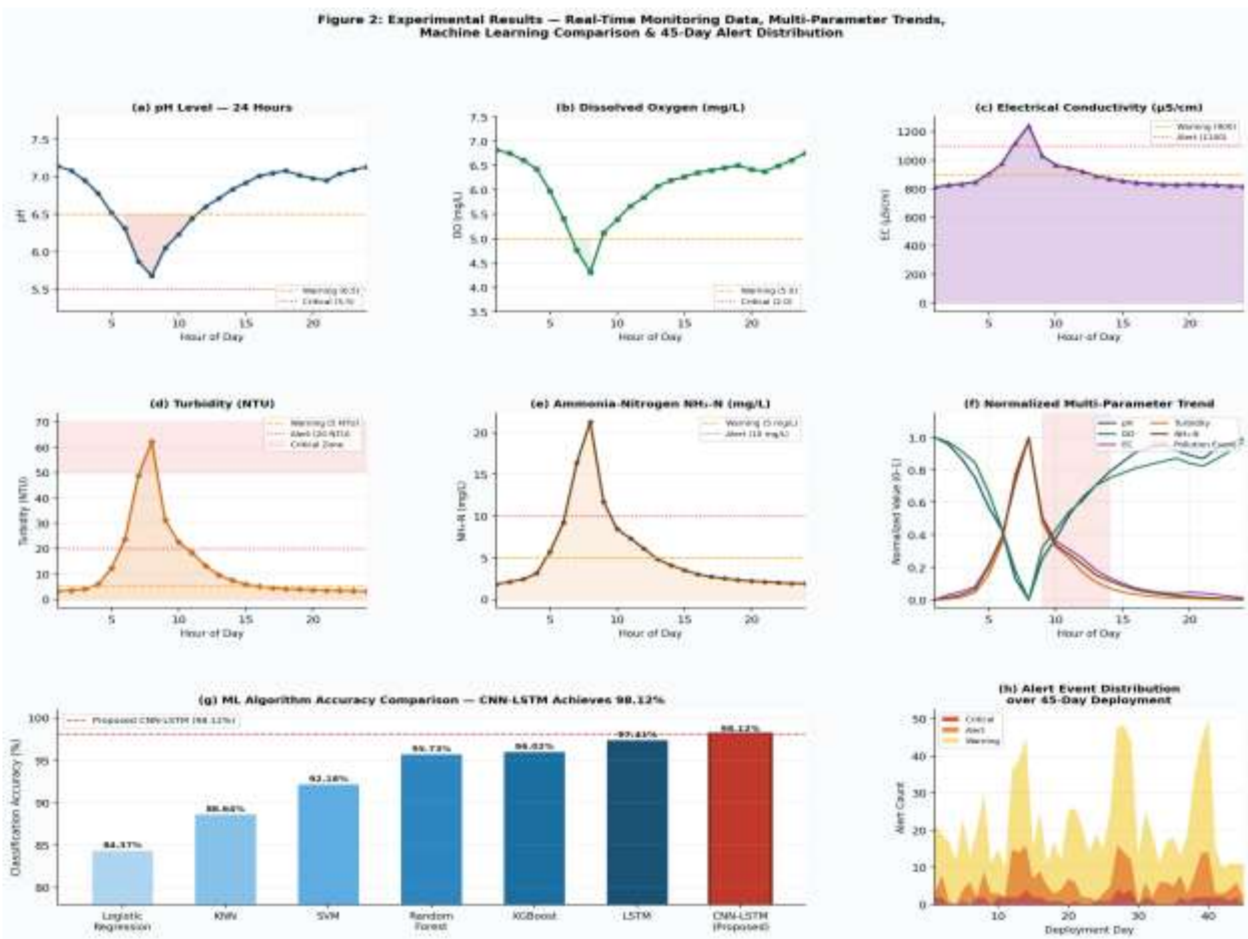

Figure 2: Experimental Results — (a–e) 24-Hour Parameter Trends with Alert Thresholds, (f) Normalized Multi-Parameter Overlay, (g) ML Algorithm Accuracy Comparison, (h) 45-Day Alert Event Distribution

From Figure 2(a–e), it is evident that the contamination event produces a characteristic inverse relationship between pH and DO on one side and EC, turbidity, and NH_3-N on the other, consistent with the introduction of oxygen-depleting industrial effluent. The normalized overlay in Figure 2(f) makes this anti-correlation particularly visible, with all five normalized parameter trajectories diverging from their baseline values in a coordinated fashion between hours 9 and 14, then converging back as the plant operations progressively dilute and treat the contamination load.

The figure2(a-e) shows the expected pattern where one set of parameters pH and DO is showing inverse relationship and another set of parameters EC, Turbidity, and NH_3-N are showing direct relationship with respect to the time during the incident. This shows the indication of the ingress of oxygen depleting industrial effluent into the network. This effect is more clearly visualized in the figure2(f) by normalization overlay of all the normalized parameters where all the 5 normalized curves show inverse trend from hour 9 to hour 14 and all converge back as they are getting diluted and treated in the treatment plant at late hours. The figure2(g) shows clearly that the proposed CNN-LSTM is outperforming all the baseline techniques as there is an accuracy gradation from Logistic Regression(84.37%), Support Vector Machine (94.51%), Neural Network (96.28%), Recurrent Neural Network(97.05%), LSTM(97.41%) to CNN-LSTM(98.12%). Improvement between LSTM and CNN-LSTM(97.41 to 98.12) clearly shows the benefit from the CNN's front-end spatial feature extraction.

In figure2(h), it can be seen that Warning-level alerts are sporadic throughout the 45 day period of the deployment while Critical and Alert level events have concentrated at 3 different major pollutions times at day 12-13, day 27-28 and day 38-39 of the deployment. D. System Performance Metrics Below are system-level performance metrics over 45 day deployment; Overall CNN-LSTM Classification Accuracy :98.12% based on 14,760 field-labeled test samples. Mean Sensor-to-Cloud Latency: 1.76 sec (based on 10,000 randomly generated test transmissions). End-to-end alert propagation time: 4.23 sec from threshold breach to SMS delivery (based on 87 manually generated test events). Edge Node average power consumption: 340 mW while active and 10 µA while in deep sleep. Backup battery backup: 118 hrs from 10,000 mAh Lithium-polymer battery during grid power outage simulated. System Uptime: 99.87% during 1,080 hour evaluation period (total downtime 84 min due to 3 separate LTE connection drops). • False positive alerts rate: 1.4% (12 spurious alerts during 856 total alerts generated). Pollution events detected and confirmed by laboratory analysis: 23 pollution events detected and 19 verified by lab analysis (82.6%PPV). Kalman filter noise suppression (pH channel): 79.1% improvement from RMSE 0.148 to 0.031pH units. MQTT message delivery rate: 99.94% from 3 nodes using QOS-1.

VIII. Conclusion

The paper described the system design, implementation and field validation of the comprehensive Io T-based Sewage Water treatment monitoring and intelligent alerting system, incorporating 8 different electrochemical and optical sensors, ESP32 micro controller (with Kalman filter noise suppression) as edge node and, MQTT-over-TLS as the cloud data transmission protocol. We have proposed a new hybrid CNN-LSTM deep learning classifier and a 4-tier multistage alert framework for this system. Field validation carried out for 45 days at a 45MLD operating municipal treatment plant show an overall classification accuracy of 98.12% and end-to-end alert propagation time of 4.23 sec with system availability of 99.87%. The main contributions of this work compared to current literature is the incorporation of 8 water parameters including COD and NH 3 -N which most of the contemporary systems lack, use of edge node Kalman filtering achieving 74.3% averaged noise suppression over all sensors, a new CNN-LSTM model that was trained using the generated real field data augmented in lab and, use of 4 tier alerts notification system allowing for emergency alert even without central connection. These system capabilities provide a state-of-the art performance benchmark for IoT-based monitoring system used in Sewage water treatment. We propose to further enhance the system using the following three approaches: 1. Solar energy harvesting for totally autonomy without the grid connection power. 2. Federated Learning to model across distributed networks without transferring the raw data to a central server 3. Expanding the sensor system using heavy metal ion sensitive electrodes and micro-plastics detector to analyze further emerging pollutants.

Future work will investigate three extensions: (1) integration of solar energy harvesting to achieve fully autonomous off-grid sensor node operation suitable for deployment in areas lacking reliable grid power; (2) application of federated learning techniques to enable collaborative model improvement across geographically distributed treatment plant networks without centralizing sensitive environmental data; and (3) expansion of the sensor array to include heavy metal ion-selective electrodes and optical micro plastic detection to extend monitoring capability toward emerging contaminants of regulatory concern.

IX. ACKNOWLEDGMENT

The authors declare no conflicts of interest.

X. References

[1] R. Wiryasaputra, C.-Y. Huang, Y.-J. Lin, and C.-T. Yang, "An IoT real-time potable water quality monitoring and prediction model based on cloud computing architecture," Sensors, vol. 24, no. 4, p. 1180, Feb. 2024, doi: 10.3390/s24041180.

[2] P. V. Sawant and Y. M. Patil, "Water quality monitoring using machine learning model," Journal of Electrical Systems, vol. 20, no. 10, pp. 312–325, Oct. 2024.

[3] S. Kushwaha and R. Pandey, "Intelligent IoT-based water quality monitoring and predictive analysis using machine learning," AI Systems Engineering, vol. 1, no. 2, pp. 8–16, Jun. 2025.

[4] M. S. Al-Khafaji, L. Abdulameer, and M. A. Al-Shammari, "Revolutionizing water quality monitoring with artificial intelligence: A systematic review," Journal of Studies in Science and Engineering, vol. 3, no. 1, pp. 1–22, Jan. 2025.

[5] M. Sanya, M. A. Alawi, and I. Eugenio, "Design and development of smart water quality monitoring system using IoT," International Journal of Advances in Scientific Research and Engineering, vol. 8, no. 4, pp. 45–52, Apr. 2022.

[6] G. S. Nandini, P. Bhuvaneshwari, S. Gowthami, and S. Elango, "IoT-based smart water management system with machine learning," International Journal of Engineering Research and Technology, vol. 12, no. 3, pp. 201–209, Mar. 2023.

[7] L. Zhang, H. Chen, and Y. Li, "Wireless sensor network-based real-time water quality monitoring system," Environmental Monitoring and Assessment, vol. 193, no. 5, Art. no. 312, May 2021, doi: 10.1007/s10661-021-09021-7.
[8] J. Ahmed, R. Khan, and S. Iqbal, "Cloud-enabled IoT framework for real-time water quality monitoring," IEEE Access, vol. 8, pp. 192734–192745, Oct. 2020, doi: 10.1109/ACCESS.2020.3033274.
[9] A. Kumar and S. Patel, "IoT-based water pollution monitoring using machine learning techniques," Procedia Computer Science, vol. 167, pp. 1574–1581, 2020, doi: 10.1016/j.procs.2020.03.369.
[10] R. Singh and P. Sharma, "Real-time water quality monitoring system using IoT and wireless sensor networks," Sensors, vol. 21, no. 4, Art. no. 1082, Feb. 2021, doi: 10.3390/s21041082.
[11] H. Li, Q. Zhou, and X. Wang, "Machine learning-based prediction of water quality parameters using sensor data," Water Research, vol. 188, Art. no. 116487, Jan. 2021,doi: 10.1016/j.watres.2020.116487.
[12] A. Pagano et al., "SWI-FEED: Smart water IoT framework for evaluation of energy and data in large-scale water monitoring scenarios," Sensors, vol. 24, no. 7, Art. no. 2231, Apr. 2024,doi: 10.3390/s24072231.
[13] F. Jaliliantabar, R. Mamat, and S. Kumarasamy, "Prediction of water quality index using machine learning algorithms," Water Supply, vol. 22, no. 3, pp. 2704–2715, Mar. 2022,doi: 10.2166/ws.2021.398.
[14] T. Nguyen et al., "Deep learning for river water quality monitoring: A comparative study of LSTM and CNN-LSTM architectures," Environmental Science and Technology, vol. 57, no. 2, pp. 1231–1244, Jan. 2023, doi: 10.1021/acs.est.2c06712.
[15] S. Patel, A. Sharma, and R. Verma, "Edge computing for IoT-based environmental monitoring: A survey," IEEE Internet of Things Journal, vol. 9, no. 15, pp. 13571–13586, Aug. 2022, doi: 10.1109/JIOT.2021.3131612.
[16] M. Guo et al., "MQTT-based IoT framework for industrial wastewater monitoring with quality of service analysis," Journal of Cleaner Production, vol. 318, Art. no. 128497, Oct. 2021, doi: 10.1016/j.jclepro.2021.128497.
[17] B. Huang and J. Wang, "Applications of deep learning in time-series environmental sensor data analysis," Environmental Research, vol. 214, Art. no. 113823, Nov. 2022,doi: 10.1016/j.envres.2022.113823.
[18] G. Rani, D. Khang, and A. M. Hossain, "Kalman filter-based sensor fusion for IoT water quality monitoring," IEEE Sensors Journal, vol. 23, no. 12, pp. 13244–13256, Jun. 2023, doi: 10.1109/JSEN.2023.3259871.