

Artificial Intelligence and IoT-Enabled Chemical Pollution Monitoring and Remediation of Rivers: A Comprehensive Review of Sensing Technologies, Machine Learning Methods, and IoT Integration

Sarthak Sunil Khandagale, Prof. Bangar Abhishek Popat, Prof. S. B. Bhosale, Dr. A.A.Khatri

Dept. of Computer Engineering
Jaihind College of Engineering, Pune, India

Abstract- Rivers receive chemical contamination from industrial effluent, domestic sewage and agricultural runoff, including nutrients such as phosphorus, organic load, and other pollutants. Conventional assessment relies on periodic sampling and laboratory analysis, which needs specialised staff, chemicals and expertise, and therefore cannot give continuous, early warning of pollution events. Artificial intelligence (AI), combined with low-cost sensing and the Internet of Things (IoT), is increasingly proposed for continuous river monitoring and for the data-driven control of chemical treatment such as coagulant dosing. This paper presents a focused narrative review of studies that use machine learning (ML) and deep learning (DL) with IoT sensing for river water quality monitoring and chemical-treatment control. It examines three connected themes: sensing technologies, learning methods, and IoT architectures. The reviewed literature reports water-quality classification accuracy of up to 98 % with an Arduino-based IoT device, XGBoost-based prediction in an industrially affected river basin, ML-assisted phosphorus sensing in river samples, and a deep-learning coagulant-dosing model that reduced dosage by about 22 % in a plant study. However, most studies are small, single-site, short-term and laboratory-oriented, and few report field validation, sensor fouling and drift, or closed-loop treatment control in rivers. The paper contributes a comparative table of representative studies, a conceptual river-to-decision architecture (not experimentally validated), and a prioritised list of research gaps.

Keywords— River Pollution, Water Quality Monitoring, Machine Learning, Internet of Things, Chemical Treatment, Edge Computing

I. INTRODUCTION AND RESEARCH OBJECTIVES

River water quality depends on physicochemical conditions such as pH, turbidity, temperature, conductivity, dissolved oxygen and nutrient and pollutant concentrations. Industrial discharges, domestic effluents and agricultural runoff degrade these conditions, as discussed for the Mahanadi River basin [5]. Traditional monitoring relies on laboratory tests that require specialised staff, chemicals and expertise, and is often insufficient for complex and dynamic pollution problems [2]. Cleaning the chemical load of polluted water, whether at an effluent outfall or a treatment plant, also depends on chemical processes such as

coagulation, where the required dose is affected by many interacting factors and is hard to determine optimally [6].

Three developments converge: low-cost sensors and microcontrollers that measure water parameters continuously, ML and DL models that map sensor signals to quality classes, concentrations or chemical demand, and IoT connectivity that moves data from the riverbank to dashboards and decision systems [2], [3], [5].

This review examines these strands together, asks how mature the evidence is for field deployment, and proposes a conceptual architecture that addresses the weaknesses identified.

II. REVIEW METHODOLOGY

A focused web and publisher-platform search (MDPI, Wiley, Royal Society of Chemistry, Elsevier, university repositories and arXiv) was carried out using combinations of the terms river water quality, IoT, machine learning, deep learning, chemical dosing, coagulation and wastewater treatment, prioritising 2023–2025 publications. Studies were included when they reported river- or treatment-specific sensing, prediction or control with a stated method. Direct logged querying of Scopus and Web of Science was not performed, so the review is narrative rather than systematic and study counts are not reported. Some sources (a student thesis, a repository record and a conference chapter) were available only as abstracts or repository pages and are marked as such in Table I.

III. SENSING TECHNOLOGIES

1. Low-Cost Physicochemical Sensor Nodes

An IoT device built on an Arduino UNO microcontroller with multiple sensors collected data from three water sources (potable water, flowing river water and stagnant puddle water) at Sungai Pusu, Malaysia [3]. The related study reports a dataset of more than 879 data points with seasonal changes considered [2]. Parameters such as turbidity, pH and total suspended solids are typical targets of such nodes [2].

2. Electrochemical Sensing of Nutrients

Lithography-free 3D-printed sensors were used to estimate phosphorus levels in river water, using samples from five locations on the Rappahannock River, Virginia, USA, many of which already contained treated-sewage inflow; ML models were trained on 2,000 sensor data points, half from standard phosphate solutions and half from river samples [4].

3. Vision and Process Signals in Treatment

In treatment systems, image-based DL models can relate floc morphology to coagulant dosage, and hybrid models that combine AI with regression or physical and kinetic models show improved prediction [9], [1].

IV. MACHINE LEARNING AND DEEP LEARNING METHODS

1. Classical Machine Learning

In the Pusu river work, ML classified water quality conditions with up to 98 % accuracy [3]. For the Mahanadi River, an IoT monitoring system was combined with the XGBoost model to address the limits of traditional evaluation, with seasonal contamination analysis across river belts [5]. Heddam used extremely randomised trees, random forest and multiple linear regression to predict coagulant dosage at a drinking-water plant, as summarised in [6].

2. Deep Learning for Dosing and Forecasting

A dual-stage attention-based recurrent neural network built on real-world multivariate time series outperformed three competing ML models for coagulant dosage prediction [6], noting that traditional ML algorithms under-perform on this problem. A DL model trained on large-scale real-time plant data predicted coagulant dosage and sedimentation-basin turbidity, validated against a physicochemical model, and reduced dosage by about 22 % while keeping turbidity under 1.0 NTU [7]. A review of AI in coagulation highlights challenges including process complexity, feedback-signal acquisition and the transfer of models from simulation to real settings [1]. A wider review identifies at least 91 peer-reviewed AI studies in water treatment since 1997: 41 on coagulation/flocculation, 21 on membrane filtration, 16 on adsorption and 13 on disinfection by-products [9].

3. Validation Practice

Environmental time series are autocorrelated, so random train-test splits of neighbouring samples can leak information and inflate accuracy. Splitting by site, season or time block is preferable. The reviewed river studies did not consistently report such splits, so reported accuracies should be read cautiously.

V. IOT INTEGRATION AND EDGE DEPLOYMENT

A typical IoT design joins sensors, a communication network and a cloud back end with a dashboard for prediction results; a Guayas River design follows this pattern and notes that the initial implementation cost may be high while long-term data-management benefits are significant [8]. The Pusu system uses a microcontroller-based device with a GUI [2], [3]. Edge inference lowers latency and dependence on riverbank connectivity, whereas cloud processing simplifies model updates and cross-site analytics.

None of the reviewed river studies reported long-term outdoor operation with biofouling, sensor drift or automated actuation of chemical dosing; the dosing-control evidence comes from treatment plants rather than rivers [1], [6], [7].

VI. COMPARATIVE ANALYSIS

Table I summarises representative studies. Reported values are not directly comparable because targets, datasets and validation differ.

Table 1: Comparison of Representative Studies

Author, year	Objective	Sensing / method	Major reported findings	Limitations
Kabbashi et al., 2024 [2]	Monitor river water quality	IoT sensors; ML	>879 data points; three water classes	Repository record only; details limited
Tahsin, 2025 [3]	Real-time river quality system	Arduino UNO IoT; ML	Up to 98 % classification accuracy	Student thesis (abstract); small scope
Mahanadi study [5]	River quality and seasonal contamination	IoT; XGBoost	Seasonal multi-belt prediction	Quantitative results not verified (abstract)
3D-printed P-sensor study [4]	Phosphorus in river water	3D-printed electrodes; ML	2,000 points; 5 river sites	Single river; preprint
Guayas River design, 2024 [8]	Forecast river quality	IoT sensors, cloud, ML	Network and dashboard design	Design-level; cost noted as high
RNN dosing study, 2023 [6]	Coagulant dosage prediction	Dual-stage attention RNN	Outperformed 3 ML baselines	Treatment plant, not river
Chemosphere study, 2024 [7]	Optimise coagulant dose	DL on large-scale data	About 22 % dose reduction; turbidity < 1.0 NTU	Single plant

VII. PROPOSED CONCEPTUAL ARCHITECTURE

Fig. 1 presents a conceptual architecture derived from the reviewed systems; it is not experimentally validated. Complementary sensors capture physical, electrochemical and optical signals (layer 2) because no single modality covers all contaminants [2], [4]. An embedded unit timestamps and filters signals (layer 3) and edge models perform classification, concentration regression and anomaly detection

with a confidence score (layer 4). Results travel over a low-power network to a cloud database and dashboard (layer 5) [8]. Decision support (layer 6) raises pollution alerts and recommends or sets treatment dosing, with a local alert path that works offline, extending dosing-prediction work [6], [7]. Periodic laboratory samples (layer 7) verify predictions, and models are retrained only after independent re-testing.

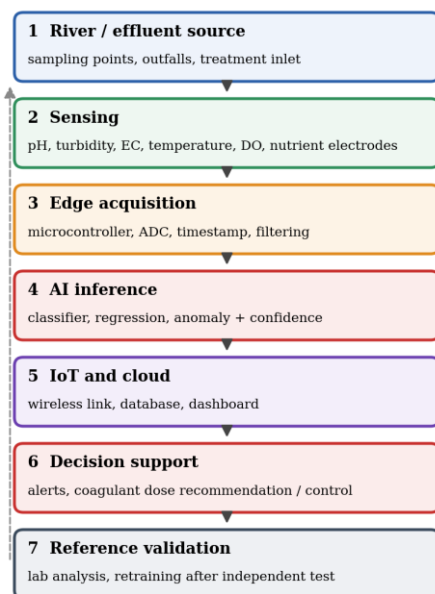


Fig. 1. Proposed conceptual architecture for AI and IoT-based river chemical pollution monitoring and treatment support (conceptual, not experimentally validated)

VIII. RESEARCH GAPS AND FUTURE DIRECTIONS

Gaps supported by the reviewed evidence. (i) Small, single-site datasets and short collection periods [2], [3], [4]. (ii) Few field trials: treatment-dosing results come from plants, not rivers [1], [7]. (iii) Limited real-time actuation and cost evidence for river-side deployment [8]. (iv) Weak validation reporting, with random splits risking leakage. (v) Limited interpretability and uncertainty reporting. (vi) Model adaptability from simulation to real settings remains difficult [1].

Directions requiring investigation. Multi-season, multi-river datasets with laboratory reference labels; field trials with biofouling and drift compensation; multi-sensor fusion; calibrated confidence and out-of-distribution detection at the edge; grouped, site-wise and time-blocked validation; hybrid physics-plus-AI models for dosing [9]; and secure, open data sharing.

IX. CONCLUSION

AI and IoT have moved river chemical monitoring from periodic laboratory sampling toward continuous, low-cost sensing with ML classification and prediction, and AI-based dosing models show clear savings in treatment plants. The evidence for rivers, however, is dominated by small, short-term, single-site studies with limited independent validation. Reliable deployment will require larger and more diverse datasets, field trials, rigorous validation, multi-sensor fusion and continuous reference-based model maintenance, as reflected in the proposed conceptual architecture.

REFERENCES

1. J. Liu, Y. Long, and G. Zhu, "Application of artificial intelligence in the management of coagulation treatment engineering system," *Processes*, vol. 12, no. 9, Art. no. 1824, 2024, doi: 10.3390/pr12091824.
2. N. A. Kabbashi, T. F. Hasan, M. Z. Alam, T. Saleh, and A. Hassan Abdalla Hashim, "Monitoring water quality in Pusu river using Internet of Things (IoT) and Machine Learning (ML)," *IJUM Repository*, 2024. [Online]. Available: <https://irep.iium.edu.my/115923/>
3. T. F. Hasan, "Water quality monitoring of Sungai Pusu using IoT technology and machine learning approaches," Thesis, International Islamic University Malaysia, 2025. [Online]. Available: <https://studentrepo.iium.edu.my/handle/123456789/34088>
4. "Machine learning-enabled river water quality monitoring using lithography-free 3D-printed sensors," *arXiv:2507.14152*, 2025. [Online]. Available: <https://arxiv.org/pdf/2507.14152>
5. "IoT-enabled machine learning for comprehensive water quality assessment in the Mahanadi River: A multibelt analysis of seasonal contamination and predictive modeling," Wiley, doi: 10.1155/je/5549990.
6. "Exploring potential dual-stage attention based recurrent neural network machine learning application for dosage prediction in intelligent municipal management," *Environ. Sci.: Water Res. Technol.*, 2023, doi: 10.1039/D2EW00560C.

7. "Optimizing coagulant dosage using deep learning models with large-scale data," *Chemosphere*, vol. 350, Feb. 2024.
8. "Comparative learning model for water quality forecasting of the Guayas River based on the Internet of Things," in *Applied Human Factors and Ergonomics International*, 2024.
9. "Machine learning (ML) applications in water treatment: Possibilities and advantages," in *Proc. CEST 2025*, Paper cest2025_00235.