

Smart Water Management System Using AI for Waterlogging and Pollution Detection

Nayana Navare¹, Samiksha Nandre², Priya Gore³, Siddhi Thorat⁴, Siddhi Thorat⁵

Students, Department of Computer Engineering¹⁻⁴

Faculty, Department of Computer Engineering⁵

Guru Gobind Singh Polytechnic, Nashik, India

Abstract: Waterlogging during heavy rainfall and the pollution of local water bodies are two environmental problems that are usually studied and monitored separately, even though both require continuous, location-specific data rather than periodic manual checks. This paper surveys five peer-reviewed studies on Internet of Things (IoT) sensing and machine learning (ML) for water-quality monitoring and urban waterlogging prediction, and proposes a Smart Water Management System that integrates both problems within a single low-cost IoT prototype. The system uses ESP32-based sensor nodes to collect rainfall, water-level, pH, turbidity and temperature data, a Django REST API and PostgreSQL database for storage, and Logistic Regression, Decision Tree and Random Forest models to classify waterlogging risk (Low/Medium/High) and pollution risk (Normal/Warning/High), displayed on a web dashboard with a risk map and early-warning alerts. This paper presents the design, related work and implementation plan for the system empirical performance evaluation is reported as part of ongoing and future work.

Keywords: Internet of Things (IoT), Machine Learning, Waterlogging Prediction, Water Pollution Monitoring, Early Warning System, Random Forest

I. INTRODUCTION

During periods of heavy seasonal rainfall, cities commonly experience two connected environmental challenges: water accumulating faster than local drains can clear it, and a decline in the quality of nearby water bodies caused by sewage, industrial waste and surface run-off. Traditional monitoring of both conditions is largely manual and periodic, which limits the ability of residents and local authorities to respond in time.

Advances in low-cost microcontrollers such as the ESP32, combined with machine-learning techniques, have made it possible to build IoT-based systems that sense environmental conditions continuously and classify risk automatically. This paper reviews existing IoT- and ML-based approaches to water-quality monitoring and urban waterlogging prediction, identifies the research gap between them, and proposes a Smart Water Management System that addresses this gap by combining both forms of monitoring within one low-cost platform. e ability of residents and local authorities to respond in time.

II. RELATED WORK

A consistent theme in the water-quality literature is the use of low-cost microcontrollers paired with pH, turbidity and temperature sensors for continuous sensing instead of manual sampling. Murti et al. [1] transmitted pH and turbidity readings over a long-range (LoRa) link to a cloud service with low reported sensor error. Bakar et al. [2] used an ESP32/NodeMCU architecture with Wi-Fi connectivity and reported reliable real-time logging, while noting that sensor calibration still needs improvement. Rahu et al. [5] added a machine-learning stage to classify sensed data into a Water Quality Index and Class, with performance varying across dataset sizes. These studies confirm that low-cost IoT sensing of water-quality parameters is workable, but none consider rainfall or water level.



A separate line of research addresses waterlogging prediction using rainfall and historical depth or exposure data rather than water quality. Zhang et al. [3] compared multiple ML strategies for predicting urban flood depth up to thirty minutes ahead with high reported short-term accuracy. Tong et al. [4] applied CatBoost and related models to thirteen disaster-related factors to map city-scale population exposure to waterlogging. Both studies show that ML improves waterlogging prediction over simple rainfall thresholds, but neither is built around a low-cost IoT sensor node, and neither incorporates water-quality sensing.

Comparing the two directions, the water-quality studies [1], [2], [5] differ mainly in communication design and the presence of an ML stage, while the waterlogging studies [3], [4] differ in spatial scale and purpose. Reported accuracy and reliability are measured inconsistently across all five studies, and none report a combined waterlogging-and-pollution outcome. Table I summarises the problem, method, dataset, key finding, limitation and research gap for each reviewed paper.

Author & Year	Problem	Method	Dataset	Key Finding	Limitation	Research Gap
[1] Murti et al., 2024	Water pollution hard to detect manually	IoT sensing (pH, turbidity) + LoRa + cloud	Live pH/turbidity readings	Low sensor error, fast cloud delivery	Only 2 parameters; no waterlogging	No combined pollution + waterlogging assessment
[2] Bakar et al., 2025	Manual water testing is slow	ESP32/NodeMCU + pH, temp, turbidity + cloud	Real-time prototype readings	Reliable real-time remote logging	Needs better calibration	Pollution + waterlogging both outside scope
[3] Zhang et al., 2022	Predicting flood depth in advance	Multiple ML models on rainfall + depth data	Urban rainfall/depth records	High 30-min-ahead accuracy	Rainfall/depth only, no water quality	No pollution or multi-sensor IoT data
[4] Tong et al., 2025	City-scale population exposure risk	CatBoost + ML on 13 disaster factors	City-wide GIS dataset	Identified high-exposure zones	Not a low-cost field IoT node	Not built for site-level real-time sensing
[5] Rahu et al., 2023	Continuous water-quality analysis needed	IoT framework + ML (WQI/WQC)	Two water-quality datasets	Better results on smaller dataset	Targets general WQI, not location risk	No rainfall/water-level for waterlogging

III. OBJECTIVES AND SCOPE OF THE LITERATURE SURVEY

Objectives:

- Identify existing research on IoT-based water-quality monitoring and machine-learning-based urban waterlogging prediction
- Classify the reviewed literature into meaningful themes rather than presenting isolated paper summaries
- Compare the methods, datasets, findings and limitations reported across the water-quality and waterlogging studies
- Identify a defensible research gap relevant to the proposed Smart Water Management System
- Translate the literature findings into a realistic system design and implementation plan without overstating novelty



Scope:

Limited to peer-reviewed studies published between 2022 and 2025

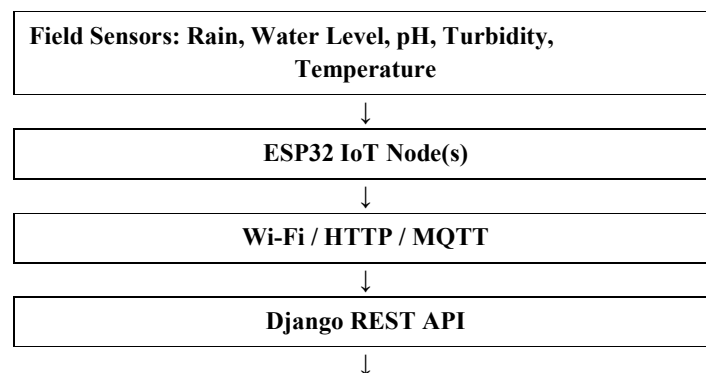
- Covers studies that apply Internet of Things (IoT) sensing, machine learning (ML), or both, to either water-quality assessment or urban waterlogging/flood-risk prediction
- Does not claim that the proposed system replaces municipal flood-warning infrastructure, certified water-testing laboratories, or official disaster-management systems
- Positions the proposed system as a low-cost, field-deployable prototype demonstrating an integrated approach to two environmental risks typically studied separately
- Notes interoperability with larger municipal or smart-city systems as a future direction, not a claim of the current prototype

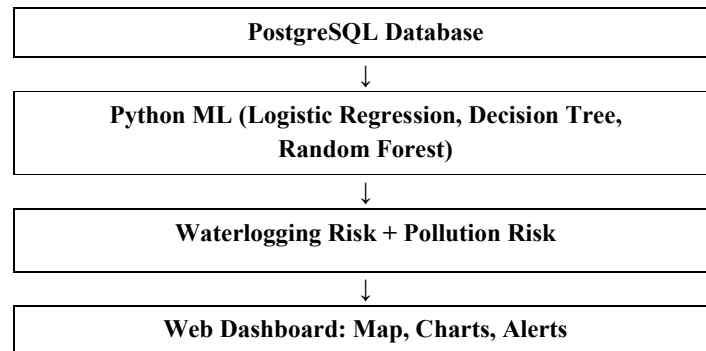
IV. PROPOSER METHODOLOGY/SYSTEM DESIGN

The proposed system follows a layered IoT-to-AI/ML architecture, shown in Fig. 1. Field sensors for rainfall, water level, pH, turbidity and temperature are read by ESP32 IoT node(s), and the readings are sent through a Wi-Fi connection using HTTP or MQTT messaging, reaching a Django-based REST API that writes each reading into a PostgreSQL database. Python-based data processing and machine-learning models then read current and historical data to classify waterlogging risk (Low/Medium/High) and water-pollution risk (Normal/Warning/High). The outcomes appear on a web dashboard that combines a location-based risk map, trend charts built from past readings, and alerts that trigger automatically when risk levels rise (dashboard notification, buzzer/LED, and optional email/SMS).

This architecture follows a layered sense-to-decision pipeline. At the sensing layer, five physical parameters, namely rainfall, water level, pH, turbidity and temperature, are continuously measured at the demonstration site, forming the raw environmental signal that drives both risk models. These readings are acquired by one or more ESP32 microcontroller nodes, chosen for their built-in Wi-Fi capability and low cost, which makes field-level deployment economically feasible for a prototype system. Each node packages the sensor readings and transmits them using HTTP or MQTT, both lightweight protocols suited to frequent, small-payload transmissions from resource-constrained devices.

On the server side, a Django REST API receives and validates incoming readings before persisting them in a PostgreSQL database, which maintains both the live feed and the historical record needed for trend analysis. This stored data is then consumed by a Python-based processing layer, where three machine-learning models — Logistic Regression, Decision Tree and Random Forest — are applied to classify the current environmental state into a waterlogging-risk category (Low/Medium/High) and a water-pollution-risk category (Normal/Warning/High). Separating the sensing, storage and inference layers in this way allows each component to be developed, tested and scaled independently, and allows the ML models to be retrained on accumulated historical data without disrupting data collection.





The classification output is finally rendered on a web dashboard, which brings together a location-based risk map, time-series charts of historical readings, and an early-warning alert mechanism. When either risk category reaches a high-severity level, the system triggers a layered response — an on-dashboard alert, a local buzzer/LED indicator at the sensor node, and an optional email/SMS notification — ensuring that the warning reaches both the monitoring dashboard and anyone physically near the site

V. SENSOR CALIBRATION AND SYSTEM RELIABILITY CONSIDERATIONS

Murti et al. [1] report a low percentage error for pH and turbidity sensors transmitted over a LoRa-based prototype, while Bakar et al. [2] demonstrate reliable real-time logging on an ESP32/NodeMCU architecture but explicitly identify sensor calibration as an open problem that affects the accuracy of low-cost water-quality sensors over time. Rahu et al. [5] further show that the performance of machine-learning classification on sensed data varies depending on the size of the dataset used for training, indicating that prediction reliability is sensitive to both sensor accuracy and the volume of historical data available.

For SmartWaterGuard, the important design implication is that low-cost pH, turbidity and water-level sensors should be periodically checked against reference readings, and that raw sensor values should be validated before being passed to the ML models rather than assumed to be accurate by default. Nevertheless, none of the cited studies validate long-term calibration drift or sensor reliability under continuous multi-season field deployment of the kind SmartWaterGuard proposes; that validation remains part of SmartWaterGuard's future testing rather than an established result in the reviewed literature.

VI. DASHBOARD, ALERTING AND EARLY-WARNING DESIGN CONSIDERATIONS

Across the reviewed literature, result delivery is handled in different and largely single-purpose ways: Murti et al. [1] and Bakar et al. [2] deliver readings to a cloud service or remote log without a described multi-channel alert mechanism, while Zhang et al. [3] and Tong et al. [4] output a predicted depth or exposure-risk value intended for planning use rather than for a real-time dashboard aimed at a local user. Rahu et al. [5] classify water quality into an index/class but do not describe how this classification is communicated to an end user.

For SmartWaterGuard, the design implication is that risk outputs should be presented through more than one channel, namely a dashboard view, a local buzzer/LED at the sensor node, and an optional email/SMS alert, so that a warning reaches both a remote viewer and anyone physically near the site. However, the cited studies do not validate a unified alerting interface that reports both waterlogging and pollution risk together; that combined-risk alerting design remains part of SmartWaterGuard's proposed integration rather than a feature demonstrated in prior work.



VII. IMPLEMENTATION

The hardware layer consists of an ESP32 development board, a rain sensor/rain gauge, a waterproof ultrasonic water-level sensor (JSN-SR04T), a pH sensor probe, a turbidity sensor, a DS18B20 waterproof temperature sensor, a buzzer and LED for local alerts, and an IP65 enclosure for outdoor protection, assembled on a breadboard with a regulated 5V/12V power supply.

The software layer uses Arduino IDE/PlatformIO for ESP32 firmware, Python with Django and Django REST Framework for the backend API, PostgreSQL for data storage, and Pandas, NumPy and scikit-learn for data processing and model training. Chart.js and Leaflet.js are used for dashboard charts and the risk map, and Git/GitHub is used for version control. The system is implemented in two stages: first, sensor data is transmitted to the Django API over HTTP and stored directly; second, the trained ML models are integrated for live inference on incoming readings.

VIII. RESULTS AND DISCUSSION

As this paper is submitted at the design and implementation stage of the project, no field-trial accuracy figures are reported here. Based on the system design, the expected outcome is a working ESP32-based sensor node transmitting live readings to the Django API, trained Logistic Regression, Decision Tree and Random Forest models that classify waterlogging and pollution risk, and a dashboard displaying live readings, historical trends, a risk map and alerts.

Quantitative evaluation, including classification accuracy, precision, recall and confusion matrices for each ML model, and tests of sensor and network failure handling, will be carried out once the prototype and dataset are complete, and will be reported in a revised version of this paper.

IX. CONCLUSION

This paper reviewed five peer-reviewed studies on IoT-based water-quality monitoring and ML-based waterlogging prediction, and found that these two research directions have not been combined in a single low-cost, field-deployable system. The proposed Smart Water Management System addresses this gap by integrating rainfall, water-level, pH, turbidity and temperature sensing with ML-based risk classification for both waterlogging and pollution, presented through one dashboard with early-warning alerts.

Future work includes completing field testing of the prototype, addressing sensor calibration as noted by Bakar et al. [2], extending the system to multiple sensor nodes with a TDS/EC sensor, evaluating MQTT or LoRa communication for sites with limited Wi-Fi coverage, and training the ML models on a larger, multi-season dataset to report quantitative accuracy results.

X. ACKNOWLEDGMENT

The authors thank Ms. Siddhi Thorat, Department of Computer Engineering, Guru Gobind Singh Polytechnic, Nashik, for guidance on this project.

REFERENCES

- [1] M. A. Murti, A. R. A. Saputra, I. Alinursafa, A. N. Ahmed, A. Yafouz, and A. El-Shafie, "Smart system for water quality monitoring utilizing long-range-based Internet of Things," *Appl. Water Sci.*, vol. 14, no. 4, p. 69, 2024. [doi: 10.1007/s13201-024-02128-z](https://doi.org/10.1007/s13201-024-02128-z)
- [2] A. A. A. Bakar, Z. A. Bakar, Z. M. Yusoff, M. J. M. Ibrahim, N. A. Mokhtar, and S. N. Zaiton, "IoT-based real-time water quality monitoring and sensor calibration for enhanced accuracy and reliability," *Int. J. Interact. Mob. Technol.*, vol. 19, no. 1, 2025. [doi: 10.3991/ijim.v19i01.51101](https://doi.org/10.3991/ijim.v19i01.51101)
- [3] Z. Zhang, J. Liang, Y. Zhou, Z. Huang, J. Jiang, J. Liu, and L. Yang, "A multi-strategy-mode waterlogging-prediction framework for urban flood depth," *Nat. Hazards Earth Syst. Sci.*, vol. 22, no. 12, pp. 4139-4165, 2022. [doi: 10.5194/nhess-22-4139-2022](https://doi.org/10.5194/nhess-22-4139-2022)



- [4] S. Tong, J. Wang, J. Qin, X. Ji, and Z. Wu, "Study on the risk of urban population exposure to waterlogging in Huang-Huai area based on machine learning simulation analysis: A case study of Xuzhou urban area," *Land*, vol. 14, no. 5, p. 939, 2025. [doi: 10.3390/land14050939](https://doi.org/10.3390/land14050939)
- [5] M. A. Rahu, A. F. Chandio, K. Aurangzeb, S. Karim, M. Alhussein, and M. S. Anwar, "Toward design of Internet of Things and machine learning-enabled frameworks for analysis and prediction of water quality," *IEEE Access*, vol. 11, pp. 101055-101086, 2023. [doi: 10.1109/ACCESS.2023.3315649](https://doi.org/10.1109/ACCESS.2023.3315649)

