

IoT-Based Waterborne Disease Monitoring System: A Review of Water-Quality Sensing, Machine Learning, and Community Early Warning

Nawale Atharva Bharat, Prof. Bangar Abhishek Popat,
Prof. Bhosale.Sachin Bajirao, Prof. Karad Vandana

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

Abstract — Waterborne diseases remain a public-health concern where drinking-water sources may be contaminated by pathogenic microorganisms, sewage, agricultural runoff, or inadequate sanitation. Conventional laboratory testing provides important reference evidence, but sampling, transport, and analysis can delay the identification of unsafe water. Internet of Things (IoT) devices can support more frequent observation of water-quality indicators, while machine learning (ML), cloud or edge computing, and community health reporting can help convert observations into timely alerts. This paper presents a narrative review of IoT-enabled water-quality monitoring and waterborne-disease early-warning approaches. It examines common sensing parameters—including pH, temperature, turbidity, electrical conductivity and total dissolved solids—along with data transmission, anomaly detection, predictive analytics, and public-health notification. The review compares representative approaches, identifies gaps in field validation, sensor calibration, pathogen-specific testing, data quality, and alert governance, and proposes a conceptual architecture that combines sensor stations, laboratory confirmation, risk analytics, and community health surveillance. The proposed architecture is a review-derived design, not an experimentally validated system.

Keywords— Waterborne Disease, Internet of Things, Water Quality Monitoring, Machine Learning, Early Warning, Public Health

I. INTRODUCTION

Safe drinking water is fundamental to public health, yet water sources can be affected by bacteria, viruses, protozoa, sewage discharge, agricultural runoff, and inadequate treatment or storage. Exposure to contaminated water can contribute to illnesses such as diarrhoeal disease, cholera, typhoid, and other infections. The level of risk depends on the contaminant, the amount of exposure, local sanitation conditions, and the effectiveness of water treatment and distribution.

Conventional water-quality assessment relies on collecting samples and analysing them using appropriate laboratory procedures. These methods remain necessary, especially when microbiological confirmation is required. However, periodic sampling provides observations only at selected times and locations. A contamination event occurring between sampling rounds may not be recognised immediately. Transport, laboratory capacity, and communication of results can add further delay.

Internet of Things technology provides an opportunity to observe selected water-quality indicators at more frequent intervals. Sensors installed at a source, treatment plant, storage

tank, or distribution point can send timestamped measurements to a gateway or server. Statistical rules and ML models can then identify unusual readings, estimate selected water-quality variables, and prioritise locations for further investigation. Community health workers and health facilities may contribute aggregated information about symptoms, adding context to environmental measurements.

A key distinction must be maintained between monitoring water quality and diagnosing waterborne disease. A sensor reading may indicate a change that deserves attention, but it does not necessarily reveal which contaminant caused the change. Similarly, an increase in reported gastrointestinal symptoms is a signal for investigation, not proof that a particular water source caused an outbreak. This review therefore considers technical approaches together with validation, uncertainty, privacy, and the response process.

II. REVIEW METHODOLOGY

This paper presents a focused narrative review based on the supplied research materials and the studies cited within them. The reviewed themes include IoT and ML-enabled water-quality analysis, automated prediction of water-related disease,

hybrid deep-learning approaches for time-series water-quality prediction, community health monitoring, and model-based approaches to waterborne-disease control. Each source is considered in terms of its objective, input data, analytical method, reported findings where available, and limitations relevant to real-world deployment.

This is not a systematic review or meta-analysis. It does not claim an independent database-wide search, formal duplicate screening, or pooled statistical estimate. Numerical values are reported as described in the supplied sources and must be interpreted in the context of each study's dataset, prediction target, data split, and evaluation method. In particular, reported model accuracy is not assumed to represent pathogen-detection accuracy or performance in every community.

III. SYSTEM ARCHITECTURE DIAGRAM



IV. BACKGROUND AND TECHNOLOGY OVERVIEW

1. Water-Quality Sensing

Common low-cost sensing parameters include temperature, pH, turbidity, electrical conductivity (EC), and total dissolved solids (TDS). These variables can reveal changes in water

conditions and may help identify anomalies that warrant investigation. Additional sensors may be selected according to the local source and risk profile. Sensor choice must be matched to the intended use, and calibration, fouling, temperature effects, and maintenance requirements should be considered.

2. IoT Communication and Data Management

A typical station combines sensors, a microcontroller, timestamping, local storage, and a communication module. Data may travel through Wi-Fi, cellular networks, LPWAN technologies, or a local gateway depending on coverage and energy constraints. A server or cloud platform stores time-series observations, device status, calibration records, and alert history. Local buffering is useful where connectivity is intermittent.

3. Analytics and Early Warning

Rule-based thresholds are simple to interpret and can be used as a first line of screening. ML models can learn relationships among measurements, identify unusual patterns, or estimate selected water-quality variables. Time-series models may support trend forecasting, while classification models can prioritise samples for further testing. Alerts should include confidence or severity, the affected location, the reason for the alert, and a recommended verification step.

4. Public-Health Information

Water-quality observations can be considered alongside anonymised or aggregated reports of symptoms from community health workers or health facilities. A rise in gastrointestinal symptoms near a water source may justify investigation, but symptoms are not specific to waterborne disease. Public-health teams should confirm suspected events using established epidemiological and laboratory procedures.

V. LITERATURE REVIEW

1. IoT and Machine Learning for Water-Quality Analysis

The study “Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality” describes a framework that combines IoT measurements with ML. The supplied paper reports temperature, pH, turbidity, and TDS as sensor inputs and reports an MLP regression R^2 of 0.93 and Random Forest classification accuracy of 0.91 for its evaluated tasks. These figures demonstrate feasibility within that study, but they should not be treated as universal performance estimates or as direct proof of pathogen detection.

2. IoT-Based Automated Water-Related Disease Prediction

The supplied 2024 Scientific Reports article, “IoT-based automated system for water-related disease prediction,” explores water-quality data and predictive methods. The source reports Random Forest accuracy of 99.66%, XGBoost accuracy of 99.52%, and AdaBoost accuracy of 99.64% for the study's evaluated prediction task; it also describes time-series modelling and methods for addressing class imbalance and reducing features. Such high scores require careful interpretation against the dataset construction, labels, split strategy, and external validation. They do not establish that a sensor-only system can diagnose individual patients.

3. Hybrid Deep Learning for Water-Quality Prediction

The supplied paper “Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models” examines hybrid models including CNN-LSTM and CNN-GRU. The paper reports that CNN-LSTM performed best among the evaluated models and datasets. Aquaculture prediction is relevant to time-series modelling and sensor analytics, but its objectives and environmental conditions may differ from drinking-water safety and human disease surveillance.

4. Community Health Monitoring and Early Warning

The supplied community-monitoring material describes combining IoT observations with community-level symptom reporting and early-warning workflows. This reflects an important system-level principle: environmental monitoring may become more useful when interpreted with local public-health observations. However, a practical deployment needs clearly defined reporting protocols, privacy protections, responsible alert review, and confirmation by public-health authorities.

5. Biodynamic and Model-Based Approaches

The supplied material on a biodynamic model for waterborne-disease control links a mathematical model with IoT-network information and public-health interventions. Model-based approaches can help explore how interventions and changing conditions affect risk. Their practical value depends on the quality of input data, the assumptions of the model, and validation against observed local outcomes.

6. Synthesis

Across the reviewed materials, three capabilities recur: frequent sensing, predictive analysis, and a response pathway. The principal distinction is between estimating water-quality conditions and predicting disease-related risk. A robust system should preserve this distinction and use laboratory or

epidemiological confirmation when the alert has public-health consequences.

VI. PROPOSED CONCEPTUAL ARCHITECTURE

The following architecture is derived from the reviewed approaches and is proposed for future implementation; it has not been experimentally validated.

1. Layer 1—Water-Source Monitoring

Install monitoring nodes at selected source, storage, treatment, or distribution points. Collect parameters appropriate to the site, such as pH, temperature, turbidity, EC, and TDS. Each record should include timestamp, sensor identifier, location, calibration status, and device-health information.

2. Layer 2—Preprocessing and Quality Checks

Check missing values, impossible readings, duplicate timestamps, sensor drift, and abrupt changes. Retain raw readings alongside cleaned data so that an alert can be audited. If a sensor is outside its calibration or maintenance period, flag the data rather than silently treating it as reliable.

3. Layer 3—Risk Analytics

Use transparent threshold rules as an initial screen and compare them with validated ML models for anomaly detection or water-quality prediction. Models should be selected according to the target and available labelled data. A risk score should not be presented as a confirmed pathogen result unless a pathogen-specific test supports that conclusion.

4. Layer 4—Communication and Dashboard

Transmit readings and alerts to a dashboard using a suitable network and protocol. Display trends, location, sensor status, alert rationale, and verification status. Provide offline buffering and a local warning path for locations with unreliable connectivity.

5. Layer 5—Laboratory and Public-Health Verification

When readings cross defined limits or the analytics identify a persistent anomaly, notify the designated operator and collect a confirmatory sample according to local procedures. Where disease signals are also reported, public-health teams should review the combined evidence. Close the alert only after documenting the investigation and corrective action.

6. Evaluation Plan

Evaluation should include sensor calibration against reference instruments, laboratory testing where relevant, sensitivity and

specificity for defined alert labels, false-alarm rates, time-to-alert, missing-data performance, network uptime, power consumption, and performance across seasons and locations. Validation should separate sites or time periods where possible to reduce data leakage.

VII. RESEARCH GAPS AND FUTURE DIRECTIONS

The reviewed material suggests several priorities. First, sensor readings are indirect indicators and may not identify pathogens; deployments need confirmatory microbiological testing. Second, models require diverse, locally representative datasets and external validation across water sources, seasons, and operating conditions. Third, sensor drift, fouling, calibration, power, and network interruptions can reduce reliability outside controlled environments. Fourth, the meaning of “disease prediction” must be defined precisely, including how outcome labels were created and whether they represent water contamination, community-level incidence, or individual illness.

Further work should also report uncertainty, false positives and false negatives, model explainability, and the cost of field maintenance. Alert protocols should define who receives an alert, who verifies it, and which actions are authorised. Community reporting should minimise personal data, use aggregation where possible, and avoid stigmatizing locations or populations on the basis of unconfirmed signals.

Limitations and Ethical Considerations

This paper is a narrative review based on the supplied materials rather than a systematic evidence synthesis. Some papers report strong performance on their own evaluated datasets, but those values cannot be assumed to transfer to other regions or operational deployments. The proposed architecture has not been built or tested as part of this review.

A monitoring dashboard must not replace certified laboratory tests, clinical assessment, or official public-health guidance. False alarms can cause unnecessary concern, while missed alerts can create a false sense of safety. Systems should use secure communication, access control, data minimisation, transparent alert rules, and documented human review.

VIII. CONCLUSION

IoT-enabled water-quality monitoring combined with ML and public-health information offers a promising way to identify

unusual water conditions and prioritise timely investigation. The reviewed studies demonstrate technical approaches for sensor-based measurement, prediction, and early warning, but performance figures remain specific to their datasets and tasks. Most importantly, routine water-quality indicators are not equivalent to pathogen detection or disease diagnosis. A credible system should combine calibrated sensors, robust data handling, locally validated analytics, laboratory confirmation, and a clear response workflow. Future deployments should prioritise field validation, uncertainty reporting, maintenance, privacy, and collaboration with water and public-health authorities.

Study / source	Focus	Reported method or finding	Key caution
IoT + ML water-quality framework (2023)	Water-quality analysis/prediction	MLP regression $R^2 = 0.93$; RF classification accuracy = 0.91 in evaluated tasks	Task-specific metrics; not direct pathogen confirmation
IoT-based automated water-related disease prediction (2024)	Disease-related prediction	RF 99.66%; XGBoost 99.52%; AdaBoost 99.64% reported by source	Requires scrutiny of labels, splits, and external validation
Smart aquaculture water-quality prediction (2022)	Time-series prediction	CNN-LSTM reported as best among evaluated hybrid models	Aquaculture setting may not transfer to drinking-water risk
Community health monitoring material (2026)	Community early warning	Combines environmental monitoring with symptom/health-worker reporting	Needs privacy safeguards and official verification
Biodynamic-model material	Disease-control modelling	Links mathematical modelling, IoT information, and interventions	Depends on assumptions and local outcome validation

REFERENCES

1. “Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality,” IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3315649.

2. “Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models,” IEEE Access, 2022, doi: 10.1109/ACCESS.2022.3180482.
3. “IoT-based automated system for water-related disease prediction,” Scientific Reports, vol. 14, Art. no. 29483, 2024, doi: 10.1038/s41598-024-79989-6.
4. “SMART COMMUNITY HEALTH MONITORING AND EARLY WARNING” [title as shown in the supplied document], Journal of Advanced Multidisciplinary Research Studies and Development, vol. 5, special issue 2(B), 2026.
5. “Smart Waterborne Disease Control for a Scalable Population using Biodynamic Model in IoT Network,” bibliographic details to be completed from the original source before submission.