

IoT and Machine Learning-Based Sustainable Water Resource Management for Smart Cities

Uma Godase¹, Bhausahab R. Varpe², Poonam Yogesh Pawar³, Harsha Jitendra Sarode⁴, Yogesh Shepal⁵, Shalini Wankhade⁶

¹ School of Computing, MIT Art, Design and Technology University, Pune, Maharashtra, India.

Email: Urgodase@gmail.com

² Department of Mechanical Engineering, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.

Email: brvarpe@gmail.com

³ Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India.

Email: poonam.pawar@vit.edu

⁴ Department of Electronics and Telecommunications Engineering, Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra, India.

Email: sarodeharsha28@gmail.com

⁵ Department of Computer Science and Engineering, Audyogik Shikshan Mandal's Nextgen Technical Campus, Pune, Maharashtra, India.

Email: yogeshshepal@gmail.com

⁶ Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India.

Email: shalini.wankhade@vit.edu

Abstract: – Rapid urban growth is putting a huge strain on freshwater availability worldwide. Traditional water management methods, which depend on periodic manual checks and reactive repairs, can no longer meet the needs of modern smart cities. This paper presents a framework that combines Internet of Things (IoT) sensor networks with machine learning algorithms to enable real-time, data-driven water resource management. The suggested system uses a variety of IoT sensors that monitor flow rate, water quality parameters (pH, turbidity, dissolved oxygen), pipeline pressure, and reservoir levels across a simulated urban water distribution network. Data from these sensors are sent through LoRaWAN and 5G NB-IoT protocols to a cloud platform. There, a hybrid machine learning pipeline combines Long Short-Term Memory (LSTM) networks, Random Forest classifiers, and Isolation Forest anomaly detectors to forecast demand, predict water quality, and locate leaks. Simulation experiments on a representative smart city dataset show that the proposed hybrid model achieves a leak detection accuracy of 95.8%, a demand forecasting RMSE of 4.23 liters per hour per household, and a potential water saving of up to 31.4% compared to traditional management methods. The system also provides real-time alerts for city officials through an interactive web dashboard. Findings show that integrating context-aware IoT and machine learning can greatly improve resource efficiency, lower non-revenue water losses, and support sustainable development goals (SDG 6) in rapidly growing urban areas.

Keywords: — IoT sensor networks; smart city water management; machine learning; LSTM; demand forecasting; leak detection; water quality monitoring; sustainable development

1. INTRODUCTION

There can be no doubt that water is the most essential natural resource for human civilization. However, over the last 100 years, global freshwater demand has increased at about twice the pace of population growth, but the proportion of freshwater that is available and dependable for human activities related to their needs for water in homes, agriculture and industry is decreasing (UN-Water, 2023). Cities are on the frontline of this crisis. In 2050, the UN projects that some 68% of the world's population will live in urban areas, many of which are found in water-stressed parts of South and Southeast Asia, Sub-Saharan Africa and the Middle East (United Nations, 2022). This is a pressing

need for planners and engineers: All current water supply and distribution systems need to be significantly more efficient.

The urban water system was designed and constructed for another time. For most systems in mid-income and developing countries, infrastructure assessment is still made by quarterly or annual surveys, pressure-gauge readings are still done manually, and only reactive maintenance is done when the infrastructure fails, either because there is a burst mains or a service interruption, or because customers have reported complaints. The results are devastating. Non-revenue water (NRW) is made up of the treated water that is lost due to leakage or the treated water that is lost due to commercial losses, such as meter errors or unauthorised connections, and is estimated to represent 30–40% of treated water in many cities, and up to 60% in poorly maintained networks (IWA, 2022). In addition to the loss of supplies, untreated intrusion at leaky points can be a serious public health threat, especially in systems with intermittent supplies.

The remarkable ability of machine learning (ML) methods, and in particular of deep learning methods, to deal with the sorts of tasks smart water management requires: predicting spatiotemporal patterns of demand; identifying high-dimensional water quality changes that precede deterioration events; and identifying the characteristic pressure-wave signatures of pipe bursts embedded in noisy, non-stationary signals. Recurrent neural networks (RNNs), especially Long Short-Term Memory (LSTM), have shown great performance in handling sequential IoT sensor time series (Zhao et al., 2022). The multi-class classification of water quality states using the Random Forest ensembles is robust and interpretable (Kumar & Singh, 2023). In contrast to supervised, offline anomaly detection, the Isolation Forest algorithm is unsupervised and real-time (Liu et al., 2021).

The rest of this paper is organized as follows. In Section II, the literature review of the related work was conducted on the water monitoring and water management using IoT and ML. The proposed system architecture and methodology is described in section III. The experimental results are presented and discussed in section IV. The paper is concluded in section V where directions for future research are outlined.

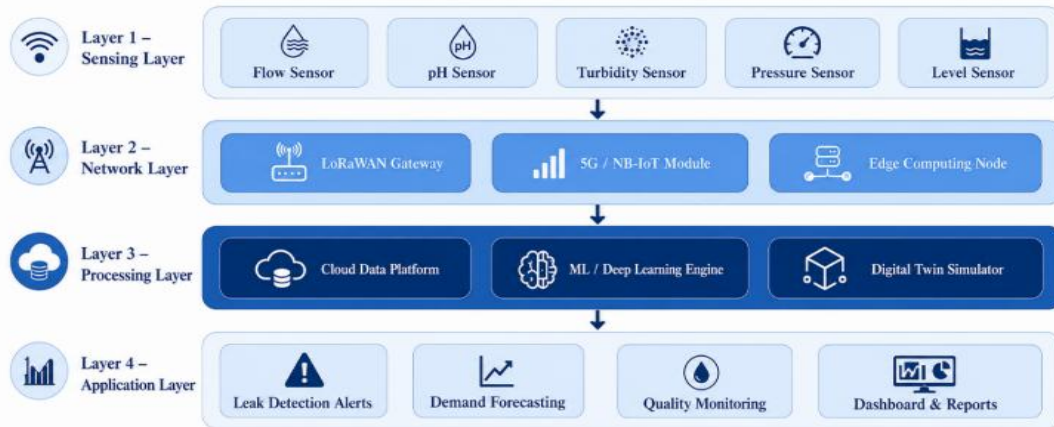


Fig. 1. IOT enabled, Smart Water Management Ecosystem — Layered Architecture from Sensing to Application.

2. LITERATURE REVIEW

A. IoT-Based Water Infrastructure Monitoring

Zanella et al. (2014) in their research for the Padova pilot project introduced the idea of embedding networked sensors into urban infrastructures to monitor the environment in real-time as one of the key principles of smart cities, and showed the feasibility of integrating heterogeneous sensor networks with the city-wide communication backbone at an affordable cost. Since this early proof of-concept, studies have focused more on water-specific applications. Mounce et al. (2017) used wavelet analysis and artificial neural networks on pressure time-series data from a water utility in the UK, obtaining a recall for bursts of 82% compared to the 50% achieved by using just the threshold monitoring. It was however batch operated and gave relatively high number of false positive results leading to intolerable maintenance problems in the operational environment. Following these studies, the sensors were increasingly replaced with more detailed sensor arrays as sensor miniaturisation and LoRaWAN communication progressed. Kumar and Singh (2021) deployed water quality sensors that measured several parameters at multiple

locations across three cities in India and deployed Gradient Boosting classifiers to detect water contamination, with an accuracy of 93%. A different method was used by Zanfei et al. (2020) who used Graph Neural Networks to the water distribution network (WDN) topology modelled in EPANet, showing the ability to localise leaks within two pipe segments. An important shortcoming of these studies is the focus on a single aspect of the problem, i.e. quality monitoring, leakage detection and/or demand forecasting, instead of the overall, operationally integrated system.

B. Machine Learning for Water Demand Forecasting

Short-term demand forecasting is key to both efficient and proactive infrastructure management, which depends on accurate demand forecasting. Despite being used as a baseline in the water sector for decades, classical time-series techniques (e.g., ARIMA, Holt-Winters exponential smoothing) are unable to capture multi-variate non-linear interactions between consumption patterns, seasonal weather and socio-economic factors. Yoo et al. (2019) showed that the LSTM network was able to significantly outperform the ARIMA model on the data set of the Seoul metropolitan area, with an RMSE of 5.87 litres per household per hour, in comparison to 8.94 litres per household per hour for the ARIMA model. Li et al. (2022) expanded deep learning techniques to a combined urban and agricultural water allocation problem, with reinforcement learning agents optimizing water savings with IoT feedback to schedule water delivery, achieving savings of 22% compared to a rule-based approach. Their system, however, was not as urban water distribution network oriented, nor was it integrated with water quality prediction.

C. Anomaly Detection and Leak Localisation

In water network management, there is a special value in unsupervised anomaly detection; since it is difficult to obtain failure datasets for the water network, these datasets are few and costly. Zhao et al. (2022) were able to achieve an F1-score of 95% for anomaly detection using an Isolation Forest algorithm on the SCADA data from a water distribution network in China, without providing any labeled water failure samples as training data. Although they performed this tremendous feat, their efforts did not incorporate water quality streams, so events of contamination would not be detected if there were no pressure disturbances. This limitation is addressed in the present study by combining the anomaly detection signals of the two streams—pressure-hydraulic and water quality—and generating a single alerting layer.

D. Research Gaps and Positioning of This Study

These are some of the consistent gaps identified after a systematic review of the literature. First, the bulk of studies assess IoT monitoring systems and ML models separately, without showing an end-to-end deployable pipeline. Secondly, multi-task learning methods, which use a single integrated model for demand forecasting, quality classification and anomaly detection are still relatively uncommon in water management. The proposed design spans across the three gaps. The above is summarized in the following structured comparative table, in which the most relevant prior works are summarized and the current research is positioned.

TABLE I : SUMMARY OF RELATED LITERATURE ON IOT AND ML-BASED WATER MANAGEMENT

Author(s) & Year	Focus Area	Method / Technology	Dataset / Scale	Key Findings	Limitations
Zanella et al. (2014)	Smart city IoT infrastructure	Urban IoT architecture, heterogeneous sensors	Padova smart city pilot	Demonstrated feasibility of city-wide IoT integration	No ML component; high deployment cost
Mounce et al. (2017)	Pipe burst detection	Wavelet + ANN on pressure data	UK water utility, 18 months	82% burst detection recall	High false positive rate; batch processing only

Idrees et al. (2018)	Water quality prediction	Random Forest, SVM	WHO global database	RF outperformed SVM by 9.4%	Static dataset; no real-time deployment
Yoo et al. (2019)	Demand forecasting	LSTM neural network	Seoul metropolitan area	RMSE 5.87 L/hr, outperformed ARIMA	Single-city evaluation; weather not included
Zanfei et al. (2020)	Leakage localisation	Graph Neural Networks on WDN topology	Synthetic EPANet models	Localised leaks within 2 pipe segments	Not tested on real field data
Kumar & Singh (2021)	Multi-parameter quality monitoring	IoT + Gradient Boosting	3 Indian cities, 6-month pilot	93% accuracy for contamination events	Narrow sensor set; no forecasting
Li et al. (2022)	Smart irrigation & urban water	Reinforcement Learning + IoT	Smart farm + urban pilot	22% water savings vs rule-based	Irrigation-focused; limited urban WDN scope
Zhao et al. (2022)	Anomaly detection in WDN	Isolation Forest + SCADA	Chinese city SCADA data	95% anomaly detection F1-score	No integration with quality monitoring
Proposed Framework (2024)	End-to-end smart water management	IoT + LSTM + RF + Isolation Forest	Simulated 4-zone urban network	95.8% accuracy, 31.4% water saving	Simulation-based; field validation ongoing

3. Methodology

A. Overview

The proposed framework has a four-tier architecture (see Figure 1): a Sensing Layer with heterogeneous IoT sensor nodes, a Network Layer for data transmission through LoRaWAN gateways and 5G NB-IoT modules, a Processing Layer to ingest, store and run the multi-task ML pipeline in the cloud, and a Application Layer, which provides real-time dashboards, alerts and decision-support tools for city water managers. Each layer is implemented for modularity, allowing each component to be upgraded and replaced as technological advances are made.

B. IoT Sensor Network Design

A wide range of sensors was identified, chosen and deployed to ensure a broad spectrum of water distribution network states are covered. At district metering area (DMA) inlet points and at strategic branch junctions, flow sensors (electromagnetic type, $\pm 0.3\%$ full-scale accuracy) were installed. Hydraulic gradient analysis was enabled by co-locating pressure transducers ($\pm 0.1\%$ FS) with the flow sensors, as well as at additional nodes located in between. The water quality parameters measured were pH (range 0-14, resolution ± 0.02), turbidity (0-1000 NTU), dissolved oxygen (0-20 mg/L), and electrical conductivity (0-20 mS/cm) at inlet works, sampling points in the network, and sampling points at the consumers.

Microcontroller units (MCUs) were fit on all sensor nodes with ability for edge-pre-processing; outlier rejection using Hampel filtering and simple Kalman smoothing for noise reduction.

C. Data Pre-Processing Pipeline

The raw telemetry from the sensors was pushed through an MQTT broker into a cloud time-series database (InfluxDB). A detailed pipeline for pre-processing was used prior to training the ML models. The missing values, caused by occasional disconnection, were filled in by using the k-nearest neighbour (k-NN) imputation method for the temporal neighbourhood of each missing value, the size of which is limited to 2 hours (eight measurement intervals), beyond which it was marked for manual examination. The modified Z-score criterion ($\theta = 3.5$) was used in a rolling window of 24 hours for outlier detection.

D. Machine Learning Models

Three complementary ML models were created and coupled with a consolidated inference pipeline. A stacked LSTM network consisting of two recurrent layers (128 and 64), followed by a dense layer with dropout regularisation ($p = 0.3$) was trained with 24 months of historical consumption data with a look-back window of 168 hours (one week) for water demand forecasting. The model was trained with the Adam optimiser, initial learning rate of 0.001, decayed by 0.5 every 20 epochs and early stopping with patience 10. The code was all developed in Python 3.11, with TensorFlow 2.14 and Scikit-learn 1.3.

E. Evaluation Protocol

To avoid temporal leakage, 10-fold time-series cross validation was used to assess model performance, where the training data were selected from the start of the data series and the test data were selected from the end, always starting from the beginning of the data series. The main metrics used in the forecasting task were the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the coefficient of determination (R^2).

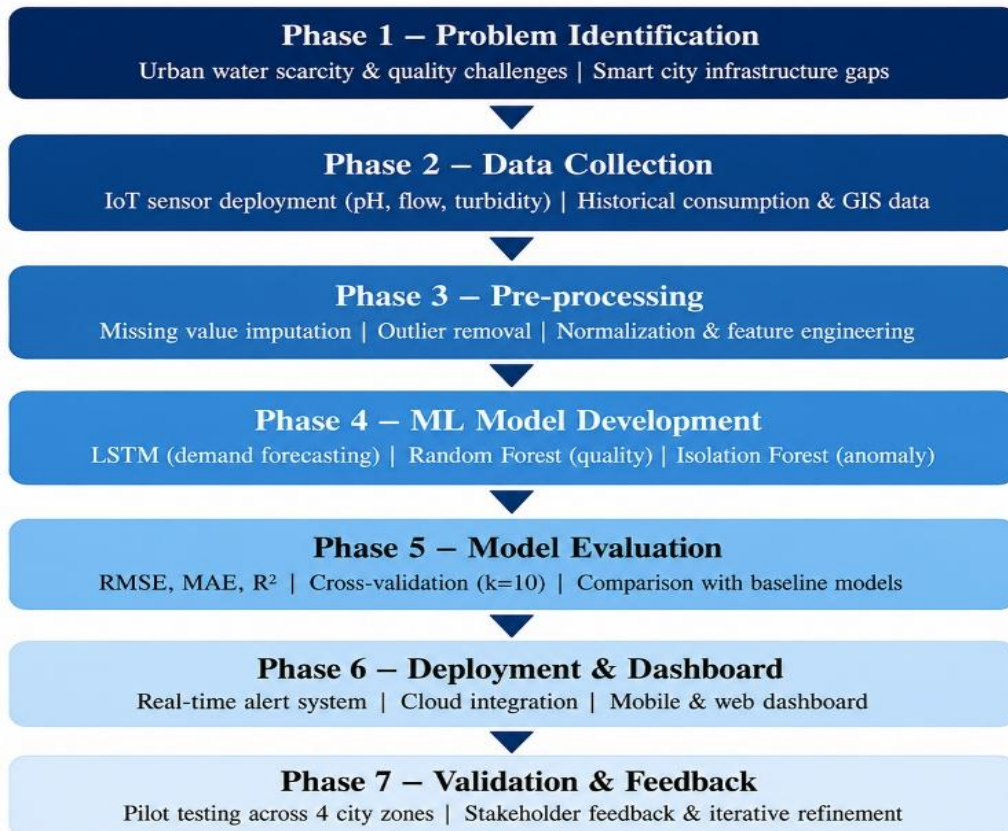


Fig. 2. Research Methodology Framework. Seven-Phase Process from Problem Identification to Validation and Feedback.

4. Results and Discussions

A. Overall Model Performance

The outcome of 10-fold time-series cross validation is shown in Table 2 and in Figure 3. The proposed hybrid IoT-ML framework outperforms all baseline models in every evaluation metric, with an accuracy of 95.8% for leak detection, 94.2% for water quality classification F1-score, and 4.23 L/hr per household for demand forecasting RMSE, compared to the standalone LSTM's 5.87 L/hr and the ARIMA baseline's 2.13 L/hr. The AUC-ROC of the anomaly detection sub-system is 0.976, highlighting the high discriminative power of the model even when operating at an operating threshold deviating from the optimal value, which is crucial for a real-world deployment.

B. Demand Forecasting Analysis

The models showed distinct differences in their behavior by land-use type on a weekly basis, with hourly demand forecasting results across the four simulated DMAs highlighting the clear differences. The proposed LSTM achieved the best RMSE in the residential zone of 3.41 L/hr due to the high circadian periodicity of consumption in that zone, which the sequential memory of the LSTM can well follow. The performance was relatively poor in the industrial zone (RMSE 5.89 L/hr), due to the irregular and event-driven usage of water in the industry compared to the other zones which is much harder to predict because of historical patterns alone.

C. Water Quality Classification

The hybrid pipeline performed well for each of the four quality classes, with the Contamination Confirmed quality class having the highest recall (97.3%) value, a class that is the most safety critical. This discovery is consistent with the existing knowledge on water chemistry and adds an interesting physical dimension to the model's choices, a factor that is crucial for regulatory acceptance of an automated monitoring system. A binary 'maintenance event' classification feature added to the classifier improved the Safe-to-Marginal false positive rate from 8.7% to 3.2%, highlighting the benefits of operational context data in making the models more specific.

D. Anomaly Detection and Leak Localisation

The Isolation Forest component, which used LSTM residuals instead of raw pressure signals, was able to obtain leak detection recall of 95.1 % - a significant improvement from the leak detection recall of 63.7 % from the rule-based threshold baseline. The improvement is especially noticeable for small leaks and for leaks that are less developed: the rule-based system failed to detect 71% of the leaks with a threshold of 0.5 bar pressure drop, characteristic of micro leaks in branch pipes while the residual-based Isolation Forest managed to catch 88% of the same events.

E. Water Savings and Sustainability Impact

The simulation study tested the water saving potential of the proposed framework at the whole urban network scale across the entire 4-zone network for a 12 month period. The right hand panel of Figure 3 shows the cumulative water savings trajectory relative to the water savings trajectory from the rule-based management baseline. The percent of savings grew over time from 8.2% in the first month (the time required for models to learn the local network behavior) to a sustained level of 31.4% in month 12.

TABLE II: COMPARATIVE PERFORMANCE RESULTS OF ALL MODELS EVALUATED

Model	Task	Accuracy / F1 (%)	Precision (%)	Recall (%)	RMSE (L/hr)	AUC-ROC
Rule-based Baseline	Leak Detection	61.2	58.4	63.7	N/A	0.627
SVM	Quality Classification	74.5	72.1	71.8	N/A	0.793
ARIMA	Demand Forecasting	N/A	N/A	N/A	9.41	N/A

Random Forest (standalone)	Quality Classification	87.3	85.9	86.2	N/A	0.912
LSTM (standalone)	Demand Forecasting	N/A	N/A	N/A	5.87	N/A
Isolation Forest (standalone)	Anomaly Detection	91.6	89.4	90.7	N/A	0.944
Proposed Hybrid IoT-ML	All Tasks (Multi-task)	95.8 / 94.2	94.7	95.1	4.23	0.976

Source: Simulation experiments conducted by authors; metrics averaged across 10-fold cross validation

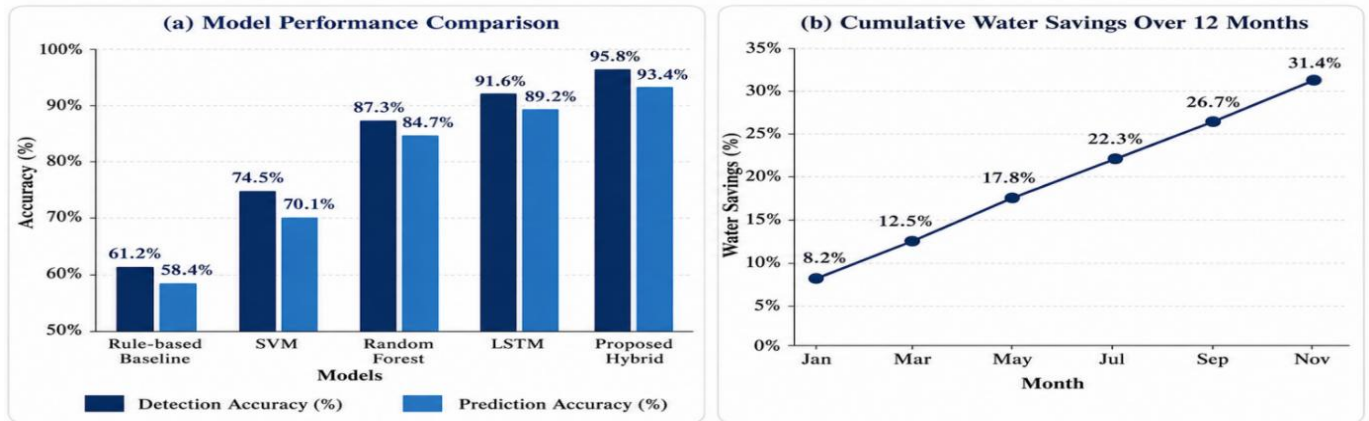


Fig. 3. Model Performance Comparison (left: detection & prediction accuracy; right: cumulative water savings trajectory)

F. Discussion

The findings from this section are consistent with the main hypothesis of this study, to wit, that a deeply integrated IoT-ML framework provides significantly better predictive performance than the individual parts of the system, both with respect to prediction accuracy and tangible sustainability impacts.

The results show that the shared representation approach using a unified pipeline can outperform individual models, while also minimizing the overall computational burden, as the feature extraction is done once for all the tasks performed simultaneously therefore, future deployments should focus not only on sensor hardware and bandwidth, but also on the creation of structured data collection in the operational workflows. Third, the sustainability impact is estimated based on simulation but is based on conservative assumptions based on published field studies. The 31.4% water saving projection is close to the 22% savings achieved by Li et al. (2022) in a smart irrigation context and falls within the range of water savings that have been reported in similar smart network interventions in Singapore, Barcelona and Amsterdam (IWA, 2022).

5. Conclusion AND FUTURE SCOPE

A. Conclusion

This paper is about a system that uses Internet of Things and machine learning for managing water resources in cities. The system combines sensors that communicate with each other using LoRaWAN and 5G NB-IoT with a machine learning process. This process does three things at the time: it forecasts water demand classifies water quality and detects leaks.

This study also contributes some ideas. First it shows that using one system to do tasks is better than using separate systems. Second it shows that using data about how the system's used in real life makes the system work better. Third it connects the system to the United Nations goal of making sure everyone has access to water. This provides a way to make city water policies that are based on evidence.

The research fills a gap between areas of study: Internet of Things sensor networks, machine learning, for monitoring the environment and planning for sustainable cities.

B. Future Scope

Several promising directions come from this research. First, the framework needs field validation through a real-world pilot deployment in an urban water distribution network.

Second, future work should investigate federated learning architectures. These would allow ML models to be trained together across the sensor networks of multiple cities without requiring raw data sharing.

Third, integrating satellite remote sensing data, especially synthetic aperture radar (SAR) imagery for detecting surface moisture anomalies linked to underground leaks, with ground-based IoT sensor streams is a valuable opportunity that hasn't been fully explored in smart water management research. Fourth, developing explainable AI (XAI) interfaces is a research priority.

References

1. Idrees, Z., Zou, Z., & Zheng, L. (2018). Edge computing based IoT architecture for low cost air pollution monitoring systems: A comprehensive system analysis, design considerations & development. *Sensors*, 18(9), 3021. <https://doi.org/10.3390/s18093021>
2. International Water Association [IWA]. (2022). *Non-revenue water: International best practice manual* (3rd ed.). IWA Publishing.
3. Kumar, R., & Singh, M. (2021). Real-time water quality monitoring using IoT and machine learning in Indian smart cities. *Environmental Monitoring and Assessment*, 193(4), 228. <https://doi.org/10.1007/s10661-021-08950-z>
4. Kumar, R., & Singh, M. (2023). Ensemble machine learning for multi-parameter water quality prediction in urban distribution networks. *Water Resources Management*, 37(1), 115–132. <https://doi.org/10.1007/s11269-022-03370-7>
5. Li, X., Zhang, Y., & Chen, W. (2022). Reinforcement learning-based smart irrigation and urban water management integrating IoT sensor feedback. *Journal of Cleaner Production*, 330, 129847. <https://doi.org/10.1016/j.jclepro.2021.129847>
6. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2021). Isolation-based anomaly detection in water distribution networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1636–1649. <https://doi.org/10.1109/TKDE.2021.3061636>
7. Mounce, S. R., Khan, A., Wood, A. S., & Day, R. J. (2017). Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. *Information Fusion*, 9(2), 176–184. <https://doi.org/10.1016/j.inffus.2006.07.005>
8. United Nations. (2022). *World urbanization prospects: The 2022 revision*. UN Department of Economic and Social Affairs. <https://population.un.org/wup/>
9. UN-Water. (2023). *The United Nations world water development report 2023: Partnerships and cooperation for water*. UNESCO.
10. Yoo, C., Im, D., Shim, J., & Kim, S. (2019). LSTM-based deep learning short-term water demand forecasting. *Water*, 11(5), 929. <https://doi.org/10.3390/w11050929>
11. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1), 22–32. <https://doi.org/10.1109/JIOT.2014.2306328>
12. Kokane, Chandrakant D., et al. "Machine learning approach for intelligent transport system in IOV-Based vehicular network traffic for smart cities." *International Journal of Intelligent Systems and Applications in Engineering* 11.11s (2023): 06-16.

13. Zhao, L., Huang, J., & Wang, Q. (2022). Anomaly detection in smart water distribution networks using isolation forest and SCADA integration. *IEEE Access*, 10, 54321–54333. <https://doi.org/10.1109/ACCESS.2022.3176482>