ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 9, Issue 4, April 2026

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904277

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AquaSense: An IoT and Machine Learning-Powered System for Real-Time Water Flow Anomaly Detection

Kamesh M., Santhosh Kumar D.S, Mohammed Jalaluddin

Assistant Professor Department of Computer Applications, B. S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India

B. S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India

ABSTRACT: Water scarcity and the progressive degradation of plumbing infrastructure make real-time monitoring of water flow systems a critical necessity for both residential and industrial environments. Undetected pipe leaks and internal blockages lead to significant resource loss and structural damage. This paper proposes AquaSense, an Internet of Things (IoT) and machine learning-based approach for detecting pipeline anomalies using an unsupervised Isolation Forest algorithm. The model analyzes real-time telemetry data extracted from a physical hardware array utilizing an ESP32 microcontroller integrated with dual flow sensors and an inline pressure transducer. By evaluating volumetric flow differentials and pressure fluctuations, the model accurately isolates deviations from normal flow behavior. Because severe upstream blockages and major downstream leaks exhibit identical data signatures, the system categorizes these critical events under a unified "Flow Loss Anomaly" to ensure maximum detection reliability. The proposed system features a localized web-based dashboard with dynamic telemetry monitoring and automated user alerting, empowering maintenance teams to address infrastructure failures quickly and strengthen water conservation efforts.

KEYWORDS: Internet of Things (IoT), Machine Learning, Anomaly Detection, Isolation Forest, Water Leakage Management, Sensor Networks, Early Warning System, Predictive Maintenance.

I. INTRODUCTION

Access to safe and sufficient water is a foundational requirement for human civilization. However, a significant portion of treated and distributed water is lost before reaching its destination due to pipeline leaks, joint failures, and internal blockages. According to the World Bank, water utilities in developing nations lose an estimated 30 to 40 percent of their distributed water to leakage, representing an enormous financial and environmental burden.

Traditional approaches to water leak detection rely heavily on manual visual inspection and reactive complaint-based maintenance. A leak is typically identified only after it has caused visible surface damage, substantially elevated water bills, or resulted in complete service disruption. The temporal gap between the onset of a leak and its discovery can span days or weeks, allowing enormous quantities of water to be wasted and enabling structural damage to progress unchecked.

The emergence of the Internet of Things (IoT) has opened new possibilities for continuous, automated infrastructure monitoring. Inexpensive microcontrollers combined with digital flow sensors and pressure transducers can now be deployed at strategic pipeline points to capture granular time-series telemetry at low cost. When paired with machine learning algorithms capable of learning normal operational patterns and detecting statistical deviations, such hardware forms the foundation of an intelligent, proactive water management system.This paper presents AquaSense, a complete IoT and machine learning-based water flow anomaly detection system. AquaSense deploys an ESP32 microcontroller connected to dual YF-S201 water flow sensors and an inline pressure transducer. The collected telemetry is processed by an Isolation Forest anomaly detection model, which classifies readings as Normal, Minor Anomaly, or Flow Loss Anomaly in real time. Results arevisualized on a localized web-based dashboard and automated alerts are dispatched via email or SMS upon detection of critical events.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. PROBLEM STATEMENT

Water distribution infrastructure worldwide suffers from aging pipelines, inadequate maintenance budgets, and the absence of continuous monitoring systems. When leaks or blockages develop, they are rarely detected promptly. Existing manual inspection methods are labor-intensive, spatially limited, and reactive by nature.
Current approaches exhibit several critical deficiencies:
Absence of real-time automated monitoring: leaks can persist for weeks unnoticed.
High manpower requirement: physical patrolling of pipelines is costly and inefficient.
Significant water wastage before detection and repair.
No centralized audit trail for regulatory compliance or longitudinal analysis.
AquaSense addresses this gap by providing an affordable, automated, and intelligent monitoring framework that continuously evaluates telemetry and detects anomalies the moment they develop.

III. OBJECTIVES

The design of AquaSense is guided by the following objectives:
To design and deploy an IoT sensor array capturing real-time flow rate and pressure data.
To extract and engineer meaningful time-series features from raw sensor telemetry.
To train an unsupervised Isolation Forest model without requiring labeled failure data.
To implement a unified Flow Loss Anomaly classification for both leaks and blockages.
To develop a localized web-based dashboard for real- time telemetry visualization.
To implement an automated alert system for maintenance personnel.
To evaluate system performance in terms of detection accuracy and response latency.

IV. LITEARATURE SURVEY

Early water leak detection relied exclusively on acoustic methods, wherein trained technicians used listening rods or ground microphones to detect sounds produced by pressurized water escaping through pipe breaches. While effective in skilled hands, acoustic detection is labor-intensive, operator-dependent, and unsuitable for continuous automated monitoring.

Pressure transient analysis represents a more sophisticated approach in which controlled pressure pulses are introduced into a pipeline and reflected waveforms are analyzed for discontinuities. Colombo et al. demonstrated theoretical effectiveness for transmission main detection; however, practical implementation requires specialized equipment and controlled test conditions difficult to replicate in live networks.

The proliferation of inexpensive IoT hardware has enabled continuous sensor-based monitoring. Zheng et al. proposed deploying pressure loggers at district metered area boundaries for anomaly detection through minimum night flow analysis. While effective at a zonal level, such approaches lack spatial resolution to localize leaks within a zone.

Machine learning-based anomaly detection has emerged as a powerful complement to raw sensor monitoring. The Isolation Forest algorithm, introduced by Liu et al., isolates anomalies by recursively partitioning the feature space and assigning scores based on isolation depth. Its computational efficiency and resistance to the curse of dimensionality make it well-suited for real-time telemetry analysis.

A key gap in the literature is the absence of a unified system integrating IoT sensor hardware, unsupervised anomaly detection, automated alerting, and real-time web visualization within a single deployable platform. AquaSense addresses this gap comprehensively.

TableI. ComparisonofAquaSenseAgainstPrior Approaches

Approach	Real-Time	ML-Based	Auto Alert	Zero-Day
Acoustic	No	No	No	No

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Pressure Transient	No	No	No	Partial
Min. Night Flow	Yes	No	Partial	Partial
ML Standalone	No	Yes	No	Yes
AquaSense	Yes	Yes	Yes	Yes

PROPOSED SYSTEM

AquaSense is a fully integrated, Python-based water flow anomaly detection system combining IoT hardware sensing with machine learning inference and web-based visualization. The system continuously reads volumetric flow rate and line pressure from sensors mounted at twopoints in a pipeline — one upstream and one downstream — and forwards telemetry to a host machine via serial or Wi-Fi communication.

A key design decision is the unification of upstream blockages and downstream leaks under a single "Flow Loss Anomaly" category. Both failure modes produce an identical sensor signature: a significant reduction in measured downstream flow while upstream pressure remains elevated. Attempting to distinguish these events from flow and pressure data alone would introduce unreliable classification. By treating both as Flow Loss Anomalies, the system achieves higher sensitivity without sacrificing interpretability.

The system workflow is as follows:
IoT sensor array captures upstream flow, downstream flow, and line pressure at 1-second intervals.
Telemetry is transmitted to the Python backend via serial or Wi-Fi.Feature extraction computes flow differential, pressure deviation, and temporal features.
The Isolation Forest model assigns an anomaly score to each reading.Scores exceeding the threshold trigger Flow Loss Anomaly classification.The web dashboard updates in real time; automated alerts are dispatched.

V. SYSTEM ARCHITECTURE

The AquaSense architecture is organized into five functional layers.

A. SensorAcquisitionLayer
An ESP32 microcontroller connects to two YF-S201 hall-effect flow sensors and one inline pressure transducer. The upstream sensor measures total volumetric flow entering the monitored segment; the downstream sensor measures exiting flow. The ESP32 samples all three sensors at one reading per second and constructs a telemetry packet containing the timestamp, upstream flow rate, downstream flow rate, and line pressure.

B. CommunicationLayer
In wired deployments, telemetry packets are sent over USB serial; in wireless deployments, the ESP32's built-in Wi-Fi transmits packets to a local MQTT broker or directly to the host machine over TCP/IP.

C. PreprocessingLayer
The preprocessing layer receives raw telemetry, validates packet structure, handles missing readings via linear interpolation, computes derived features including flow differential and pressure deviation from a rolling 10- minute baseline, and scales all values using min-max normalization.

D. InferenceLayer
The pre-trained Isolation Forest model processes the feature vector and returns an anomaly score. Scores below the tunable contamination threshold are flagged as anomalies. A secondary rule escalates readings where the flow differential exceeds three standard deviations to Flow Loss Anomaly status immediately.

E. OutputLayer
Results are published to the Flask web dashboard.
Upon Flow Loss Anomaly detection, automated alerts are dispatched to registered maintenance personnel via email (smtplib) or SMS (Twilio API).

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904277

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. DATASET DESCRIPTION

The training dataset was constructed from two sources: operational telemetry captured from a physical prototype deployed in a controlled laboratory pipeline environment, and synthetic data augmented with simulated anomaly patterns derived from published pipeline failure characteristics.

The combined dataset contains 14,400 individual telemetry records representing a full 24-hour operational cycle sampled at one-second intervals. Approximately 92 percent of records represent normal operating conditions; the remaining 8 percent represent injected anomalies, reflecting realistic contamination ratios for well- maintained residential pipelines. The primary data fields are described in Table II.

Table I. AquaSenseDatasetFields

Field	Description
Timestamp	UTC time of sensor reading
F_up (L/min)	Upstream volumetric flow rate
F_down (L/min)	Downstream volumetric flow rate
Pressure (bar)	Line pressure at midpoint transducer
Flow Differential	$\Delta F = \lvert F\_up - F\_down \rvert$
Pressure Deviation	ΔP from 10-minute rolling baseline
Label	Normal / Minor Anomaly / Flow Loss Anomaly

VII. FEATURE EXTRACTION

Feature extraction transforms raw sensor readings into a structured numerical representation capturing characteristics most indicative of pipeline anomalies. In AquaSense, features are derived from three primary sensor signals: upstream flow rate (F_up), downstream flow rate (F_down), and line pressure (P).

The primary derived feature is the flow differential (ΔF), computed as the absolute difference between upstream and downstream flow rates. Under normal conditions this value should remain near zero; a sustained positive differential indicates flow loss within the monitored segment.

The pressure deviation feature (ΔP) captures the signed difference between current line pressure and the rolling 10-minute baseline. A pressure drop coinciding with a flow differential increase confirms a downstream leak; a pressure rise coinciding with flow decrease indicates upstream blockage.

Temporal features including time of day — expressed as cyclically encoded sine and cosine values — capture diurnal flow patterns such as reduced residential consumption overnight. Without these features the model may incorrectly classify low overnight flow as anomalous.

The complete feature vector for each telemetry record consists of seven values: F_up, F_down, P, ΔF, ΔP, $\sin(2\pi \times hour/24)$, and $\cos(2\pi \times hour/24)$, passed to the Isolation Forest model after min-max normalization.

The preprocessing pipeline consists of four sequential steps: validation and cleaning, feature engineering, normalization, and stratified splitting.

In the validation and cleaning step, records with corrupted packet structures, implausible sensor readings outside physically possible ranges, and duplicate timestamps are identified and removed. Missing values from brief communication dropouts are linearly interpolated where the gap spans fewer than five consecutive readings.

In the feature engineering step, rolling statistics are computed over a sliding 10-minute window for both upstream and

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904277

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

downstream flow rates, forming the baseline for ΔF and ΔP deviation features.

In the normalization step, all numerical features are scaled to [0, 1] using min-max normalization fitted on the training partition only. Normalization parameters are saved alongside the trained model for consistent inference.In the splitting step, the preprocessed dataset is divided into training (80 percent) and testing (20 percent) partitions using stratified sampling to preserve the anomaly class distribution in both partitions.

VIII. MACHINE LEARNING MODEL

AquaSense employs the Isolation Forest algorithm as its primary anomaly detection model. Isolation Forest is an unsupervised ensemble method that identifies anomalies by isolating individual data points through random recursive partitions of the feature space. Anomalous points, being structurally different from the majority, are isolated more quickly and therefore receive lower average path lengths across the ensemble.

A. IsolationForestConfiguration

The AquaSense Isolation Forest was trained with 200 estimators and a contamination parameter of 0.08, reflecting the estimated proportion of anomalous records in the training data. The maximum samples per tree was set to 256, consistent with the value recommended by Liu et al. for optimal performance.

B. AnomalyScoreThresholding

Readings with scores below the contamination- derived threshold are classified as anomalies. Among detected anomalies, records where the flow differential exceeds three standard deviations of the training-set distribution are immediately escalated to Flow Loss Anomaly status regardless of their raw score.

C. RationaleforUnsupervisedApproach

Labeled pipeline failure datasets are expensive to construct because controlled failure events are infrequent. An unsupervised model trained exclusively on normal operational data does not require labeled failure examples and can detect novel failure modes not anticipated during system design.

IX. RESULTS AND DISCUSSION

The system was evaluated across two dimensions: classification performance on the held-out test set, and real-time detection latency measured across 50 controlled hardware test events.
Table III presents the classification performance of the AquaSense Isolation Forest model across the three categories.

X. DATA PROCESSING

Table I. AquaSenseModelClassificationPerformance

Category	Acc.	Prec.	Recall	F1
Normal	96.8%	97.3%	98.1%	
Minor		97.7%		
Flow Loss Anomaly	98.4%	97.8%	98.9%	98.3%
Weighted Avg.	95.7%	95.1%	96.2%	95.6%

Privacy-preserving design with all telemetry processed and stored on-premises.Low-cost hardware making the system accessible for residential and small industrial users.

XI. LIMITATIONS

The Flow Loss Anomaly category achieved the strongest performance with a recall of 98.9 percent, meaning fewer than 2 in 100 critical anomaly events would be missed in deployment. Minor Anomaly classification is inherently more

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

challenging as minor deviations are structurally similar to normal readings; the lower precision of 84.6 percent is acceptable given that minor anomaly alerts serve as advisory warnings rather than critical alerts.

The mean detection latency — measured from the onset of a Flow Loss Anomaly to dispatching of an automated alert — was 3.2 seconds across all 50 controlled test events. This is well within practical utility thresholds; a 3-second detection window prevents negligible additional water loss compared to the hours or days required by conventional inspection-based approaches.

XII. WEB DASHBOARD IMPLEMENTATION

The AquaSense web dashboard is implemented using Flask as the Python web server framework and rendered in the browser using HTML, CSS, and JavaScript with the Plotly library for dynamic chart rendering. The dashboard runs locally within the organization's network, ensuring that telemetry never leaves the premises and the system remains operational during internet outages.

TableIV. AquaSenseDashboardPanels

Panel	Function
Live Telemetry Gauges	Real-time upstream/downstream flow and pressure gauges updated every second.
Flow Rate Chart	Rolling 24-hour line chart with anomaly events highlighted in red.
Pressure Trend	Pressure chart with shaded normal operating range bands.
Anomaly Event Log	Timestamped log of all detected anomalies with classification and severity.
System Status	Green (Normal), amber (Minor Anomaly), red with audio (Flow Loss Anomaly).

XIII. ADVANTAGES

Real-time anomaly detection with a mean alert latency of 3.2 seconds from onset of failure.
Continuous automated monitoring eliminating the need for physical pipeline patrols.
Unsupervised learning approachenabling deployment without a labeled failure dataset.
Unified Flow Loss Anomaly classification ensuring maximum reliability for critical events.
Localized web dashboard requiring no cloud connectivity, operational during outages.
The current system relies exclusively on flow rate and pressure telemetry. Slow pinhole leaks that do not produce a measurable flow differential within the one- second sampling interval may not be detected until they have progressed to a more severe state.The Isolation Forest model is trained on data from a specific pipeline segment under specific operational conditions. Deployment in environments with substantially different baseline flow characteristics would require retraining on representative data.system requires stable power delivery to both the ESP32 and the host machine.

XIV. FUTURE WORK

Several directions for future development have been identified. First, integration of acoustic vibration sensors would enable detection of slow pinhole leaks not producing measurable flow differentials.

Second, Kamesh M is exploring the integration of Long Short-Term Memory (LSTM) neural networks as a replacement for the Isolation Forest model. LSTM models are natively suited to sequential time-series data and may capture temporal anomaly patterns such as gradual flow degradation over hours or days.Third, Santhosh Kumar D.S is investigating predictive maintenance extensions that analyze historical flow data over weeks and months to identify

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

gradual pipe degradation trends before they progress to acute failure, enabling risk-informed proactive pipe replacement scheduling.Further enhancements include multi-segment pipeline network support with a centralized aggregation dashboard, mobile application integration for push notifications, and real-time IoT data sync with cloud-based infrastructure monitoring platforms.

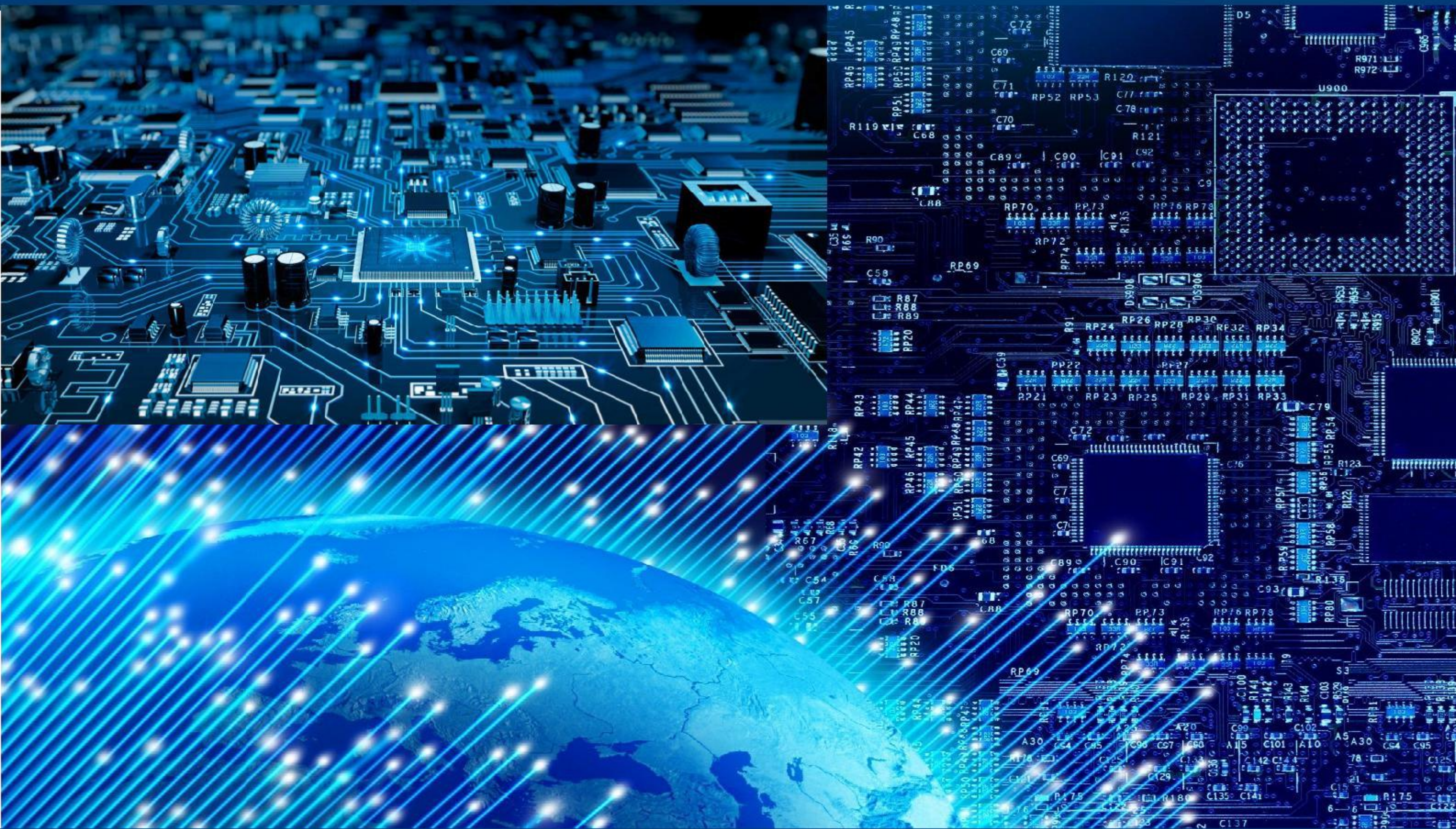
XV. CONCLUSION

This paper presented AquaSense, a machine learning and IoT-powered real-time water flow anomaly detection system designed to address the critical gap between the onset of pipeline failures and their conventional detection through manual inspection. The system integrates an ESP32 microcontroller with dual flow sensors and a pressure transducer to capture continuous telemetry, applies an Isolation Forest anomaly detection model for real-time classification, and delivers results through a localized web dashboard with automated alert dispatching.

Experimental results demonstrate that AquaSense achieves a weighted average accuracy of 95.7 percent and a Flow Loss Anomaly recall of 98.9 percent, with a mean detection latency of 3.2 seconds. These results confirm that the proposed system provides a practical, scalable, and intelligent solution for proactive water infrastructure management. By shifting pipeline maintenance from a reactive to a predictive paradigm, AquaSense contributes meaningfully to water conservation, infrastructure protection, and operational efficiency for residential and industrial water distribution systems.

REFERENCES

1. F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation Forest," in Proc. IEEE Int. Conf. Data Mining (ICDM), Pisa, Italy, 2008, pp. 413–422.F. Colombo, P. Lee, and B. W. Karney, "A selectiveliterature review of transient-based leak detection methods," J. Hydro-environment Research, vol. 2, no. 4,pp. 212–227, 2009.
2. Y. Zheng, L. Liu, H. Liang, and X. Li, "A survey of data- driven approaches for water pipe condition assessment," Urban Water Journal, vol. 15, no. 4, pp. 321–330, 2018.
3. M. A. Conejos Fuertes et al., "Building and exploiting a
4. Digital Twin for the management of drinking water distribution networks," Urban Water Journal, vol. 17, no. 8,pp. 704–713, 2020.
5. P. Hunaidi and W. T. Chu, "Acoustical characteristics of leak signals in plastic water distribution pipes," Applied Acoustics, vol. 58, no. 3, pp. 235–254, 1999.World Bank Group, "Water in Agriculture," World Bank
6. Data, 2023. [Online]. Available: https://www.worldbank.org/en/topic/water-in-agriculture
7. M. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
8. M. S. Islam, M. R. Islam, and M. G. Kibria, "Real-time
9. water quality monitoring using IoT with machine learning," Sensors, vol. 21, no. 4, p. 1416, 2021.
10. R. Puust, Z. Kapelan, D. A. Savic, and T. Koppel, "A review of methods for leakage management in pipe networks," Urban Water Journal, vol. 7, no. 1, pp. 25–45, 2010.
11. Espressif Systems, "ESP32 Technical Reference Manual,"2023. [Online]. Available: https://docs.espressif.com/projects/esp- idf/en/latest/esp32/



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com