Ai Smart Water Leak Detection

Rehema Malenda, Christina Longo, Happyness Paul, Sharom Mwakisopile, Shakila Sadi, Shukuru Thobias, Simon James, Abdallah Ali.

Department of Computer Science, Faculty of Information Communication Technology.
Ruaha Catholic University (RUCU)
Iringa, Tanzania
Rehemaismail030@gmail.com

Abstract: Water loss in distribution systems, commonly referred to as non-revenue water (NRW), represents a major global challenge for water utilities, contributing to resource wastage, economic losses, and sustainability issues. This review paper critically examines the integration of Artificial Intelligence (AI) and related digital technologies for smart water loss detection, with a particular emphasis on addressing the constraints of aging infrastructure, limited resources, and variable operational conditions in developing and urbanizing regions. Through a systematic synthesis of recent advancements in machine learning models, sensor-based monitoring, acoustic analysis, pressure data analytics, and ensemble approaches, the analysis identifies persistent gaps in data scarcity, model generalization, noise handling, real-time deployment, and integration with existing systems. While individual AI techniques demonstrate high detection accuracy in controlled settings, their practical application remains fragmented, often reliant on high-quality labeled data or cloud infrastructure that may not suit intermittent connectivity or low-resource environments. In response, this paper proposes a conceptual framework for a unified "AI Smart Water Loss Detection" platform that synergistically combines anomaly detection algorithms, edge computing for on-device processing, IoT sensor fusion, and adaptive learning within a resilient, hybrid architecture. The paper concludes with strategic recommendations for researchers, utilities, and policymakers to advance effective, scalable, and equitable AI-driven solutions that minimize water losses, enhance infrastructure resilience, and support sustainable water management

Keywords: **Smart Water Management, Artificial Intelligence, Water Loss Detection, Non-Revenue Water, Leak Detection, Machine Learning, IoT in Water Systems, Anomaly Detection.**

1. INTRODUCTION

The escalating global demand for water, coupled with aging infrastructure and climate variability, has intensified the urgency of minimizing losses in water distribution networks. Non-revenue water encompassing physical leaks, apparent losses from metering inaccuracies, and unauthorized consumption accounts for substantial portions of supplied water in many regions, often exceeding 30-50% in developing utilities. Smart water management, empowered by digital transformation, offers transformative potential to detect, locate, and mitigate these losses efficiently.

This review focuses on Artificial Intelligence as a pivotal enabler for smart water loss detection. AI techniques, including machine learning, deep learning, and anomaly detection algorithms, analyze diverse data streams from pressure sensors, flow meters, acoustic devices, and smart meters to identify deviations indicative of leaks or inefficiencies. Framed around the conceptual "AI Smart Water Loss Detection" approach, this analysis explores why current solutions often fall short in real-world deployment and proposes principles for resilient, context-aware systems that bridge the divide between technological capability and practical implementation in resource-constrained settings.

1.1 BACKGROUND

Water is a finite and increasingly stressed resource globally, with demand projected to exceed sustainable supply in many regions by 2030 due to population growth, urbanization, industrialization, and climate change-induced variability in precipitation patterns. Water utilities worldwide are under mounting pressure to deliver reliable, equitable, and affordable services while simultaneously reducing operational inefficiencies and financial losses [1]. A major source of inefficiency in drinking water supply systems is non-revenue water (NRW) the difference between the volume of water supplied into the distribution network and the volume billed to customers [2]. NRW comprises three principal components:

• Physical (real) losses — mainly leaks and bursts in pipes, joints, valves, and reservoirs;

• Commercial (apparent) losses — metering inaccuracies, data handling errors, theft, and unauthorized consumption;

• Unbilled authorized consumption — water used for firefighting, flushing, public facilities, and operational purposes that is not charged.

According to international benchmarks published by the International Water Association (IWA), NRW levels in well-managed utilities in high-income countries typically range between 5–15%, whereas in many developing and middle-income countries including large parts of Sub-

Saharan Africa, South Asia, and Latin America NRW frequently exceeds 30–50%, and in some urban systems surpasses 60%. These elevated loss rates translate into enormous volumes of treated and pumped water being wasted every year, imposing severe economic, social, and environmental costs [3]. Economically, utilities lose revenue that could fund infrastructure maintenance, network expansion, or tariff stabilization. Environmentally, the energy consumed in abstracting, treating, and pumping lost water contributes unnecessarily to greenhouse gas emissions. Socially, high NRW exacerbates water scarcity for low-income households, informal settlements, and peri-urban communities that already face intermittent or inadequate supply [4].

Historically, water loss management relied on labour-intensive and episodic methods: ground listening with analogue acoustic devices, step-testing of district metered areas (DMAs), night-flow analysis, correlators, and large-scale leak detection surveys conducted every few years. While effective in some contexts, these approaches are slow, coverage-limited, costly per detected leak, and reactive rather than preventive. They also struggle to detect small but cumulatively significant background leaks, which can account for 50–70% of physical losses in aging networks.

The convergence of low-cost sensors, wireless communication, big data infrastructure, and Artificial Intelligence has given rise to the concept of smart water management. Within this paradigm, continuous, high-resolution monitoring of hydraulic and acoustic parameters combined with advanced analytics enables utilities to shift from periodic, manual interventions to proactive, data-driven loss control. AI techniques now play a central role in this transition. Supervised and unsupervised machine learning models identify anomalous patterns in pressure, flow, and consumption time series. Deep learning architectures (particularly convolutional and recurrent neural networks) classify leak signatures from acoustic and vibration signals. Ensemble methods and graph-based algorithms localize leaks within complex pipe networks by fusing multi-sensor data. Anomaly detection frameworks based on autoencoders, isolation forests, or one-class SVMs flag deviations from normal behavior even when labelled leak examples are scarce.

Despite these advances, the translation of laboratory-demonstrated or pilot-scale AI performance into sustained, city-wide operational systems remains limited, especially in utilities that face the most acute NRW challenges. Common barriers include sparse and low-quality sensor coverage, intermittent power and communication networks, high levels of background noise, heterogeneous pipe materials and ages, limited historical labelled data, shortage of in-house data science expertise, and high initial capital costs relative to constrained utility budgets. Moreover, most published AI models have been developed and validated using datasets from high-income countries with modern, well-instrumented networks, raising questions about transferability to infrastructure typical of the Global South asbestos-cement, cast iron, PVC, and galvanized steel pipes laid decades ago under variable soil conditions and subjected to frequent third-party damage.

Complementary technologies further shape the landscape. Internet of Things (IoT) platforms enable dense, low-power sensor deployments. Edge computing reduces latency and bandwidth dependence by performing preliminary inference close to the data source. Cloud-based analytics support model training and long-term trend analysis. Digital twins and hydraulic simulation models provide contextual baselines against which anomalies can be judged. Yet the literature reveals a persistent fragmentation: acoustic AI solutions rarely integrate pressure transients, pressure-based localization seldom incorporates consumption patterns, and most edge prototypes stop short of full end-to-end decision support. This creates a clear research and engineering opportunity to design integrated, context-aware AI platforms that are resilient to the infrastructural, financial, and operational realities of high-NRW utilities while delivering measurable reductions in lost water volume, energy use, and repair costs.

1.2 PROBLEM STATEMENT

Despite growing academic and commercial interest in AI-driven water loss detection, the majority of water distribution systems particularly in developing and rapidly urbanizing regions continue to experience persistently high levels of non-revenue water and delayed leak identification. Current detection practices and digital solutions suffer from several interlocking deficiencies that collectively undermine effectiveness, scalability, and equity:

Prolonged Detection Times and High-Volume Losses

Most utilities still depend on periodic manual surveys, consumer complaints, visible surface breaks, or infrequent night-flow measurements to discover leaks. Small-to-medium background leaks, which often constitute the largest share of physical losses, can remain undetected for months or even years, resulting in cumulative water wastage measured in millions of cubic metres annually per city.

Fragmentation of Monitoring and Analysis Capabilities

Data sources and analytical tools remain largely siloed. Pressure loggers may be installed in one part of the network, acoustic sensors in another, smart meters in selected zones, and SCADA data confined to treatment plants and major pumping stations. There is seldom a unified platform capable of ingesting, correlating, and jointly interpreting these heterogeneous streams in real time to produce actionable intelligence. As a result, operators must manually cross-

reference outputs from multiple disconnected systems, leading to delayed or missed opportunities for intervention.

Connectivity, Power, and Infrastructure Constraints

Many high-loss zones peri-urban extensions, informal settlements, rural supply schemes, and aging inner-city districts suffer from unreliable electricity, intermittent 2G/3G/4G coverage, or complete absence of communication infrastructure. Cloud-dependent AI models become inoperable during outages, while high-bandwidth continuous data streaming is economically and technically unfeasible. This creates a sharp divide between what is demonstrated in well-equipped pilot areas and what can be reliably deployed network-wide.

Data Scarcity, Noise, and Localization Bias

Effective supervised AI models require large volumes of accurately labelled leak events, which are expensive and time-consuming to collect. In practice, most utilities possess only sparse, noisy, or unlabeled operational data. Models trained on clean datasets from modern European or North American networks frequently exhibit degraded performance when applied to pipes made of different materials, buried in different soils, subjected to different pressure regimes, or affected by different consumption behaviors (e.g., intermittent supply, illegal connections, standpipe usage).

Limited Contextual Awareness and Decision Support

Many existing AI solutions focus narrowly on binary classification ("leak / no leak") or basic anomaly flagging without providing utilities with interpretable, prioritized, and spatially contextualized recommendations. Operators receive alerts without sufficient information about likely leak size, severity ranking, probable location accuracy, expected water savings if repaired promptly, or integration with work-order and asset-management systems.

Economic and Institutional Barriers to Adoption

Even when technically promising prototypes exist, deployment is hindered by high upfront costs, lack of local technical capacity to maintain and update models, absence of performance-based procurement frameworks, and insufficient evidence of long-term return on investment under real operating conditions.

Collectively, these shortcomings mean that the transformative potential of AI to enable near-real-time, network-wide, proactive water loss management remains largely unrealized in the very contexts where the need is greatest. There is therefore an urgent need for integrated, resilient, and contextually intelligent AI platforms that overcome fragmentation, function reliably under constrained conditions, generalize across diverse infrastructure types, and deliver clear, quantifiable benefits to utilities, consumers, and the environment.

1.3 Objectives

Main Objective:
To design and develop an AI-based smart system for real-time detection and mitigation of water losses in water distribution networks to reduce Non-Revenue Water and improve overall system efficiency.

Specific Objectives:
1. To collect and analyze real-time data from sensors (flow, pressure, acoustic/vibration) for anomaly detection.
2. To implement machine learning models (e.g., anomaly detection, classification) capable of identifying leak patterns with high accuracy.
3. To develop a system prototype that provides alerts, leak localization estimates, and visualization dashboards for utility operators.
4. To evaluate the system's performance in terms of detection accuracy, false positive rate, response time, and water savings potential.

2.0 Related Work

Research in AI for water loss detection spans multiple domains, each advancing specific aspects of the problem

2.1 AI powered Anomaly Detection and Machine Leaning Model

Extensive studies validate machine learning for identifying irregularities in flow, pressure, and consumption data [5]. Ensemble methods and classifiers like Random Forest, XGBoost, and neural networks achieve high accuracies in controlled or simulated environments by learning patterns from historical and real-time datasets. However, many operate as offline analyzers or require extensive labeled data, limiting real-time adaptability to dynamic events such as bursts or gradual leaks [6].

2.2 Acoustic and Vibration-Based Detection with AI

AI enhances acoustic monitoring by classifying leak signatures from sensor data using convolutional neural networks or feature-engineered models [7] [8]. These approaches excel in pinpointing leaks in metallic pipes but face challenges with noise interference, non-metallic materials, and computational demands that restrict field deployment without cloud support [9] [10].

2.3 Pressure and Flow Analytics with Geospatial Integration Pressure

Transient monitoring and hydraulic modeling combined with AI enable zone-based leak localization [11]. Machine learning refines predictions by correlating anomalies across network segments. Limitations include dependency on dense sensor coverage and assumptions of stable hydraulics, reducing effectiveness in variable or poorly instrumented systems [12].

2.4 IoT and Edge Computing for Resilient Monitoring

IoT deployments facilitate continuous data acquisition, while edge processing enables preliminary detection in low-connectivity scenarios [13]. Emerging works explore hybrid

models that prefetch data or run lightweight inference locally [14]. Gaps persist in orchestrating edge-cloud interactions for dynamic, intelligent decision-making without excessive latency or power consumption [15].

2.5 Towards Integrated and Context-Aware Systems

Advanced frameworks attempt holistic monitoring by fusing multimodal data (acoustic, pressure, consumption) and adapting to context [16]. These represent progress toward smart systems but often remain centralized, vulnerable to connectivity failures, and insufficiently tailored to local infrastructure variations [17].

3.0 Observation

Synthesis of the literature on AI-driven water loss detection reveals several persistent systemic barriers that prevent most solutions from moving beyond prototypes or pilots into reliable, network-wide operational use particularly in high-NRW utilities in developing and middle-income contexts [18].

The Silo Effect and Lack of True System Integration

Research remains strongly compartmentalized: acoustic studies focus on signal classification, pressure research on transient analysis and localization, flow/consumption work on time-series anomaly detection, and IoT/edge efforts on low-power data acquisition [19]. Very few publications demonstrate end-to-end, multi-modal systems that fuse acoustic, pressure, flow, and geospatial data to produce unified, prioritized alerts with estimated leak size, location confidence, and economic impact. The near-total absence of published reference architectures, modular middleware, open interoperability standards, or documented integration patterns hinders utilities from combining best-in-class components into cohesive platforms. This fragmentation leads to duplicated effort, inconsistent performance across data types, and failure to leverage complementary detection strengths (e.g., acoustic for small leaks + pressure for bursts), ultimately blocking holistic, network-wide situational awareness and proactive loss control [20].

Data Quality, Scarcity, and Localization Challenges

Most effective AI models require large volumes of accurately labelled real leak events data that the majority of utilities lack. Operational datasets are noisy, incomplete, imbalanced, and highly context-specific (pipe materials, soil conditions, intermittent supply, illegal connections). Models trained on clean laboratory, simulated, or high-income-country data often perform poorly when transferred to heterogeneous, aging networks common in the Global South. Progress in domain adaptation, few-shot learning, physics-informed models, realistic synthetic data generation, and active labelling remains limited. Critically, there is a severe shortage of open, curated, region-specific benchmark datasets, which prevents fair evaluation, reproducible advances, and development of truly localized, generalizable AI solutions for diverse utility environments.

Connectivity Dependence and Insufficient Edge/Offline Intelligence

Many proposed AI solutions assume continuous, reliable, high-bandwidth connectivity for inference, retraining, and alerting an unrealistic expectation in peri-urban, rural, informal, or aging urban areas with intermittent power and patchy mobile coverage. Cloud-centric designs fail completely during outages, leaving high-loss zones unmonitored precisely when detection is most critical. While some studies explore lightweight edge models or local pre-processing, truly offline-first designs are underdeveloped: ultra-compact architectures for low-power sensors, intelligent caching/prefetching, conflict-free synchronization upon reconnection, and graceful degradation hierarchies receive insufficient attention. Without a fundamental shift toward inherent resilience and autonomy at the edge, AI detection will remain confined to well-connected pilot zones and fail to deliver equitable, network-wide coverage.

Limited Explainability, Trust, and Actionable Decision Support;

Many deep learning models act as near-black boxes, offering high accuracy but little insight into decision rationale. Operators need interpretable, domain-grounded explanations (e.g., wave arrival times, frequency signatures, deviation magnitude) to trust alerts, justify crew dispatch, and defend prioritization. Adoption of tailored explainable AI techniques (SHAP, LIME, attention visualization) adapted to water physics and workflows is growing but still inadequate. Most evaluations stop at detection metrics (precision, recall, localization error) and rarely assess real operational value: false-positive rates in live use, crew efficiency gains, actual NRW volume reduced, repair cost savings, or long-term infrastructure benefits. This gap in explainability and impact measurement erodes trust, slows adoption, and complicates funding justification.

Underdeveloped Longitudinal Impact Assessment and Socio-Economic Evaluation;

The vast majority of studies are lab-based, simulation-driven, or short-term pilots, with very few reporting multi-year, network-wide results under real conditions. Key long-term questions remain unanswered: Do AI platforms deliver sustained, cost-effective NRW reductions? Do they generate verifiable economic returns (energy savings, deferred capex, tariff stability)? Do they improve equity of supply, extend asset life, reduce environmental footprint, or enhance climate resilience? Rigorous longitudinal studies using before-after comparisons, control groups, or full cost-benefit analyses especially in developing-country contexts with different governance, funding, and capacity constraints are notably scarce. Without this broader evidence base, it remains difficult to convince utility boards, regulators, donors, and governments to prioritize and scale AI-based loss management over conventional infrastructure investments

4.0 Figures

Fig. 1. *AI Smart water loss detection*

5.0 CONCLUSION

This review underscores AI's transformative role in water loss detection while highlighting systemic barriers: fragmentation, data limitations, connectivity dependence, and insufficient localization. Individual technologies show maturity, yet integrated deployment lags, particularly in challenging environments.

The proposed conceptual "AI Smart Water Loss Detection" platform merging anomaly detection, multimodal fusion, edge processing, and adaptive learning in a hybrid framework addresses these gaps directly. It shifts from disjointed tools to a cohesive, intelligent system capable of proactive intervention, resource optimization, and enhanced utility resilience. Realizing this vision requires focused efforts on integration, robust edge AI, and localized validation to deliver sustainable water management benefits.

6.0 Recommendations and Future Works

For the Research Community:

For the Research Community:

1. Develop modular integration architectures and open benchmarks for multimodal AI water loss systems

Researchers should prioritize the publication of detailed, replicable system architectures that enable seamless integration of acoustic, pressure, flow, and geospatial data streams. Open benchmarks with standardized evaluation protocols and shared datasets would allow fair comparison of multi-modal approaches and accelerate progress toward unified platforms.

2. Advance lightweight, on-device models through compression and transfer learning for field viability

Focus on model compression techniques (quantization, pruning, knowledge distillation), efficient architectures (Mobile Net-style, TinyML), and transfer/few-shot learning to create compact AI models that run reliably on low-power edge devices commonly deployed in water networks, ensuring functionality in low-connectivity and resource-constrained environments.

3. Create and share region-specific datasets to improve model relevance and fairness

Curate, annotate, and openly release high-quality, geographically and infrastructurally diverse datasets (e.g., intermittent-supply systems, mixed pipe materials, tropical/subtropical conditions) to enable training and validation of models that generalize beyond high-income-country networks and reduce localization bias.

4. Prioritize longitudinal field studies measuring NRW reduction, economic impacts, and sustainability outcomes

Design and conduct multi-year, real-world deployments with rigorous before-after analysis, control groups, and full cost-benefit evaluation to quantify sustained NRW volume reduction, energy savings, repair cost avoidance, equity improvements, and environmental benefits, providing the evidence base needed for scaled adoption.

For Practitioners and Policymakers:

1. Adopt open standards and interoperable platforms to enable flexible, vendor-agnostic deployments

Mandate or incentivize the use of open APIs, standardized data formats (e.g., WaterML, SenML), and modular platforms so utilities can integrate best-available sensors, models, and analytics without vendor lock-in, fostering local innovation and reducing long-term costs.

2. Invest in public digital infrastructure: sensor networks, validated datasets, and shared hydraulic models

Fund and maintain national or regional open-access assets dense reference sensor networks in pilot zones, authoritative geospatial pipe data, verified hydraulic models, and curated historical datasets to dramatically lower entry barriers for AI developers and utilities.

3. Engage stakeholders in co-design to ensure cultural, operational, and contextual fit

Involve local engineers, community water committees, informal settlement representatives, maintenance crews, and small-scale vendors from the earliest design stages to guarantee that solutions address real operational workflows, cultural acceptance, affordability, and practical maintainability.

4. Support complementary investments in connectivity and metering upgrades alongside AI initiatives

Recognize that AI effectiveness depends on foundational infrastructure; therefore, coordinate public-private investments in reliable power, expanded mobile coverage in high-loss zones, district metering, and smart meter roll-out to create the enabling environment for resilient, network-wide AI-based loss detection.

8.0 REFERENCES

[1] H. T. A. K. S. &. A. K. M. AbuEltayef, Addressing non-revenue water as a global problem and its interlinkages with sustainable development goals. Water Practice & Technology, 18(12), 3175–3202., London,UK: IWA Publishing, 2023.

[2] Kingdom, B., Liemberger, R., & Marin, P., The challenge of reducing non-revenue water in developing countries., Washington Dc: World Bank (Water Supply and Sanitation Sector Board)., 2026.

[3] Farouk, A. M., Rahman, R. A., & Romali, N. S, Non-revenue water reduction strategies: A systematic review. Smart and Sustainable Built Environment, 12(1), 181–199., Bingley,UK: Emerald Publishing., 2023.

[4] Ogata, R., Tsutsui, N., Bahige, J. B., & Murakami, S, Cost-effective non-revenue water reduction: Analysis through pilot activities in Kigali City, Rwanda. Water Practice & Technology, 19(10), 4328–4337., London,UK: IWA Publishing, 2024.

[5] O. Ayeni, Edge computing in smart water networks: Reducing costs and improving real-time decision-making., Nigeria: Medium / Independent., 2025.

[6] R. H. H. X. K. &. T. T. Li, A review of methods for burst/leakage detection and location in water distribution systems. Water Supply, 15(3), 429–441., London UK: IWA Publishing, 2015.

[7] Liu, R., Zayed, T., & Xiao, R, A critical systematic review of machine learning based models for water leak detection using vibroacoustic technology. Engineering Applications of Artificial Intelligence, 132, 108434., Amsterdam, Netherlands: Elsevier, 2025.

[8] R. e. a. Vanijjirattikhan, AI acoustic leak detection algorithm using machine learning techniques., Thailand: Conference/Technical Report; often cited in IEEE or similar, 2022.

[9] S. e. a. Rahimi, Frequency convolutional neural network for acoustic leak detection in water pipes., Netherlands: Elsevier, 2020.

[10] R. e. a. Xiao, Deep learning-based acoustic leak detection using time-frequency representations., Netherland: Elsevier, 2021.

[11] B. Sithole, S. Rimer, K. Ouahada, C. Mikeka and J. Pinifolo, Smart water leakage detection and metering device, South Africa: IST-Africa Week Conference, 2016.

[12] A. Salam, Water Leakage Detection Using TinyML, morocco: Othmane Atanane 1, Asmaa Mourhir 1 ORCID ,Nabil Benamar 1,2ORCID andMarco Zennaro 3ORCID, 2023.

[13] E. Farah and I. Shahrour, Smart water for leakage detection: Feedback about the use of automated meter reading technology, Beiriut,: IEEE, 2017.

[14] Thenmozhi, K. Sumathi, A. Asokan, B. Priyanka, R. Maheswar and P. Jayarajan, IoT Based Smart Water Leak Detection System for a Sustainable Future., Chennai, India: IEEE, 2021.

[15] F. Abate, M. Carratù, C. Liguori and A. Pietrosanto, Smart meters and water leakage detection: a preliminary study, Auckland, New Zealand: IEEE, 2019.

[16] B. Sithole, Water Research Commission, "Water loss-SA needs to do more, Durban, South Africa: IEEE, 2016.

[17] C. Liguori and A. Pietrosanto, Towards a Context Awareness System Using IoT, AI, and Big Data Technologies, Francesco Abate; Marco Carratù; : IEEE, 2025.

[18] T. K. Chan, C. S. Chin and X. Zhong, Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection, IEEE, 2012.

[19] M. R. Islam, S. Azam, B. Shanmugam and D. Mathur, A Review on Current Technologies and Future Direction of Water Leakage Detection in Water Distribution Network, IEEE, 2022.

[20] Z. K. a. D. A. S. M. Romano mr277@exeter.ac.uk, Real-Time Leak Detection in Water Distribution Systems, 2010.