

Artificial Intelligence and Data-Driven Leakage Management for Sustainable and Resilient Water Distribution Systems: A Critical Review

Amir Noori

anoor074@uottawa.ca

University of Ottawa

Ehsan Roshani

National Research Council

Hossein Bonakdari

University of Ottawa

Research Article

Keywords: Water Distribution Systems, Leakage Management, Artificial Intelligence, Machine Learning, Urban Water Infrastructure, Infrastructure Resilience

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11089934/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Water leakage threatens the sustainability and resilience of water distribution systems (WDSs) by contributing to treated-water losses, unnecessary energy consumption, service disruption, and deteriorating infrastructure performance. Effective leakage management is therefore essential for resource-efficient and resilient urban water services. This critical review synthesizes statistical and probabilistic data models, classical machine learning, neural and deep learning (DL), physics-informed and hybrid learning, and AI-enabled decision and control for leak detection, localization, prediction, and mitigation. It draws on 41 core comparative studies and 59 contextual sources identified through a transparent and reproducible screening process. Emerging directions include acoustic and multisensor learning, graph neural networks and transformer models, digital twins, physics-informed machine learning (ML), reinforcement learning (RL), and retrieval-augmented large language model decision support, reflecting a transition toward connected and intelligent water infrastructure. A two-axis taxonomy separates analytical method families from leakage-management tasks, while dataset provenance and class-imbalance-aware evaluation guide interpretation of reported performance. No analytical family is uniformly superior: performance depends on sensing modality, leak characteristics, class imbalance, domain shift, uncertainty, and validation design, and often declines from simulation to laboratory or field conditions. Persistent gaps include limited field validation, inconsistent reporting, insufficient imbalance-aware evaluation, incomplete integration with utility information systems, and limited research on emerging generative and agentic artificial intelligence (AI). The resulting decision framework links utility objectives with data, models, metrics, and validation requirements. By connecting leakage analytics with water conservation, operational efficiency, infrastructure resilience, and intelligent decision support, this review supports more sustainable and reliable urban WDSs.

1. Introduction

Water distribution systems (WDSs) are critical urban infrastructure whose sustainability and resilience depend partly on limiting avoidable losses. Leakage and pipe bursts waste treated water, reduce pressure, increase treatment and pumping costs, and may increase contamination risk (Fontanazza et al. 2015). Liemberger and Wyatt (2019) estimated global losses of about 126 billion m³/year, while reports found average non-revenue water of 39% among the three largest suppliers in 56 reporting countries (WHO and UNICEF, 2026). Leakage management is therefore directly linked to water conservation, energy efficiency, service reliability, and infrastructure resilience.

Leakage remains difficult to detect because signals may be small relative to demand variability and most pipes are buried (Yu et al. 2021). Physical methods use acoustic devices, hydrophones, vibration sensors, ground-penetrating radar, tracers, or robotic and non-destructive inspection, but require equipment and field access (Rajeswaran et al. 2018). Computational approaches include statistical and probabilistic data models, AI and machine-learning models, hydraulic model-based methods, and hybrid physics–data approaches. Increasing availability of supervisory control and data acquisition (SCADA), advanced metering infrastructure (AMI), Internet of Things (IoT) sensing, Geographic Information

System (GIS), and digital twins has expanded their role from isolated detection toward integrated monitoring, prediction, localization, and decision support.

Hydraulic model-based approaches use conservation laws and calibrated network models to simulate normal and leakage states and infer leak location or magnitude from measurement–model discrepancies (Almandoz et al. 2005; Young and Tych 1996). Data-driven approaches learn patterns from observed or simulated data, whereas hybrid approaches combine hydraulic constraints, topology, or simulation with statistical or AI models. High predictive accuracy alone does not ensure operational value; models must also be robust to field noise, transferable across networks and time periods, computationally practical, interpretable where required, and compatible with utility systems. Here, localization denotes estimation of the position, pipe, node, or search zone associated with a leak. This review has four objectives: (1) organize leakage analytics using a two-axis taxonomy separating analytical methods from operational tasks; (2) evaluate how dataset provenance, sensing modality, and preprocessing affect performance and transferability; (3) compare metrics and validation practices, particularly for class imbalance and field generalization; and (4) translate the findings into a utility-oriented decision framework linking data, models, and performance criteria with water conservation, operational efficiency, infrastructure resilience, and sustainable urban water management. Figure 1 summarizes the taxonomy.

2. Research methodology

This structured critical review synthesizes representative peer-reviewed studies without claiming full compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). It covers statistical and probabilistic models, classical machine learning (ML), neural and deep learning (DL), physics-informed and hybrid learning, and AI-enabled decision and control for WDS leak detection, localization, prediction, and mitigation. Five questions guide the synthesis: (RQ₁) How have these approaches evolved? (RQ₂) Which sensing and data regimes are used, and how comparable are they? (RQ₃) Which models suit specific tasks and constraints? (RQ₄) Which metrics and validation designs enable credible comparison under class imbalance? (RQ₅) Which gaps hinder transfer from simulations and testbeds to utility-scale deployment?

2.1. Search sources and study selection

The search covered Google Scholar, IEEE Xplore, Elsevier's ScienceDirect, an institutional library discovery service, and backward and forward citation chasing. Up to 200 results per source were retained for screening. Because coverage overlapped, raw hit totals were not treated as unique records; Fig. 2 reports the search and screening flow.

The sources returned 15,962 raw hits; 914 were retained under the cap. Deduplication removed 191, leaving 723 records, and initial screening excluded 125. Of the remaining 598, 31 were unavailable and 293 were excluded during title and abstract screening, leaving 274. A further relevance screen excluded

153; 121 underwent full-text assessment, where 21 were excluded. The final corpus contained 41 core comparative studies and 59 contextual sources.

Eligible studies addressed leakage in potable WDSs; examined statistical or probabilistic models, machine or DL, physics-informed or hybrid learning, or AI-enabled decision and control; provided sufficient methodological detail; and were peer-reviewed English-language journal articles or full conference papers published during 2010–2026. Earlier foundational studies were retained when necessary. Exclusions covered duplicates, non-water applications, studies without a leakage-related output, insufficiently documented or non-peer-reviewed records, unavailable full texts, and conventional optimization-only studies. Core studies supported comparative synthesis; contextual sources provided background and evidence on emerging directions.

2.2. Keyword co-occurrence and thematic clusters

Keyword co-occurrence analysis maps recurring concepts and relationships within a literature corpus (van Eck and Waltman, 2010; Su and Lee, 2010). In Fig. 3, node size represents term frequency, edge thickness represents co-occurrence strength, and degree centrality indicates direct connections. The clusters describe the broader literature landscape rather than the analytical taxonomy adopted in this review. Optimization-related terms may therefore appear in the keyword map even though conventional optimization-only methods are outside the final review scope.

2.3. Datasets and experimental settings

Leakage-model performance depends strongly on dataset provenance. Simulation and benchmark resources, including LeakDB, L-Town/BattLeDIM, C-Town, and Net3, often implemented through EPANET or WNTR, provide complete labels and controllable scenarios but inherit assumptions about demand, roughness, sensor noise, leak hydraulics, and calibration. Laboratory and pilot systems provide measured hydraulic, acoustic, or vibration signals under controlled conditions but cover limited pipe materials, diameters, and operating regimes. Field and utility datasets based on SCADA, AMI, acoustic monitoring, or excavation-confirmed events provide stronger deployment evidence but are usually sparse, imbalanced, noisy, and difficult to share.

Simulation-based evidence remains dominant (Zuñiga-Urbe et al., 2026). Basnet et al. (2023), Rajabi et al. (2023), and Ranganathan et al. (2026) used hydraulic simulations; Noori et al. (2026) combined pressure-driven simulation with infrastructure screening and GIS variables in a large WDS and is therefore classified as hybrid simulation-based. Joseph et al. (2025) used a laboratory network, whereas Sitaropoulos et al. (2023), Fares et al. (2023), Xu et al. (2025), and Glynis et al. (2023) used field or utility data. Wu et al. (2023b) applied a utility-scale digital twin, while Rajan and Li (2025) emphasized high-frequency pressure and flow acquisition and system integration. Performance should therefore be interpreted in relation to data origin, leak prevalence and magnitude, sensor density, train–test separation, and evaluation-environment independence.

Dataset availability also affects reproducibility. Public benchmarks support repeatable comparison, whereas many field datasets remain inaccessible. Oberascher et al. (2026) identified substantial under-reporting of preprocessing, hyperparameter optimization, explainability, and uncertainty in supervised urban-water ML studies. Accordingly, this review distinguishes simulation, laboratory, and field evidence, reports study-specific limitations where available and does not treat their performance values as directly comparable.

3. AI and Data Driven Leakage Analytics

A data-driven leakage model derives predictive or diagnostic relationships from measured or simulated observations rather than relying exclusively on calibrated hydraulic equations. Relevant information may include hydraulic residuals, temporal trends, acoustic and vibration spectra, network topology, asset characteristics, environmental variables, and learned representations. This review organizes the analytical approaches into five broad families: statistical and probabilistic data models; classical ML; neural, deep, and representation learning; physics-informed and hybrid learning; and AI-enabled decision and control. Figure 1 maps these families against the principal leakage-management tasks of detection, localization, severity estimation, prediction and early warning, and mitigation and control. Conventional mathematical and metaheuristic optimization is not classified as an AI or data-driven model family, although selected studies may be discussed contextually when they clarify how analytical outputs support management decisions.

3.1. Taxonomy and Artificial Intelligence Context

Machine-learning is a subset of AI, and DL is a subset of machine learning. Leakage analytics also uses statistical and probabilistic models that are data-driven but are not necessarily AI, as well as hydraulic models derived principally from physical equations. Classical ML methods include support vector machine (SVM), k-nearest neighbour (KNN), decision trees (DT), random forests (RF), boosting, clustering, and one-class models. Neural and representation-learning methods include artificial neural networks (ANN), extreme learning machine (ELM), convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, contrastive learning, graphical neural networks (GNN), and Transformers. Physics-informed and hybrid approaches incorporate hydraulic equations, calibrated model residuals, network topology, or model-data assimilation into statistical or learning-based analysis. AI-enabled decision and control includes methods such as reinforcement learning (RL) and deep reinforcement learning (DRL) that learn operational policies from data or interactions with an environment. These categories describe each method's dominant role; individual studies may combine methods from several families.

Recent reviews show rapid AI development alongside persistent deployment gaps. Rajan and Li (2025), reviewing 71 flow-data studies, found that 58% addressed only detection and few integrated acquisition, alert, and interface functions. Zuñiga-Urbe et al. (2026) synthesized 53 studies from 2018–2025, identifying pressure as the most common hydraulic input, widespread simulation dependence,

competitive SVM performance for small and noisy datasets, frequent CNN use for multiclass localization tasks, and increasing hybridization. Javed et al. (2025) found DRL increasingly useful for pressure management and pump and valve control but still immature for leak-specific detection and prevention. Farah and Shahrouz (2024) and Serafeim et al. (2024) similarly treat leakage management as detection, estimation, and mitigation rather than a single classification task.

Joseph et al. (2023) reviewed software- and hardware-based leak and burst technologies, while Nimri et al. (2023) and Romero-Ben et al. (2023) compared data-driven and model-based localization. Negm et al. (2023a) contrasted hardware and software, intrusive and non-intrusive, steady and transient, and single and hybrid methods, concluding that no economical technique reliably captures both background leakage and bursts. These reviews connect earlier algorithm-centred research with multisensor fusion, digital twins, hybrid learning, and AI-enabled control. Oberascher et al. (2026), reviewing more than 100 highly cited supervised-ML applications in urban water infrastructure, found that only about one third reported feature scaling or automatic hyperparameter optimization and fewer than one quarter reported explainability or uncertainty, reinforcing the need for transparent preprocessing, tuning, and uncertainty reporting.

3.2. Evolution of AI and Data Driven Model Families

Machine-learning models learn parameters, representations, or decision rules from data. Classical methods include SVM, KNN, tree-based models, boosting, clustering, and one-class learning, whereas neural approaches range from shallow ANN and ELM models to CNN, RNN, GNN, and Transformer architectures. Supervised, semi-supervised, self-supervised, and unsupervised learning describe how these methods use labels rather than defining mutually exclusive model families (El-Abbasy et al., 2016; Butterfield et al., 2018; Cody and Narasimhan, 2020; Guo et al., 2020; Fares et al., 2023; Noori et al., 2026). Statistical forecasting, bayesian inference, change detection, and process-control methods remain useful when their assumptions are appropriate. Physics-informed and hybrid approaches incorporate hydraulic knowledge or network structure, while AI-enabled control methods learn actions or operational policies. Figure 4 summarizes the occurrence of these approaches in the reviewed leakage literature.

The expanded 2023 coverage corrects the apparent decline caused by incomplete coverage of that year. Studies published from 2024 to 2026 extend established classical ML and neural-learning models through contrastive learning (R. Liu et al. 2024b), physics-informed transient leakage prediction (Garzón-Orduña et al. 2026), unified benchmarking of ML, DL, GNN, and Transformer models (Ranganathan et al. 2026), and retrieval-augmented large language model (LLM) decision support (Liu et al. 2026). Because individual studies may evaluate multiple approaches, the bars in Fig. 4 represent methodological occurrences rather than publication counts. Wang et al. (2026) is excluded because it presents a simulation and analysis architecture rather than a directly validated leakage-detection model.

The suitability of an analytical method depends on the type, quality, and availability of its input data. Figure 5 maps the dominant input modalities to the statistical and probabilistic, classical ML, neural and

representation-learning, and physics-informed and hybrid approaches used for leakage detection, prediction, and localization. Studies published from 2023 to 2026 expand the use of field acoustic and vibration sensing, multisensor hydraulic monitoring, digital twins, and hybrid learning. AI-enabled control studies are discussed separately because their inputs include system states, actions, rewards, and operational constraints rather than only leakage observations. Figure 5 characterizes the studies reviewed in this paper and is not intended as a complete bibliometric census.

The reviewed leakage-analytics literature includes statistical and probabilistic models, classical ML, neural and representation learning, and physics-informed and hybrid approaches. These families differ in their dependence on labelled events, ability to represent temporal and spatial relationships, use of hydraulic knowledge, computational requirements, and interpretability. Their suitability therefore depends on the sensing modality, data volume, label availability, network topology, and operational task. The following sections evaluate each family in that context rather than ranking models by isolated accuracy values.

3.3. Statistical and Probabilistic Data Models

Statistical and probabilistic data models detect deviations, forecast expected behaviour, or estimate leakage risk without requiring complex machine-learning architectures (Wu and Liu 2017; Chan et al. 2018; Palau et al. 2012; Jung and Lansey 2015). Relevant approaches include regression, time-series forecasting, change detection, forecast-residual analysis, Gaussian mixture models, Bayesian dynamic models, and statistical process control. Statistical process control (SPC) applies thresholds or control limits to pressure and flow measurements or residuals using techniques such as Western Electric rules, Hotelling's T^2 , cumulative sum (CUSUM), and exponentially weighted moving average (EWMA). These approaches are generally transparent and computationally efficient but remain sensitive to nonstationarity, threshold selection, unexpected demand, and distributional assumptions.

Probabilistic approaches explicitly represent uncertainty. Hutton and Kapelan (2015) compared measurements with predictive distributions, Jung and Lansey (2015) used nonlinear kalman filtering (KF)-based state estimation, and Karray et al. (2016) applied predictive KF for data compression. Wang et al. (2020) used long short-term memory (LSTM) flow prediction but evaluated only a few months and limited seasonal variability, indicating the need for broader validation.

Statistical analysis is often combined with dimensionality reduction, signal transformation, or clustering. Palau et al. (2012) used principal component analysis to reduce redundancy in flow data, while Loureiro et al. (2016) applied robust statistics to asymmetric flow patterns. Quiñones-Grueiro et al. (2018) combined wavelet transformation with preprocessing based on network dynamics, and Nam et al. (2019) combined principal component analysis, K-means clustering, and sensitivity analysis to support sensor configuration. Such unsupervised methods do not require labelled leakage histories, but they still require representative operational data and may interpret unusual demand or sensor behaviour as leakage.

Statistical methods also have application-specific limitations. Loureiro et al. (2016) did not explicitly account for unexpected demand or large individual users, the approach of Quiñones-Grueiro et al. (2018) can require substantial detection time, and Nam et al. (2019) did not fully evaluate sensitivity to measurement noise. SPC is therefore often used as one component of a broader workflow, including residual screening, preprocessing, or alert generation (Wu and Liu 2017).

Recent studies demonstrate that classical statistical methods remain competitive for online applications. Henriques-Silva et al. (2023) used a Bayesian dynamic linear model (DLM) to quantify forecast uncertainty, accommodate missing data and adjust the detection–false-alarm trade-off. Wan et al. (2023) combined EWMA and robust Tukey statistics for gradual leakage, while Barrientos-Torres et al. (2023) used seasonal autoregressive integrated moving average (ARIMA) as the expected-flow model before anomaly screening. These methods are transparent and generally easier to deploy than large deep networks, but unusual demand or operational changes may resemble leakage unless the baseline and thresholds adapt appropriately.

3.3.1. Kalman Filtering

Kalman filtering (KF) is a recursive state-estimation method rather than a ML algorithm. It combines a dynamic state model with noisy observations to update estimates of unobserved hydraulic conditions. In the reviewed WDS applications, it is therefore treated as model-data assimilation within the physics-informed and hybrid family. Ye and Fenner (2011) used KF residuals derived from pressure and flow measurements for automatic burst detection and leakage estimation. The method was computationally efficient and responsive to sudden events but less effective for slowly developing leaks. Kang and Lansey (2009) found a detection state estimator more reliable in their comparison, while Okeya et al. (2014) reported stronger performance from flow-based than pressure-based residual indices. KF also appears in hybrid forecasting systems such as McMillan et al. (2023). Its performance depends on the validity of the state-transition and noise assumptions, and gradual leakage may be absorbed into an adapting baseline.

3.4. Classical Machine Learning

3.4.1. Support vector machines

Support vector machines (SVMs) are supervised margin-based learners used for classification and regression. They perform well with moderate-dimensional and relatively small datasets and nonlinear kernels but are sensitive to scaling, hyperparameters, large training sets, and class imbalance. WDS studies have used pressure, flow, and acoustic features for detection, size estimation, and localization (Mounce et al. 2011; Mamo 2014; Zhang et al. 2016; Cody et al. 2017; Xu et al. 2021). Mashford et al. (2009, 2012) used EPANET pressure scenarios for leak size and location inference, Cody et al. (2017) reported 97% detection accuracy from laboratory acoustic features, and Zhang et al. (2016) combined hydraulic simulation, k-means zoning, and multiclass SVM for leak-zone identification. Xu et al. (2021)

combined variational mode decomposition and mel-frequency cepstral coefficient (MFCC) acoustic features with SVM, achieving 93% classification accuracy.

Reported SVM performance varies substantially across datasets and sensing conditions. Alves Coelho et al. (2020) compared SVM, RF, DT, and neural-network models using IoT flow data. SVM achieved at most 57% accuracy, whereas RF reached approximately 75%. Lin et al. (2020), however, reported 92% SVM accuracy using a similar flow-monitoring framework. These differences likely reflect variation in data representation, noise, hyperparameter selection, class balance, and validation design rather than an inherent performance limitation of SVM. The method remains sensitive to feature scaling, kernel and regularization choices, large training sets, and strongly imbalanced classes.

McMillan et al. (2024) combined a domain-informed variational representation with a binary SVM so that leak and non-leak flow sequences were better separated in latent space. Komba et al. (2024) paired SVM with RF in an IoT architecture for WDS leak detection and localization, while Xie et al. (2025) coupled variational mode decomposition, wavelet denoising and acoustic feature extraction with SVM classification in a database-centered monitoring framework. Siddique et al. (2024) used continuous wavelet transform (CWT) scalograms, a deep belief network whose hyperparameters were optimized using a Genetic algorithm (GA) and least-squares support vector machine (LSSVM) in a more complex hybrid pipeline. In a campus WDS, Onukwube et al. (2024) reported 79% SVM accuracy versus 70% for KNN and 61% for ANN, while Joseph et al. (2025) obtained 0.94 accuracy in a controlled multiclass experiment. Sourabh et al. (2023) found multilayer perceptron (MLP)-ANN consistently stronger than SVM for pressure- and flow-based classification and regression. Yussif et al. (2023) reported 82.50% SVM localization accuracy versus 83.75% for ensemble KNN, and Choi and Im (2023) obtained F1-Score = 80.99% for SVM versus 94.82% for CNN. SVM therefore remains a useful baseline, but its relative performance depends strongly on representation, sensing conditions, class structure, and validation design.

3.4.2. K-nearest neighbours

The k-nearest neighbour (KNN) algorithm is a supervised, instance-based learning method used for classification and regression. Unlike parametric models, KNN primarily stores and preprocesses labelled training instances rather than estimating an explicit predictive equation. A new observation is assigned an output from its nearest training instances. Rashid et al. (2015) used negative-pressure-wave measurements for binary detection (78.51% accuracy) and five-class leak-size estimation (90.1%). Fereidooni et al. (2021) compared KNN with RF and Bayesian networks on flow data and found KNN more sensitive to the number of leak examples in the training set. Pham Vu Hong and Nguyen Thanh (2023) used the LeakDB benchmark to compare MLP, logistic regression (LR), KNN and SVM across simulated networks. These results illustrate KNN's advantages, simple decision logic and flexible nonlinear boundaries, but also its sensitivity to scaling, irrelevant dimensions, local class density and computational cost at inference when the stored dataset becomes very large.

Recent studies show that KNN can be very strong when class boundaries are clean but substantially weaker in more heterogeneous networks. Onukwube et al. (2024) reported 70% accuracy for KNN in joint detection and localization, while Chandanwala et al. (2025) retained KNN as one of several classical baselines for acoustic MFCC-based distance classification. In the controlled Melton experimental system, Joseph et al. (2025) reported test-set precision, recall, F1-Score, and accuracy of 1.00 for KNN; importantly, the authors explicitly cautioned that the balanced, low-noise experimental design may overstate generalization to field data. Yussif et al. (2023) is particularly relevant to low-cost localization because the ensemble KNN model slightly exceeded SVM for leak-position prediction (83.75% versus 82.50% validation accuracy), while KNN direction classification reached 97.50% compared with 78.75% for a DT. This reinforces that KNN can be highly effective when acoustic features are carefully denoised and the output space is constrained, even though later studies show weaker performance in more heterogeneous hydraulic localization problems. Overall, KNN can be effective when feature scaling and local class structure are favorable, but noise, dimensionality and sparse leak examples can increase false alarms or reduce transferability. Table 1 summarizes representative findings and algorithm-specific limitations.

Table 1
Summary of KNN-based leak detection studies

Reference	Approach	Key Findings	Study/data limitations
Rashid et al. (2015)	KNN using measured negative pressure wave (NPW) data for binary (leak vs. no-leak) and multi-class leak size estimation	78.51% accuracy (binary); 90.1% accuracy (multi-class)	Data-quality dependent; large-scale and real-time validation needed
Fereidooni et al. (2021)	Compared KNN, RF, and Bayesian networks for leak detection using flow data	Best results with Naive Bayesian and RF; KNN performance sensitive to the number of leaks in training	Sensitive to limited and rare leak examples
Ravichandran et al. (2021)	Compared KNN with gradient boosting trees (GBT) for binary classification of acoustic signal features	GBT significantly reduced false positives relative to KNN	Higher KNN false alarms; limited real-time validation
Kammoun et al. (2022)	Surveyed AI-based leak detection methods in WDSs; compared MLP, LR, KNN, and SVM using the LeakDB benchmark	KNN achieved high detection rate but also produced more false alarms; SVM offered a better balance (accuracy vs. false alarms). Seasonal or nighttime data improved performance, especially with pressure signals	Higher false positives; field testing under demand variability needed
Pham Vu Hong and Nguyen Thanh (2023)	Assessed MLP, LR, KNN, and SVM (applied to the LeakDB benchmark) for improving leak detection in WDSs	Demonstrated varying performance among AI algorithms; KNN sensitive to noisy data	Scalability and real-time robustness need further study
Yussif et al. (2023)	Ensemble KNN regression for acoustic leak localization and KNN classification for leak direction	83.75% validation accuracy for localization; 97.50% for direction classification	Single-noise-logger performance depends on acoustic propagation and noise
Onukwube et al. (2024)	SVM, KNN and ANN for binary detection and two-zone localization	KNN accuracy 70%; SVM was higher at 79%	Moderate accuracy; performance is network and data dependent

Reference	Approach	Key Findings	Study/data limitations
Chandanwala et al. (2025)	KNN among ML baselines using acoustic MFCC features for leak-distance classification	Used in a broad ML and DL benchmark; XGBoost was best classical model	Acoustic benchmark; KNN-specific accuracy not reported in supplied review
Joseph et al. (2025)	KNN for three-class no-leak, minor-leak, and major-leak classification using pressure and flow features	Precision, recall, F1-Score and accuracy all 1.00 in controlled experiment	Overfitting and class-balance risk; field validation needed

3.4.3. Tree Ensembles and Boosting

Tree-based models and ensembles are widely used for tabular hydraulic, acoustic, and infrastructure data because they capture nonlinear relationships, require limited feature scaling, and can provide measures of variable importance. Individual DTs are readily interpretable but can overfit, whereas RF and boosting methods improve stability or predictive performance by combining multiple trees. Alves Coelho et al. (2020) found RF stronger than SVM in their IoT flow-data experiments, and Fereidooni et al. (2021) found RF competitive with Bayesian and KNN models. Marvin et al. (2023) used RF and histogram-based gradient boosting to estimate leakage location and event parameters under uncertain simulated pressure conditions. Xu et al. (2024) reported strong results from RF and extreme gradient boosting (XGBoost) after acoustic feature selection, while Chandanwala et al. (2025) identified XGBoost as the strongest classical model in their acoustic benchmark. Ranganathan et al. (2026) likewise found RF, XGBoost, and light gradient boosting machine (LightGBM) highly competitive for binary detection, although top-1 localization remained more difficult. These methods remain vulnerable to simulation-specific correlations, class imbalance, overfitting, and poor probability calibration when validation data do not represent operational conditions.

3.4.4. Clustering and One Class Learning

Clustering, one-class learning, and unsupervised pattern analysis identify structure or anomalies when confirmed leakage labels are limited. Classical approaches compare observations with clusters, distances, densities, or patterns derived from normal or mixed operational data. Deep representation methods may also use clustering to interpret learned features. Xu et al. (2025), for example, clustered multi-channel gradient-weighted class activation mapping (MGrad-CAM) frequency-significance patterns from acoustic CNN outputs and related the resulting groups to pressure, leak flow, and sensor distance. Leite et al. (2024) instead used dynamic shape similarity directly for online flow-based burst detection. Clustering can therefore support either anomaly detection or interpretation, but an anomalous cluster does not necessarily represent leakage.

Wu et al. (2016) used cosine distance to calculate the dissimilarity between vectors combined with the information from different sensors. Aksela et al. (2009) proposed a method based on the SOM to detect leakage by finding similarities between flow data from other weeks, facilitated by a leak function that describes the relationship between the confidence in the existence of a leak and the distance between flow meters and leak localizations. Wu and Liu (2020) proposed a Shape Similarity-Based (SSB) method that detected bursts by analyzing the shape of flow time series data within the same period from different days. Huang et al. (2018) applied dynamic time warping (DTW) to study the similarity of daily water demand and found the most unusual daily pattern.

The performance of clustering methods depends strongly on the reference library. Weekday, weekend and holiday variability may require extensive historical data, while discord-discovery methods (Wu and Liu, 2017; and Huang et al., 2018) may require event-duration assumptions that are not known in advance. These methods are therefore generally more effective for abrupt bursts than slowly developing leakage, which may preserve the overall shape of the monitored signal.

3.5. Neural Deep and Representation Learning

3.5.1. Artificial Neural Networks and Extreme Learning Machines

Artificial neural network (ANN) learns nonlinear mappings through interconnected weighted units. WDS applications include MLPs, probabilistic and radial-basis networks, and ELM variants, while deeper convolutional and recurrent architectures are discussed separately. Early studies combined pressure and flow data with MLP, probabilistic neural network (PNN), and radial basis function (RBF) models (Caputo and Pelagagge 2002) and fuzzy inference (Mounce et al. 2010). Bohorquez et al. (2020) reported localization error below 3 m and leak-size error of about 0.31 mm for a defined pipe configuration, while Shekofteh et al. (2020) progressively narrowed leaking subnetworks using generated pressure data. Recent studies extend ANN applications but retain concerns about data demand, tuning, overfitting, and field transfer. M. Liu et al. (2024) proposed Gaussian ELM (GELM) with reported accuracies reaching 100% in some tests but cautioning against extrapolation to larger urban networks. Mohan Doss et al. (2024) evaluated encoder–decoder networks on BattLeDIM, obtaining near-immediate detection for some abrupt leaks but longer delays for incipient events. Joseph et al. (2025) reported 0.96 MLP accuracy in a controlled multiclass experiment. These results preserve the central concerns of data quality, computational cost and field transfer.

In 2023, Sourabh et al. directly compared MLP-ANN with SVM for both classification and regression under pressure-only and flow-only scenarios and found ANN superior in both settings. Basnet et al. (2023) likewise compared an MLP with a CNN across four increasingly complex L-Town leak-localization cases incorporating realistic leak characteristics and model uncertainty; both models performed well, but the CNN consistently outperformed MLP because it could exploit spatial relationships among hydraulic measurements. These findings support the use of ANN as a flexible baseline while also

illustrating why spatial or convolutional structure becomes increasingly important as localization complexity grows. ANN performance generally improves with representative data but can degrade under small, noisy or shifted datasets; larger architectures also increase tuning and computation. Table 2 therefore emphasizes both reported performance and architecture-specific limitations rather than treating all neural networks as equivalent.

Table 2
Summary of neural-network-based leakage studies

Reference	Approach	Key Findings	Study/data limitations
Feng and Zhang (2006)	Fuzzy neural network model for pipeline leak fault detection	Improved robustness over traditional methods	Sensitive to fuzzy parameters; computationally complex
Aksela et al. (2009)	Combined leak function with Self-Organizing Map (SOM)	Adaptive approach accounting for detection confidence over time	Sensor-placement dependent; may miss small and slow leaks
Mounce et al. (2010)	Combined ANN for short-term forecasting with fuzzy inference systems	Enabled online leak and burst detection in near real time	Dependency on accurate flow data; complex tuning of combined systems
Tao et al. (2014)	Proposed burst detection based on an artificial immune network principle	Adapted immune concepts to identify anomalies in WDSs	Needs domain tuning; limited field validation
Zhou et al. (2019b)	Hybrid Deep Neural Network (DNN) and CNN to generate dual event indicators for leak detection	Improved detection performance without high computational cost	Requires careful handling of transient wave samples; potential overfitting
Bohorquez et al. (2020)	Utilized ANN with transient pressure wave data for leak location	Leak localization error < 3.0 m; leak size error about 0.31 mm	Focused on single-pipe topology; scalability to larger WDSs not explored
Shekofteh et al. (2020)	Iterative ANN-based segmentation of WDS to isolate leaks in sub-networks	Enhanced localization by progressively narrowing down the leak region	Multiple iterations needed; heavily reliant on data quality
Fares et al. (2023)	ANN and DL within a 10-model acoustic field benchmark	Strong generalization on new, unlabeled field cases in Hong Kong	Pipe material and ambient noise strongly influence acoustic features
Sourabh et al. (2023)	MLP-ANN classification and regression using pressure-only and flow-only scenarios	ANN outperformed SVM in both classification and regression; input optimization improved performance	Primarily modeled and single-leak conditions; field transfer needed
M. Liu et al. (2024)	Improved GELM	Accuracy reached 100% in some reported tests	May not transfer unchanged to larger urban WDSs
Mohan Doss et al. (2024)	Encoder–decoder neural network on BattLeDIM	Near-instantaneous detection for some abrupt leaks	Incipient leaks delayed; high training-data and processing demands

Reference	Approach	Key Findings	Study/data limitations
Joseph et al. (2025)	MLP within a 14-model multiclass benchmark	0.96 overall accuracy	Controlled, balanced, low-noise data

3.5.2. Convolutional Neural Network

Convolutional neural networks (CNNs) provide translation-equivariant feature extraction from locally structured inputs. In WDSs, researchers often treat sensor measurements (e.g., vibration, pressure, flow, or acoustic signals) as one-dimensional (1D) or two-dimensional (2D) inputs to CNNs. The convolutional layers automatically learn local similarities such as abrupt increases or decreases in flow or pressure, while pooling layers reduce computational complexity compared to fully connected networks. CNNs studies demonstrate strong feature extraction from acoustic, vibration and hydraulic representations. Rahimi et al. (2020) used a 1D CNN on acoustic-emission signals (over 75% accuracy), Hu et al. (2021) combined multiscale convolution with density-based spatial clustering of applications with noise (DBSCAN) for leak-zone detection, Kang et al. (2017) coupled 1D CNN and SVM (99.3%), Chuang et al. (2019) used 2D MFCC inputs (99.4%), and Cody et al. (2020b) combined a variational autoencoder (VAE) with 2D CNN for hydroacoustic anomaly detection (97.2%). Chen et al. (2020) further demonstrated a physics-guided CNN using satellite-derived environmental variables. These values are not directly comparable because sensing, labels and validation regimes differ.

Key 2023 studies established this trajectory: Choudhary et al. (2023) reported 94.32% detection, 91.91% size-estimation and 89.85% localization accuracy with a 1-D CNN on five-minute multisensor windows; Choi and Im (2023) obtained F1-Score = 94.82% and matthews correlation coefficient (MCC) = 94.47% from vibration spectra; Basnet et al. (2023) found CNN consistently stronger than MLP as L-Town localization scenarios became more realistic; and Jun and Lansey (2023b) coupled a 2-D CNN with hydraulic pressure-response surfaces to exploit spatial structure. Rajabi et al. (2023) extended image-based leakage analysis through conditional deep convolutional generative adversarial networks, transforming hydraulic demand and pressure information into image representations for leak detection and localization. The study illustrates how generative models can exploit spatial hydraulic structure, although performance remains dependent on the fidelity of the underlying hydraulic simulations and image-generation process.

In recent years CNNs research shifted toward field deployment, multisensor fusion and richer outputs. Gunatilake and Miro (2024) developed an M-TFCNN for live multiclass classification of vibro-acoustic field signals and reported 98% and 95% accuracy for HWM and Von Roll sensor data, respectively. Rajasekaran and Kothandaraman (2024) reported 99.08% accuracy combined detection and localization with a 1D DenseNet. Chandanwala et al. (2025) reported 95.98% region-based convolutional neural network (RCNN) test accuracy (with XGBoost at 98.75% among classical baselines), and Satterlee et al. (2025) combined CNN-based few-shot learning with parallel multisensor fusion, reporting 97.1% detection and 95.5–97.4% localization accuracy depending on sensor combination.

Xu et al. (2025) added interpretability by combining a four-channel acoustic CNN with MGrad-CAM, power spectral density and k-means analysis. The model reached 95.40% field accuracy and 93.61% F1-Score (91.92% accuracy for the FCNN benchmark) and remained above 90% accuracy at -10 dB signal-to-noise ratio (SNR) in laboratory tests. Interpretation separated high-frequency broadband from low-frequency narrowband leak patterns and related informative frequency content to pressure, leak flow and monitoring distance. Related acoustic studies include the integrated generative approach of R. Liu et al. (2024a), the data-centric DL framework of Wu et al. (2024), and the time-delay-based DL localization of Liu et al. (2025), collectively demonstrating the move from simple detection toward robustness, interpretation and localization. Table 3 summarizes key studies using CNNs for WDS leak detection.

Table 3
Summary of CNN-based leakage-detection and localization studies

Reference	Approach	Key Findings	Study/data limitations
Hu et al. (2021)	MFCN with DBSCAN clustering	Efficient leak zone detection using pressure and flow data	Potential issues scaling to complex or large networks
Kang et al. (2017)	1D CNN and SVM for classifying vibration data	99.3% accuracy; robust leak identification	Requires substantial labeled data; vibration-based only
Chuang et al. (2019)	2D CNN on MFCCs (plus first and second-order differences) of acoustic signals	99.4% classification accuracy; CNN outperformed KNN and SVM	Applicable primarily to acoustic sensor systems
Rahimi et al. (2020)	1D CNN on acoustic emission signals (FFT-based features)	Over 75% accuracy for detecting leaks in a tank	Reliance on acoustic data quality; limited real-world scope
Cody et al. (2020b)	2D CNN and VAE analyzing hydroacoustic signals in a test bed	97.2% accuracy for detecting a 0.25 L/s leak	Laboratory scale; real pipeline conditions may differ
Chen et al. (2020)	Physics-guided CNN on satellite-derived variables (land surface temperature, fractional vegetation cover, soil moisture)	Effective classification of canal sections into leaking vs. non-leaking	Context-specific approach; remote sensing data availability
Choi and Im (2023)	CNN using FFT magnitude spectra of vibration sound	F1-Score 94.82%; MCC 94.47%; exceeded SVM benchmark	Testbed and material transfer needs validation
Choudhary et al. (2023)	1-D CNN using 5-min multisensor smart-grid windows	94.32% detection; 91.91% size; 89.85% location accuracy	Prototype and artificial-leak setting
Basnet et al. (2023)	CNN versus MLP for four L-Town localization and size cases	CNN consistently outperformed MLP as scenario complexity increased	Simulation-based; field scalability untested
Rajabi et al. (2023)	CDCGAN, pix2pix and structural similarity index measure (SSIM) on demand, pressure images	About 70% real-time leak detection; low online cost	Depends on hydraulic model, interpolation
Jun and Lansey (2023b)	2-D CNN on hydraulic pressure-response surfaces	Spatial hydraulic representation improved burst-detection formulation	Requires calibrated model and rich AMI sensing

Reference	Approach	Key Findings	Study/data limitations
Gunatilake and Miro (2024)	M-TFCNN on live vibro-acoustic field data	98% (HWM) and 95% (Von Roll) multiclass accuracy	Sensor/deployment dependent; cross-network validation needed
Rajasekaran and Kothandaraman (2024)	Custom 1D time-series DenseNet using a single acousto-optic sensor	99.08% average detection and localization accuracy	Needs broader field validation
Chandanwala et al. (2025)	MFCC features with CNN, VGG16, RCNN and ML baselines	RCNN test accuracy 95.98%; XGBoost 98.75% among ML	Generalization gap; dataset dependent
Satterlee et al. (2025)	CNN plus few-shot learning with multi-sensor fusion	97.1% detection; 95.5–97.4% localization	Sensor-combination and deployment-cost dependent
Xu et al. (2025)	Four-channel MCNN plus MGrad-CAM on acoustic spectra	95.40% field accuracy; 93.61% F1-Score; robust at low SNR	Single-sensor; broader spatiotemporal validation needed

3.5.3. Recurrent Neural Networks

Recurrent neural networks (RNNs) model sequential data through recurrent hidden states, making them suitable for hydraulic and acoustic time series. Standard RNNs can suffer from vanishing or exploding gradients, whereas LSTM and gated recurrent unit (GRU) architectures use gating mechanisms to improve learning of longer temporal dependencies. Bjerke (2019) applied GRU-based binary leakage detection to LeakDB pressure and flow sequences and found stronger performance with flow data. Wang et al. (2020) used stacked LSTM layers for district metered area (DMA) flow prediction and residual-based burst detection, reporting 99.80% accuracy on simulated data and 99.77% in synthetic experiments; these values should be interpreted in relation to class balance and synthetic-data conditions.

RNN-type models continue to appear mainly in comparative and sequence-oriented studies. Chandanwala et al. (2025) included RNN among the DL models for acoustic MFCC-based leak-distance classification, although RCNN performed best in that comparison. Glynis et al. (2023) trained LSTMs on burst-free flow and pressure data and transferred learned channel weights when new sensors were added, achieving precision up to 98.1% across simulated hydrant bursts and real bursts in 10 UK DMAs. McMillan et al. (2023) developed the FLUIDS framework across more than 2,000 DMAs by combining RNN and LSTM mean-flow forecasting with Kalman-filter residual forecasts for proactive leakage-burst intervention. RNN-type models are therefore useful when temporal dependence is important, although training cost, sequence drift, and changing sensor configurations can constrain deployment.

3.5.4. Deep and Representation Learning

Representation-learning methods reduce dependence on manually designed features by learning compact structures from hydraulic or acoustic data. Autoencoders and VAEs have been used for anomaly detection, dimensionality reduction, and preprocessing before classification, while contrastive and self-supervised methods exploit unlabelled observations to improve representation quality. McMillan et al. (2024) combined a variational representation with SVM classification, and R. Liu et al. (2024b) used contrastive learning for acoustic leakage analysis. Graph-based methods incorporate network connectivity directly: Brahmbhatt et al. (2023) combined a calibrated digital twin with graph convolution for localization, while Ranganathan et al. (2026) compared GNN and Transformer models with classical ML and DL baselines. These architectures can represent spatial, temporal, or topological dependence, but their benefits depend on sensor encoding, data volume, computational resources, and validation under noise and network transfer.

3.6. Cross Cutting Workflows and Leakage Tasks

The following workflows cut across the principal analytical families. Forecast-based anomaly detection may combine statistical, neural, or hybrid models; feature engineering may support either classical or DL; and direct classification and localization may use several model families. These workflows are therefore organized by analytical process or operational task rather than treated as separate model families.

Across leak detection and localization studies, model performance depends strongly on data regime, output definition, and validation design. Simulated hydraulic datasets enable systematic leak scenarios but may overstate field transfer, whereas acoustic and field datasets introduce material, noise, and sensor effects. Multisensor and digital-twin systems provide richer information but increase deployment and calibration requirements. Model comparisons should therefore use matched datasets and task-appropriate metrics rather than isolated accuracy values. Table 4 compares the principal analytical families and their representative model groups in terms of strengths, limitations, and recommended evaluation criteria.

Table 4
Comparison of data-driven and AI-enabled models used in WDS leakage analytics

Analytical family	Representative models	Principal strengths	Main limitations	Representative studies
Statistical and probabilistic models	Time-series forecasting, statistical process control, regression and Bayesian dynamic models	Transparent forecasting, residual monitoring and uncertainty quantification	Sensitive to scaling, distance metrics and high dimensionality.	Palau et al. (2012); Henriques-Silva et al. (2023); Wan et al. (2023); Barrientos-Torres et al. (2023)
Model–data assimilation	Kalman filtering and Bayesian state estimation	Recursive updating and explicit treatment of state uncertainty	Susceptible to overfitting, imbalance and domain shift.	Ye and Fenner (2011); Henriques-Silva et al. (2023); McMillan et al. (2023)
Classical supervised learning	SVM and SVR	Effective in high-dimensional settings and with limited training data	Sensitive to contamination, representation and operating-regime changes.	Mounce et al. (2011); Cody et al. (2017); Xu et al. (2021); Onukwube et al. (2024)
Classical supervised learning	KNN	Simple, interpretable and useful for neighborhood-based localization	Requires careful tuning and offers limited interpretability.	Rashid et al. (2015); Yussif et al. (2023); Onukwube et al. (2024)
Classical supervised learning	DTs, RFs, and boosting	Handles nonlinear relationships, mixed inputs and feature interactions; often provides useful feature importance	Data-intensive and sensitive to noise and sensor placement.	Marvin et al. (2023); Chandanwala et al. (2025); Ranganathan et al. (2026)
Unsupervised and limited-label learning	Clustering, SOM, autoencoder-based anomaly detection and one-class methods	Useful when labelled leak data are scarce; supports anomaly discovery and baseline modelling	Computationally demanding and vulnerable to concept drift.	Aksela et al. (2009); Fan and Yu (2022); Farajzadeh et al. (2024); Leite et al. (2024)
Neural learning	ANN and ELM	Flexible nonlinear modelling and efficient inference	Data-intensive, difficult to interpret and uncertain in transfer.	Bohorquez et al. (2020); M. Liu et al. (2024); Joseph et al. (2025)

Analytical family	Representative models	Principal strengths	Main limitations	Representative studies
Deep spatial and acoustic learning	CNN and acoustic deep-learning models	Learns spatial or spectral signatures directly from images and acoustic signals	Dependent on model fidelity and computationally demanding.	Choi and Im (2023); Xu et al. (2025); Satterlee et al. (2025)
Deep temporal learning	RNN, LSTM and GRU	Captures temporal dependence, gradual changes and multivariate sequences	Challenged by reward design, safety and sim-to-real transfer.	Wang et al. (2020); Glynis et al. (2023); McMillan et al. (2023)
Representation and topology-aware learning	Autoencoders, contrastive learning, GNNs and Transformers	Learns compact representations and can exploit network connectivity and long-range dependencies	Vulnerable to hallucination, poor grounding and unsafe actions.	Brahmbhatt et al. (2023); McMillan et al. (2024); R. Liu et al. (2024b); Ranganathan et al. (2026)
Physics-informed and hybrid learning	Physics-informed neural networks, differentiable hydraulic models and digital twins	Combines physical consistency with the flexibility of data-driven learning	Sensitive to scaling, distance metrics and high dimensionality.	Wu et al. (2023b); Garzón-Orduña et al. (2026); Noori et al. (2026)
AI-enabled decision and control	Q-learning, PPO, RL, DRL, MADDPG and MARL	Learns adaptive pressure, pump and valve-control policies under dynamic conditions	Susceptible to overfitting, imbalance and domain shift.	Negm et al. (2023b); Hu et al. (2023a, 2023b); Javed et al. (2025)
Generative and agentic AI	RAG-LLMs, tool-using agents and multimodal decision support	Integrates heterogeneous evidence and supports operator interaction and explanation	Sensitive to contamination, representation and operating-regime changes.	Liu et al. (2026); Wang et al. (2026)

3.6.1. Prediction–classification workflows

Prediction-classification workflows estimate expected hydraulic behaviour and then evaluate deviations between forecasts and observations. Statistical regression, Bayesian models, ANN, RNN, and hybrid state-estimation methods may generate the forecasts, after which residual thresholds, statistical tests, or classifiers identify potential leakage (Han et al. 2019; Wu and Liu 2017). These workflows can often be trained primarily on normal-operation data, which are more readily available than confirmed leakage events. However, demand nonstationarity, sensor drift, operational changes, and forecasting errors can

produce leak-like residuals. Evaluation should therefore report leak recall, false-alarm frequency, and detection delay in addition to forecast error.

These methods typically predict short-term flow and pressure behavior (Romano et al. 2014; Wang et al. 2020) or short-term water demand (Mounce et al. 2011; Bakker et al. 2014), with anomalous demand creating potential false positives. Recent studies broaden this framework beyond conventional forecast-residual thresholds. Leite et al. (2024) used dynamic shape similarity to compare flow time-series patterns for online burst detection, representing a pattern-recognition alternative to explicit demand forecasting. Joseph et al. (2024) combined SCADA-driven logic and ML in the Sunbury WDS, where rules were effective for straightforward hydraulic anomalies and ML provided adaptation to historical patterns. Parajuli and Shin (2024) extended classification to cyber-physical failure types by training ML models on pressure, flow and tank-level SCADA variables, illustrating that operational leak detectors increasingly need to distinguish leaks from cyberattacks and other physical disturbances.

Recent forecasting research also emphasizes adaptive and probabilistic approaches. Henriques-Silva et al. (2023) applied Bayesian dynamic linear models to two years of Halifax DMA flow data, detecting up to 75% and approximately 71% of historical breaks, with false-alarm rates of 5.15% and approximately 12% depending on threshold selection. Barrientos-Torres et al. (2023) found seasonal ARIMA more accurate than transfer-function forecasting, Wan et al. (2023) combined EWMA with robust Tukey preprocessing for gradual and abrupt leakage, and McMillan et al. (2023) combined LSTM and KF to forecast leakage-related flow probability. These methods remain sensitive to demand shifts, network interactions, and threshold selection.

Alternative methods include adaptive inflow approximation and fault detection (Eliades and Polycarpou, 2012), evolutionary polynomial regression (EPR) (Laucelli et al., 2016), and DTW (Huang et al., 2018). EPR reproduces WDS behavior from online pressure and flow measurements, while DTW detects pattern anomalies and can reduce false positives, although it is more suited to bursts than background leakage.

Bakker et al. (2013) developed an adaptive demand-forecasting model that generated 48-hr predictions from measured demand and calendar information and updated its database continuously during operation. Reported average forecast errors ranged from 1.44% to 5.12% for hourly predictions and from 3.35% to 10.44% for 15-minute predictions. Blázquez-García et al. (2021) used a self-supervised framework with the Random Interval Spectral Ensemble (RISE) classifier to analyse nightly flow time series. The method sought a practical balance between detecting leakage and limiting false alarms. These studies illustrate that forecast-based leakage detection depends both on prediction quality and on the decision rule applied to the resulting residuals or time-series patterns.

Fares et al. (2023) provide a useful field benchmark: two years of wireless acoustic monitoring in metal and non-metal Hong Kong WDSs were used to compare SVM, ANN and DL models. The study showed strong performance on new and unlabeled cases but also demonstrated that pipe material and ambient noise materially change acoustic features, reinforcing the need for cross-condition validation rather than relying on a single aggregate accuracy. Overall, classical and deep models can all perform well under

appropriate conditions; the practical distinction lies in data demand, interpretability, computational burden, robustness to noise and domain shift and the consequences of false alarms. These criteria guide the comparative synthesis that follows.

3.6.2. Feature Engineering and Sensor Fusion

Feature engineering constructs, transforms, or selects variables before classification, regression, or anomaly detection. Early leakage studies combined statistical, spectral, hydraulic, and spatial features with ANN, SVM, SOM, and Bayesian models (Caputo and Pelagagge 2002; Mounce et al. 2010, 2011; Romano et al. 2014). Tao et al. (2014) used a multilevel neural workflow in which the first stage detected leakage and the second estimated its location and intensity. Soldevila et al. (2019) combined Kriging-based spatial interpolation with pressure measurements from normal operation, although the resulting spatial assumptions may not remain valid under irregular leakage patterns. These approaches demonstrate that preprocessing and feature design can improve dimensionality, interpretability, and model performance, but the selected features must remain stable across operating conditions and validation environments.

Recent studies combine feature selection, learned representations, and sensor fusion. Xu et al. (2024) used MDMR-ISFFS to select discriminative acoustic features; five classifiers achieved approximately 94–98% accuracy, with RF and XGBoost producing the strongest results. Martinez-Ríos et al. (2024) combined wavelet-scattering classifiers for acceleration, dynamic-pressure and hydrophone measurements through late fusion. McMillan et al. (2024) learned a compact variational representation before SVM classification, while Satterlee et al. (2025) used few-shot CNN learning with multisensor fusion. Xu et al. (2025) used MGrad-CAM to relate multi-channel convolutional neural network (MCNN) decisions to frequency bands, pressure, leak flow and sensor distance. Table 5 summarizes these methods and their data requirements.

Table 5
Feature-engineering, representation-learning, and sensor-fusion workflows

References	Technique adopted	Input data
Sun et al. (2019)	Linear discriminant analysis; neural networks	Pressure
Zhou et al. (2019c)	Kernel principal component analysis; cascade support vector data description	Pressure
Soldevila et al. (2019)	Kriging spatial interpolation; Bayesian reasoning	Pressure
Cody et al. (2020a)	Linear prediction	Acoustic signals
Cody et al. (2020b)	CNN; VAE	Acoustic signals
Marvin et al. (2023)	RF plus histogram-based gradient boosting	Pressure
Mazaev et al. (2023)	Hybrid AI microservices plus GIS and utility integration	Pressure, flow
Choi and Im (2023)	FFT magnitude spectrum plus CNN, SVM	Vibration, acoustic signals
Basnet et al. (2023)	CNN plus MLP localization and size prediction	Pressure, flow
Sourabh et al. (2023)	MLP-ANN plus SVM classification and regression	Pressure or flow
Brahmbhatt et al. (2023)	Calibrated digital twin plus GCN	Pressure, flow, network topology
Jun and Lansey (2023a)	Comparison of SCADA and AMI detection-localization configurations	Flow, pressure, demand
Choudhary et al. (2023)	1-D CNN	Multisensor pressure, flow, quality data
Nascimento and Gomes-Jr (2023)	Semi-supervised AutoML plus SOM and local outlier factor (LOF)	Flow
Jun and Lansey (2023b)	2-D CNN plus hydraulic pressure responses	Pressure, demand
Yussif et al. (2023)	Empirical mode decomposition (EMD) and denoising plus SVM, KNN, DT	Acoustic signals
R. Liu et al. (2024b)	Contrastive and self-supervised acoustic representation learning	Acoustic signals
McMillan et al. (2024)	VAE and variational neural representation plus SVM	Flow

References	Technique adopted	Input data
Xu et al. (2024)	MDMR–ISFFS plus DT, RF, XGBoost, SVM, MLP	Acoustic signals
Gunatilake and Miro (2024)	M-TFCNN	Vibro-acoustic signals
Rajasekaran and Kothandaraman (2024)	1D time-series DenseNet	Acousto-optic signal
Parajuli and Shin (2024)	ML failure-type classification	Pressure, flow, tank level
Joseph et al. (2024)	Logic-based plus ML detection	SCADA pressure and flow
Komba et al. (2024)	SVM-RF plus IoT architecture	Multisensor, IoT data
Liu et al. (2025)	Time-delay DL	Acoustic signals
Xie et al. (2025)	variational mode decomposition (VMD) plus wavelet denoising plus SVM	Acoustic signals
Molina et al. (2025)	SHIELDS probabilistic and Bayesian network	Pressure, flow, velocity, elevation, head loss, demand
Satterlee et al. (2025)	Parallel multi-layer sensor fusion plus CNN and few-shot learning (FSL)	Multisensor signals
Joseph et al. (2025)	14 supervised ML classifiers	Pressure and flow
Garzón-Orduña et al. (2026)	Physics-informed transient model plus supervised ML	Valve state, transient hydraulic, leakage variables
Ranganathan et al. (2026)	ML, DL, GNN and Transformer benchmark	Simulated hydraulic, sensor data (Net3)
Liu et al. (2026)	Deep acoustic retrieval plus RAG-LLM diagnosis	Acoustic signals, retrieved historical cases

3.6.3. Direct Classification and Localization

Classification-based approaches have gained increasing popularity for leak detection in WDSs. By training on either historical or simulated data, ML algorithms can automatically recognize leakage patterns based on pressure, flow, or acoustic signals. Early studies used supervised models such as support vector regression (SVR) (Kemba et al. 2017) and SVM, while later work adopted DenseNet, CNN, probabilistic models and clustering (Zhou et al. 2019b; Wu et al. 2022), frequently using EPANET-generated training data. Persistent challenges include detecting small or multiple leaks, managing class imbalance, reducing false alarms and transferring models to complex or noisy networks.

Classification targets have also expanded beyond current leak detection. Joseph et al. (2025) compared fourteen supervised models for no-leak, minor-leak, and major-leak classification. Farajzadeh et al. (2024) used one-class (OC)-SVM, Isolation Forest, and Elliptic Envelope with 1,467 burst-pipe records

and asset variables, while Karadirek et al. (2024) applied LR to pipe-failure risk. These studies extend classification toward proactive maintenance and pipe prioritization.

Localization research increasingly combines multiple measurements and network structure. Mashhadi et al. (2021) evaluated six supervised and unsupervised methods using 215 simulated leakage scenarios across five hydraulic zones in a 15-km campus WDS. LR, RF, and ANN achieved perfect classification metrics when pressure and flow were combined, whereas unsupervised methods were weakened by overlapping clusters. Fan and Yu (2022) introduced the topology-aware CtL-SSL framework, achieving approximately 95% detection and 83% localization accuracy despite leak samples comprising less than 10% of the data. Under such imbalance, however, accuracy should be complemented by leak-class recall, precision, F1-Score, area under the precision–recall curve (PR–AUC), false-alarm rates, and class-wise localization metrics.

Nagaraj et al. (2021) combined EPANET simulations and IoT sensors in a WDS containing 93 nodes and 117 pipes divided into eight sections. DT and RF achieved near-perfect accuracy with combined pressure and flow data, although small leaks remained difficult to detect. Basnet and Ranjithan (2022) generated 100,000 L-Town scenarios containing up to three leaks across 33 zones under demand, measurement and hydraulic-model uncertainty. CNN generally outperformed MLP for leak location and size, particularly under noise and multiple-leak conditions, but additional flow measurements and real-network validation were recommended.

Zhang et al. (2016) combined hydraulic simulations, leak-zone partitioning and multiclass SVM for large-network localization, providing a nonlinear approach where real leak records are limited. Fan (2024) extended CtL-SSL through principal component analysis (PCA) or autoencoder dimensionality reduction, modified k-means zoning, model-informed sensor placement and separate detection and localization models. The framework achieved approximately 95% detection and 83% localization accuracy on the C-Town and Rancho Solano Zone III networks despite class imbalance. Islam et al. (2022) combined hourly Automated Meter Readings, laboratory experiments and EPANET simulations at Lille University. Flow responses became less predictable with distance from the source, whereas combined flow and pressure measurements improved detection and localization across five zones.

Field research further shows the importance of sensing conditions. Fares et al. (2023) compared ten ML algorithms using two years of wireless acoustic monitoring in metal and non-metal Hong Kong WDSs. Performance varied with leak size, pressure, pipe material, and ambient noise, although several models correctly classified 10 of 11 new field sites. Other 2023 studies used histogram-based gradient boosting (Marvin et al. 2023), hybrid-AI microservices (Mazaev et al. 2023), attention-based autoencoders (Karimanzira 2023), acoustic XGBoost and DL (Wu et al. 2023a), and calibrated digital twins with graph convolution (Brahmbhatt et al. 2023), reflecting a shift toward integrated, topology-aware, and operationally oriented localization. Table 6 summarizes representative approaches, findings, and limitations.

Table 6
Supervised and hybrid leakage-detection and localization studies

Study	Model or approach	Main finding	Main limitation
Zhang et al. (2016)	K-means zoning and multiclass SVM	Localized leaks by network zone	Sensitive to clustering choices
Kemba et al. (2017)	SVR with simulated pressure data	Achieved localization within 20 m	Simulation-dependent validation
Zhou et al. (2019a)	DenseNet and probabilistic modelling	Extracted features from limited pressure signals	Restrictive uncertainty assumptions
Mashhadi et al. (2021)	LR, RF and ANN	Combined pressure and flow improved classification	Dependent on calibrated simulations
Nagaraj et al. (2021)	DT, RF and SVM	Tree models achieved near-perfect detection	Small leaks remained difficult
Fan and Yu (2022)	Topology-aware self-supervised learning	Performed well under class imbalance	Limited field validation
Basnet and Ranjithan (2022)	MLP and CNN	CNN performed better under complex leak conditions	Sensitive to noise and leak complexity
Fares et al. (2023)	Acoustic ML models	Demonstrated detection in operational WDSs	Sensitive to material and noise
Choi and Im (2023)	CNN using vibration spectra	Achieved an F1-Score of 94.82%	Limited cross-site validation
Barros et al. (2023b)	Graph signal processing	Achieved 86% accuracy across seven events	Small validation dataset
Yussif et al. (2023)	Acoustic SVM and k-NN	Supported localization and direction classification	Single-sensor validation
Joseph et al. (2025)	Fourteen supervised classifiers	KNN, RF and DT achieved perfect classification	Controlled, low-noise dataset
Ranganathan et al. (2026)	ML, DL, GNN and transformer benchmark	Detection exceeded 95%; localization was lower	Simulation-based validation
Liu et al. (2026)	Retrieval-augmented LLM	Produced high recall and low false-positive rates	Grounding and transfer remain uncertain

3.6.4. Cross Study Performance Synthesis

Reported performance values are comparable only when class prevalence, data source, sensor configuration, leak magnitude and output definition are sufficiently aligned. Deployment assessment should give field performance and false-alarm burden greater weight than a single benchmark accuracy.

Matched validation explains many apparent performance differences. Xu et al. (2025) reported 99.58% accuracy in clean laboratory data but 95.40% accuracy and 93.61% F1-Score on excavation-confirmed field acoustic data, directly quantifying a domain gap. Joseph et al. (2025) reported perfect test-set precision, recall, F1-Score and accuracy for KNN, RF and DT in a controlled three-class experiment, but explicitly cautioned that balanced, low-noise and clearly separated classes can overstate field generalization. Onukwube et al. (2024) reported 79% for SVM in a campus detection and localization problem, while Gunatilake and Miro (2024), Rajasekaran and Kothandaraman (2024), Chandanwala et al. (2025) and Satterlee et al. (2025) reported approximately 95–99% in acoustic and sensor-fusion settings. Ranganathan et al. (2026) further showed that over 95% binary detection can coexist with only about 75–77% top-1 localization. Thus, the most informative synthesis is not a high and low accuracy label but the combination of task difficulty, real-world relevance, class balance, localization error and robustness. Table 7 compares the performance metrics of different leak detection approaches reported in recent studies.

Table 7
Leak detection performance comparison

References	Data type	Network/ setting	Analytical method	Reported Accuracy (%)
Kang et al. (2017)	Measured data	City of Seoul WDS	1D-CNN	98.8
			1D-CNN-SVM	99.3
Chuang et al. (2019)	MFCCs	Underground pipe	2D-CNN	99.4
Bjerke (2019)	Generated data	Net1 WDS	GRU	99.31
			LSTM	99.26
		Hanoi WDS	GRU	90.68
Wang et al. (2020)	Simulated data	South of China DMA	LSTM	99.8
	Synthetic data		LSTM	99.77
Cody et al. (2020b)	Collected data	Laboratory test bed	CNN-VAE	97.2
Rahimi et al. (2020)	Collected data	Laboratory test bed	1D-CNN	75
Xu et al. (2025)	Measured field acoustic data	Dalian WDS	MCNN	95.40
Gunatilake and Miro (2024)	Measured field vibro-acoustic data	Active urban WDS	M-TFCNN	95–98
Rajasekaran and Kothandaraman (2024)	Acousto-optic time series	Water pipeline	1D DenseNet	99.08
Onukwube et al. (2024)	Hydraulic network data	Choba campus WDS	SVM	79
Chandanwala et al. (2025)	Acoustic MFCC features	Water pipeline	RCNN	95–98
Satterlee et al. (2025)	Multisensor data	Pipe leak test system	CNN plus FSL	97.1 detection
Joseph et al. (2025)	Measured pressure and flow	Melton experimental system	KNN, RF, DT	100
Choudhary et al. (2023)	IoT multisensor data	Smart-water- grid prototype	1-D CNN	94.32 detection

References	Data type	Network/ setting	Analytical method	Reported Accuracy (%)
Barros et al. (2023b)	Pressure correlations	WDS graph and 7 leaks	Graph signal processing (GSP) plus PageRank	86
Yussif et al. (2023)	Acoustic noise-logger data	Urban water mains	Ensemble KNN	83.75 localization
Choi and Im (2023)	Vibration magnitude spectra	Water pipeline test system	CNN	F1-Score 94.82
Ranganathan et al. (2026)	Simulated hydraulic data	Net3 (over 47,000 samples)	RF, XGBoost, LightGBM; DL; GNN; Transformer	About 75–77 top-1 localization for best ensembles

3.7. Physics Informed and Hybrid Learning

Physics-informed and hybrid learning combines statistical or ML inference with hydraulic equations, calibrated model outputs, physical constraints, network topology, or model-data assimilation (Ferrandez-Gamot et al. 2015; Soldevila et al. 2016, 2017; Zhang et al. 2016). Hybridization may occur through residual analysis, sensitivity features, constrained learning, digital twins, or physically informed training data. Using hydraulically simulated observations for training does not by itself make a model hybrid; the physical model or its constraints must contribute directly to feature construction, learning, inference, or decision validation. These approaches can reduce the feasible localization space and improve physical consistency, although their performance remains sensitive to calibration, demand assumptions, and simulation-to-field transfer.

Barros et al. (2025) combined monitoring data and network structure using a temporal pressure-correlation graph for anomaly detection, a multilayer graph to reduce the localization search space, and DTW to compare monitored and simulated responses. On the BattLeDIM and L-Town benchmark, abrupt leaks were detected within 15 min with some localization errors of 42–43 m, whereas gradual leaks could remain undetected for hundreds of hours or be localized more than 300 m from the actual pipe. These findings illustrate both the value of network-aware analysis and the difficulty of incipient leakage.

Hybrid physics-data studies use several forms of model coupling. Cholewa et al. (2024) combined network modelling with iterative mobile-sensor relocation and reported an average localization error of approximately 180 m using three pressure sensors. Molina et al. (2025) developed the SHIELDS probabilistic framework using hydraulic and demand variables. Digital-twin studies combine calibrated hydraulic simulations with continuously updated sensor data, while physics-informed models incorporate equations or physical constraints directly into training or prediction. By contrast, combining two classifiers or joining ML with an IoT platform constitutes model or system integration but should not automatically be classified as physics-informed learning.

Other studies use different forms of hybridization. Cholewa et al. (2024) reported approximately 180 m average localization using three pressure sensors and iterative mobile-sensor relocation. Molina et al. (2025) developed the SHIELDS Bayesian framework using hydraulic and demand variables, while McMillan et al. (2024) combined variational representation learning with SVM and Komba et al. (2024) integrated SVM-RF classification within an IoT architecture. Hybridization can therefore involve physical constraints, topology, learned representations, probabilistic inference, classifier fusion, or sensing infrastructure.

Noori et al. (2026) integrated pressure-driven EPANET simulation, GIS-based Fuzzy analytic hierarchy process (AHP), pressure-sensor siting, and supervised ML in a full-scale urban WDS. Screening reduced more than 39,000 nodes to 4,699 candidates, and XGBoost achieved accuracy = 0.961 and area under the receiver operating characteristic curve (ROC–AUC) = 0.989; however, the simulation-derived labels require independent operational validation. Garzón-Orduña et al. (2026) generated physically consistent transient leakage data from 16 valve-closure scenarios and evaluated tree, Gaussian-process, and neural-network models. All exceeded $R^2 = 0.999$, with root mean square error (RMSE) as low as 0.0019 L/s, supporting physics-informed surrogate modelling while still requiring utility-scale validation.

Sitaropoulos et al. (2023) used low-cost hydrant hydrophones in an urban WDS and identified a repeatable 150–450 Hz leak-related power increase despite site-dependent noise. Marvin et al. (2023) used histogram-based gradient boosting on uncertain simulated pressure measurements to estimate leak location and other event parameters, whereas Jun and Lansey (2023a) showed that conventional SCADA configurations can be inadequate for realistic-size leaks and that AMI data, particularly pressure-supplemented AMI, substantially improve small-leak detection and localization. Wu et al. (2023b) integrated sensing, calibration, analytics and hydraulic simulation for a high-fidelity digital twin for more than 1,000 km of Singapore network and obtained greater than 80% F1-Score for anomaly detection with localization within about 400 m, demonstrating how sensing, calibration, data analytics and hydraulic simulation can be integrated at field scale.

Hybrid methods may improve performance through hydraulic calibration, residual features, or learning-based inference (Soldevila et al. 2017; Zhou et al. 2019c), but additional components also introduce potential failure modes. Calibration uncertainty can propagate into ML features and should therefore be reported alongside independent validation. Prediction models may also support hybrid workflows (Geng et al. 2019), although prediction-classification performance ultimately depends on forecast accuracy and subsequent decision rules (Wu and Liu 2017).

3.8. AI Enabled Decision and Control

AI-enabled decision and control methods learn operational actions or policies from data, hydraulic simulations, or repeated interaction with a modelled environment. RL and DRL are increasingly investigated for pump scheduling, valve operation, and pressure management, but leak-specific applications remain limited relative to general energy and pressure control (Negm et al. 2024; Javed et al. 2025). Important barriers include simulator latency, large state and action spaces, reward-function

design, uncertainty, sparse feedback, and simulation-to-real transfer. These methods are therefore better supported as leakage-prevention and operational-control tools than as replacements for established leakage-detection models.

Recent studies illustrate the potential of learning-based WDS control. Negm et al. (2023b) used Q-learning for pressure-reducing valve (PRV) control in D-Town; a 26 m setting reduced mean pressure by approximately 2% while maintaining the regulatory minimum of 10 m. Hu et al. (2023b) used exploration-enhanced proximal policy optimization (PPO) for real-time pump scheduling, reducing online computation to approximately 0.42 s and reporting energy-cost reductions of up to 11.14%. Hu et al. (2023a) coordinated pumps and valves through a multi-agent deep deterministic policy gradient (MADDPG) framework and obtained faster solutions than GA, particle swarm optimization (PSO), and differential evolution in the evaluated scenarios. Zaman et al. (2023) also investigated DRL for real-time WDS control. These studies concern preventive pressure and operational management rather than direct leakage detection, and their deployment requires hydraulic safety constraints and independent evaluation under changing demands.

Table 8
AI-enabled control methods relevant to leakage management

Reference	Decision task	AI method	Reported outcome	Main limitation
Negm et al. (2023b)	PRV pressure control	Q-learning	A 26 m setting reduced mean pressure by approximately 2% while maintaining the 10 m minimum	Simulation based; limited evidence of field transfer
Hu et al. (2023b)	Real-time pump scheduling	Exploration-enhanced PPO	Approximately 0.42 s online computation and up to 11.14% energy-cost reduction	Reward and simulator dependent; not primarily a leakage-detection study
Hu et al. (2023a)	Coordinated pump and valve operation	Multi-agent MADDPG	Faster solutions than GA, PSO, and differential evolution in the evaluated scenarios	Large state/action spaces and simulation-to-real transfer
Zaman et al. (2023)	Real-time WDS control	DRL	Demonstrated a learning-based control framework	Utility-scale and leak-specific validation remain limited

4. Discussion

The central contribution of this review is a decision framework that links utility problems to data, analytical methods, validation, and broader infrastructure-management objectives rather than ranking algorithms by headline accuracy. Existing studies address different tasks, including event screening, localization, leak-size estimation, future-risk prediction and AI-enabled pressure control, using data that

range from fully labelled simulations to noisy and imbalanced field records. Method selection should therefore begin with the operational task and sensing context, match model assumptions to those conditions, and use metrics that reflect missed leaks, false alarms, detection delay, and localization burden. From an urban sustainability perspective, these methodological choices ultimately determine whether leakage analytics can support measurable reductions in water loss and energy demand while improving service reliability and infrastructure resilience. Consequently, model performance should be evaluated not only as a computational outcome but also in relation to its capacity to support practical and resource-efficient utility decisions.

4.1. Data Validity, Generalization and Reproducibility

Research through 2026 confirms a persistent benchmark-to-field gap. Zuñiga-Urbe et al. (2026) found that simulation-derived datasets remain dominant and that high reported accuracies are difficult to compare because sensing, leak scenarios, and validation differ. Rajan and Li (2025) similarly found that few studies implement the full chain from acquisition to alert and interface. Field-based investigations, including Xu et al. (2025), Fares et al. (2023), and Sitaropoulos et al. (2023), provide more deployment-relevant assessments by incorporating noise, pipe-material effects, and operational variability. Barros et al. (2025) showed that abrupt leaks may be detected within minutes, whereas gradual leaks can require hundreds of hours before producing sufficiently strong hydraulic signatures. Joseph et al. (2025) achieved perfect classification scores in a balanced, controlled testbed but cautioned that noise-free conditions may overestimate real-world performance. Performance evaluation should therefore extend beyond accuracy to include detection delay, leak growth rate, localization distance, false alarms, sensor density, and training and test data provenance, with independent field validation remaining essential.

Reproducibility is also a major deployment issue. Oberascher et al. (2026) found that many influential urban-water ML studies omit essential preprocessing, model-selection, and hyperparameter-tuning details, while fewer than one quarter report explainability or uncertainty. For leakage analytics, train and test separation by event, time, or site should be stated explicitly; scaling and feature selection should use training data only; hyperparameter tuning should remain separate from final testing; and leakage prevalence should be reported. When field data are unavailable, simulation should support controlled stress testing rather than be treated as equivalent to independent field validation. Noori et al. (2026) illustrates this distinction: although developed for a full-scale WDS with detailed hydraulic and infrastructure information, leakage-prone labels were generated through pressure-driven simulation with a random node-level train and test split. Its reported XGBoost performance should therefore be interpreted as simulation-based prioritization rather than direct field leak-detection capability, with spatially separated and field-confirmed validation still required.

Several 2023 studies help bridge the simulation-dominated research base. Sitaropoulos et al. tested hydrophones in a real urban WDS; Fares et al. (2023) used two years of Hong Kong acoustic field data; Glynis et al. (2023) evaluated real bursts across 10 UK DMAs; Henriques-Silva et al. (2023) used two years of Halifax flow records; and Wu et al. (2023b) deployed a high-fidelity digital-twin framework

across more than 1,000 km of Singapore network. These studies expose difficulties often suppressed in idealized experiments, including background noise, pipe-material effects, demand nonstationarity, sensor-configuration changes, and false alarms. Conversely, simulation and benchmark studies such as Rajabi et al. (2023), Basnet et al. (2023), Brahmabhatt et al. (2023), and Jun and Lansey (2023a) provide controlled environments for testing uncertainty, spatial representations, and sensor architectures. The strongest methodological basis therefore combines systematic simulation-based stress testing with independent field validation.

Class imbalance also changes the operational meaning of model performance. Leak events are usually a minority of time steps and assets, so accuracy can remain high even when rare leaks are missed. Utility-facing evaluation should therefore include leak recall, precision, F1-Score or PR-AUC, false positive rate (FPR) or false alarms per day, and T_{det} , together with localization distance and top-k accuracy where applicable. Saito and Rehmsmeier (2015) showed why precision-recall analysis is more informative than ROC alone under strong imbalance. These metrics also have direct operational consequences: False negative (FN) implies lost water and delayed intervention; false positive (FP) consumes inspection resources; large T_{det} allows events to grow; and localization error increases excavation and search costs.

Model selection should remain conditional on the data, task, and deployment environment. Statistical and probabilistic models provide transparent baselines for forecasting and residual monitoring. SVM, KNN, RF, and boosting remain strong and auditable choices for engineered hydraulic, acoustic, and asset variables. CNN and contrastive-learning models are attractive for high-frequency acoustic and vibration data, while LSTM and GRU models suit sequential hydraulic monitoring. GNN, Transformer, digital-twin, and physics-informed approaches are increasingly relevant when topology, spatial relationships, or hydraulic constraints are important. ELM and GELM can provide rapid fitting but have more limited field validation. RL and DRL are most relevant to AI-enabled pressure and operational control rather than direct leakage detection. Across all families, independent validation by site, time, and event should receive greater weight than isolated benchmark accuracy.

4.2. Sustainable and Resilient Leakage Management

The framework proposed in this review uses four stages to translate analytical capability into operationally relevant leakage management. Stage 1 defines the operational task, including screening, localization, severity estimation, prediction and early warning, or AI-enabled mitigation and control. Stage 2 defines the data context through sensing modality, sampling rate, topology and asset information, label availability, class prevalence, and data provenance. Stage 3 selects an analytical family and determines whether statistical inference, ML, neural representation, or hydraulic and physical constraints are required. Stage 4 specifies validation through independent testing by event, time, or site; uncertainty and calibration analysis where relevant; and task-appropriate performance measures. Table 9 translates these stages into practical choices. The framework is descriptive and literature-informed and does not imply that one model is optimal for every WDS. Its purpose is to balance analytical performance with water-loss reduction, implementation requirements, operational reliability, and infrastructure resilience.

Table 9
Decision framework for selecting and evaluating leakage-management analytics

Utility objective	Preferred inputs	Candidate analytical approaches	Minimum evaluation	Deployment caution
Rapid burst screening	High-frequency flow and pressure; DMA and SCADA context	Statistical residuals, model–data assimilation, RF, boosting, SVM, and lightweight sequence models	Leak recall, FPR or false alarms per day, and detection delay	Demand and operational transients can resemble bursts
Small and background leakage detection	Acoustic, vibration, dense pressure, and repeated normal-operation data	CNN, contrastive learning, one-class methods, autoencoders, and multisensor hybrids	PR–AUC, F1-Score, FPR, SNR testing, and field transfer	Signal attenuation, environmental noise, and class imbalance can dominate performance
Leakage localization	Pressure and flow with topology; spatial acoustic measurements	CNN, GNN, Transformer, digital twins, and model-assisted classification	Distance error, top-k pipe accuracy, zone accuracy, and search-area reduction	Resolution depends on sensor density and hydraulic-model fidelity
Scarce confirmed leakage labels	Large normal or unlabeled archives and limited confirmed events	One-class, semi-supervised, self-supervised, contrastive, and generative augmentation methods	Event-level recall, FPR, label-efficiency curves, and external validation	Synthetic or augmented samples must remain separated from final test data
Leakage prediction and early warning	Longitudinal hydraulic, asset, operational, and environmental variables	Regression, Bayesian models, tree ensembles, LSTM, GRU, and physics-informed learning	MAE, RMSE, R^2 , calibration, lead time, and alarm performance	Nonstationarity requires time-separated validation
AI-enabled pressure and leakage mitigation	Calibrated hydraulic or digital-twin states; pumps, valves, and service constraints	RL, DRL, MARL, and safety-constrained learned control policies	Water loss, energy, pressure violations, reliability, policy stability, and safety	Learned policies require hydraulic constraints, fail-safe operation, and simulation-to-real testing
Operator decision support	Validated detector outputs, historical cases, GIS context, and hydraulic tools	RAG-LLM and agentic orchestration layered over verified analytical models	Factual grounding, tool-task success, uncertainty, human override, and conventional leakage metrics	Language-model output should not directly actuate infrastructure without deterministic verification

Climate and resilience variables should be included only when they provide plausible, testable incremental information. Temperature cycles, drought-related soil movement, extreme precipitation, and demand and pressure changes can affect pipe loading and operating conditions, but environmental covariates do not automatically improve leak detection. Bui et al. (2025) situates AI within climate-resilient urban-water infrastructure. Future studies should therefore test whether climate variables improve recall, lead time, or calibration beyond pipe age and material, pressure, demand, and seasonality using time- and site-separated validation.

Leakage also affects outcomes beyond hydraulic loss. Barros et al. (2023a) investigated its effects on water-quality parameters in distribution networks, supporting more integrated hydraulic and quality monitoring. This is relevant to climate-resilience research because temperature, source-water conditions, and demand shifts may alter both pipe-failure likelihood and leakage consequences. Future detection frameworks should therefore assess whether environmental and water-quality covariates add predictive value without excessive dimensionality.

4.3. Emerging Physics Informed Generative and Agentic AI

Ranganathan et al. (2026) benchmarked classical ML, DL, GNN, and Transformer models within a reproducible framework; Garzón-Orduña et al. (2026) coupled transient hydraulic physics with supervised ML to generate physically consistent leakage scenarios and improve prediction plausibility; and Liu et al. (2026) used retrieval-augmented LLMs to ground acoustic diagnoses in similar historical cases. These developments do not invalidate classical models but clarify where newer architectures add value: GNN and Transformer models can represent spatiotemporal structure; physics-informed machine learning (PIML) can impose physical coherence; contrastive learning can exploit unlabeled signals; and retrieval-augmented generation (RAG) can support source-grounded reporting and human interaction.

Generative and agentic systems also introduce failure modes not captured by conventional accuracy measures. A hydraulic detector may fail through classification error, whereas an LLM layer may additionally hallucinate an asset state, misinterpret a retrieved case, or invoke an inappropriate tool. Future agentic leakage systems should therefore separate deterministic sensing and model calculations from language-model orchestration, log retrieved sources and tool calls, enforce permissions, quantify uncertainty, and retain human or rule-based confirmation for consequential control actions. EPANET-Agentic (Wang et al. 2026) demonstrates how a multi-agent architecture can orchestrate specialized tools for natural-language-controlled WDS simulation and analysis. However, it should be regarded as an enabling architecture rather than a validated autonomous leak-management system because operational actions still require deterministic hydraulic checks, controlled tool permissions, traceable outputs, and human or rule-based safeguards. This distinction is essential when translating AI advances into infrastructure practice.

4.4. Smart Water Infrastructure, System Integration and Utility Implementation

The most recent literature also links leakage management and infrastructure resilience to broader system integration. Bui et al. (2025) discusses AI for urban water networks in the context of climate change and emphasizes factors extending beyond conventional hydraulic inputs. Rajan and Li (2025) show that automated leak management increasingly depends on IoT and AMI acquisition, reliable preprocessing, real-time interfaces, GIS context, and digital-twin-type integration, while only a minority of reviewed systems close the loop from data collection to alerting and user interaction. These technologies position leakage analytics as one component of an increasingly connected smart-water infrastructure rather than an isolated prediction problem. Future climate-aware leakage models should therefore be evaluated as part of a sensing, communications, prediction, and decision-support chain. For utilities, the practical value of such integration lies in enabling earlier intervention, more targeted inspection and repair, improved pressure management, and more efficient allocation of limited operational resources. These capabilities connect AI-enabled leakage management directly with water conservation, service continuity, and the resilience of urban water infrastructure.

4.5. Research gaps and future directions

The findings synthesized above point to recurring unresolved gaps rather than an absence of capable algorithms. Two of the most consequential concern the credibility of reported findings. Most reported performance still originates from simulation or benchmark datasets (e.g., LeakDB, BattLeDIM, C-Town, Net3); Zuñiga-Urbe et al. (2026) found these results difficult to pool because sensing, leak scenarios, and validation protocols differ, while field studies such as Xu et al. (2025), Fares et al. (2023), and Sitaropoulos et al. (2023) report lower, more decision-relevant performance under noise, material effects, and operational variability. Few studies test the same model across simulation, laboratory, and field regimes, leaving uncertainty regarding deployment performance. This benchmark-to-field gap is compounded by a reproducibility gap: Oberascher et al. (2026) found that many influential urban-water ML studies omit preprocessing, model-selection, and hyperparameter-tuning details, while explainability and uncertainty are reported in fewer than one quarter of the reviewed papers. Such omissions limit independent replication and fair cross-study comparison.

A second cluster of gaps concerns performance measurement and comparison. Leak events are a minority class in most operational datasets, yet accuracy remains frequently reported. Saito and Rehmsmeier (2015) show that precision–recall analysis is more informative than ROC-based metrics under strong imbalance, but leak recall, PR–AUC, false alarms per day, and detection delay remain inconsistently reported. This limits utilities’ ability to compare models using operationally meaningful outcomes. The absence of a shared benchmark protocol further complicates comparison because studies differ in leak magnitude, sensor density, network topology, and evaluation window. A community-agreed benchmark suite spanning simulation, laboratory, and field conditions would materially improve comparability.

A third set of gaps concerns the translation of capable models into trustworthy operational systems. Recent physics-informed models, GNNs, Transformers, retrieval-augmented systems, and agentic architectures show technical promise but have not yet been evaluated fully against the safety, reliability,

and human-oversight requirements of consequential infrastructure decisions (Garzón-Orduña et al. 2026; Ranganathan et al. 2026; Liu et al. 2026; Wang et al. 2026). Hallucination, misretrieval, and inappropriate tool invocation remain potential risks for language-model components, and leakage-specific studies have not yet reported systematic evaluation of these failure modes. Rajan and Li (2025) also found that only a minority of reviewed systems complete the full sequence from sensing and preprocessing through prediction to alerting and operator interaction. Most models are still evaluated as isolated predictors rather than as components of deployed sensing-to-decision systems, leaving unresolved questions concerning latency, alert fatigue, cybersecurity, and integration with SCADA and GIS infrastructure.

Two further gaps concern context and resources. Bui et al. (2025) and Barros et al. (2023a) motivate incorporating climate and water-quality covariates into leakage analytics, but whether these variables improve recall, lead time, or calibration beyond established predictors such as pipe age, material, pressure, and demand has not been demonstrated using time- and site-separated validation. This therefore remains a testable hypothesis rather than an established benefit. A related data-sharing and cost-of-sensing gap persists because field-quality leakage datasets are expensive to acquire and rarely released publicly, limiting independent verification and concentrating field-validation experience among relatively few utilities and research groups.

Addressing these gaps does not require abandoning established model families. It requires treating reproducibility, imbalance-aware evaluation, field validation, uncertainty, and safe system integration as research outcomes alongside predictive performance. Modern approaches, including contrastive learning, GNNs, Transformers, physics-informed learning, generative models, and agentic systems, should therefore be evaluated according to whether they improve deployment reliability rather than architectural novelty alone.

Collectively, these gaps indicate that the next stage of leakage research should move from algorithm-centred performance improvement toward system-level demonstration of urban infrastructure benefit. Future studies should evaluate whether improved detection, localization, forecasting, or AI-enabled control translates into measurable reductions in water loss, energy consumption, intervention time, service disruption, or lifecycle operating burden. Such evaluation would provide a stronger basis for assessing the contribution of AI-enabled leakage management to sustainable and resilient urban water systems and align methodological advances more directly with utility and societal outcomes.

Conclusion

Water leakage remains a persistent challenge to the sustainability and resilience of water distribution systems (WDSs) because it increases treated-water losses, energy requirements, operational costs, service disruption, and potential contamination risk. This review organizes leakage analytics into statistical and probabilistic data models, classical machine learning (ML), neural and representation learning, physics-informed and hybrid learning, and AI-enabled decision and control. These analytical

families support different tasks, including detection, localization, severity estimation, failure prediction, early warning, and operational mitigation. Conventional mathematical and metaheuristic optimization may use analytical outputs to support management decisions, but it is not classified here as an AI or data-driven model family. The reviewed evidence does not identify a universally superior approach. Performance depends on data provenance, sensing modality, network topology, leak magnitude, class prevalence, noise, model assumptions, and validation design. Simulation and benchmark datasets enable controlled testing but cannot replace independent field validation. Laboratory systems provide intermediate physical fidelity, while utility data expose demand variability, sensor limitations, material effects, class imbalance, and nonstationarity that may be absent from controlled experiments. Recent studies extend the field beyond conventional SVM, KNN, ANN, and statistical comparisons through acoustic and multisensor fusion, contrastive learning, GNN and Transformer models, digital twins, physics-informed learning, AI-enabled pressure control, and retrieval-augmented decision support. These advances require task-specific evaluation: imbalance-aware metrics and false-alarm rates for detection; distance error and top-k accuracy for localization; error, calibration and lead time for prediction; and water loss, energy use, pressure compliance, reliability and safety for learned control. The proposed framework connects utility objectives with appropriate data, analytical approaches, validation designs, and operational measures. Important gaps remain in benchmark-to-field transfer, transparent reporting, standardized evaluation, end-to-end utility integration, safe use of generative and agentic AI, and empirical evaluation of climate and water quality variables. Future research should determine whether analytical improvements produce measurable reductions in non-revenue water, energy use, intervention time, service disruption, and lifecycle burden. Demonstrating those outcomes is necessary for AI and data-driven leakage management to contribute meaningfully to sustainable and resilient WDSs.

Abbreviations

Abbr.	Full Term
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DMA	District Metered Area
DL	Deep Learning
DRL	

Deep Reinforcement Learning
DT
Decision Tree
DTW
Dynamic Time Warping
ELM
Extreme Learning Machine
EWMA
Exponentially Weighted Moving Average
FFT
Fast Fourier Transform
FPR
False Positive Rate
GA
Genetic Algorithm
GELM
Gaussian Extreme Learning Machine
GIS
Geographic Information System
GNN
Graph Neural Network
GRU
Gated Recurrent Unit
IoT
Internet of Things
KF
Kalman Filter
KNN
K-Nearest Neighbours
LLM
Large Language Model
LR
Logistic Regression
LSTM
Long Short-Term Memory
MADDPG
Multi-Agent Deep Deterministic Policy Gradient
MCC
Matthews Correlation Coefficient
MCNN

Multi-Channel Convolutional Neural Network
MFCC
Mel-Frequency Cepstral Coefficient
MGrad-CAM
Multi-Channel Gradient-Weighted Class Activation Mapping
ML
Machine Learning
MLP
Multilayer Perceptron
PPO
Proximal Policy Optimization
PR–AUC
Area Under the Precision–recall Curve
RAG
Retrieval-Augmented Generation
RCNN
Region-Based Convolutional Neural Network
RF
Random Forest
RL
Reinforcement Learning
RMSE
Root Mean Square Error
RNN
Recurrent Neural Network
SCADA
Supervisory Control and Data Acquisition
SNR
Signal-to-Noise Ratio
SOM
Self-Organizing Map
SVM
Support Vector Machine
SVR
Support Vector Regression
VAE
Variational Autoencoder
WDS
Water Distribution System
XGBoost

Declarations

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research was supported by the National Research Council of Canada (Grant ID: GR007556).

Author Contribution

A.N. completed the preliminary study, gathered resources and prepared the draft manuscript. E.R. and H.B. acquired the funding, supervised the study, validated the work and reviewed the article. All the authors contributed to the final version of the article.

Data Availability

No datasets were generated or analysed during the current study.

References

1. Aksela K, Aksela M, Vahala R (2009) Leakage detection in a real distribution network using a SOM. *Urban Water J* 6(4):279–289
2. Almandoz J, Cabrera E, Arregui F, Cabrera Jr E, Cobacho R (2005) Leakage assessment through water distribution network simulation. *J Water Resour Plan Manag* 131(6):458–466
3. Alves Coelho J, Glória A, Sebastião P (2020) Precise water leak detection using machine learning and real-time sensor data. *IoT* 1(2):474–493
4. Bakker M, Vreeburg JHG, Van De Roer M, Rietveld LC (2014) Heuristic burst detection method using flow and pressure measurements. *J Hydroinformatics* 16(5):1194–1209
5. Bakker M, Vreeburg JHG, Van Schagen KM, Rietveld LC (2013) A fully adaptive forecasting model for short-term drinking water demand. *Environ Model Softw* 48:141–151
6. Barrientos-Torres D, Martinez-Ríos EA, Navarro-Tuch SA, Pablos-Hach JL, Bustamante-Bello R (2023) Water Flow Modeling and Forecast in a Water Branch of Mexico City through ARIMA and Transfer Function Models for Anomaly Detection. *Water* 15:2792. <https://doi.org/10.3390/w15152792>

7. Barros D, Almeida I, Zanfei A, Meirelles G, Luvizotto E Jr., Brentan B (2023a) An Investigation on the Effect of Leakages on the Water Quality Parameters in Distribution Networks. *Water* 15:324. <https://doi.org/10.3390/w15020324>
8. Barros D, Zanfei A, Menapace A, Meirelles G, Herrera M, Brentan B (2025) Leak detection and localization in water distribution systems via multilayer networks. *Water Res X* 26:100280. <https://doi.org/10.1016/j.wroa.2024.100280>
9. Barros DB, Souza RG, Meirelles G, Brentan B (2023b) Leak detection in water distribution networks based on graph signal processing of pressure data. *J Hydroinformatics* 25:2281–2290. <https://doi.org/10.2166/hydro.2023.047>
10. Basnet L, Brill D, Ranjithan R, Mahinthakumar K (2023) Supervised Machine Learning Approaches for Leak Localization in Water Distribution Systems: Impact of Complexities of Leak Characteristics. *J Water Resour Plan Manag* 149:04023032. <https://doi.org/10.1061/JWRMD5.WRENG-6047>
11. Basnet L, Ranjithan R (2022), January Supervised Machine Learning Models for Leak Detection in Water Distribution Systems. In *2nd International Joint Conference on Water Distribution Systems Analysis and Computing and Control in the Water Industry*
12. Bjerke M (2019) Leak detection in water distribution networks using gated recurrent neural networks (Master's thesis, NTNU)
13. Blázquez-García A, Conde A, Mori U, Lozano JA (2021) Water leak detection using self-supervised time series classification. *Inf Sci* 574:528–541
14. Bohorquez J, Alexander B, Simpson AR, Lambert MF (2020) Leak detection and topology identification in pipelines using fluid transients and artificial neural networks. *J Water Resour Plan Manag* 146(6):04020040
15. Brahmbhatt P, Maheshwari A, Gudi RD (2023) Digital twin assisted decision support system for quality regulation and leak localization task in large-scale water distribution networks. *Digit Chem Eng* 9:100127. <https://doi.org/10.1016/j.dche.2023.100127>
16. Bui MT, Yáñez-Godoy H, Elachachi SM (2025) Assessment of the Implications and Challenges of Using Artificial Intelligence for Urban Water Networks in the Context of Climate Change When Building Future Resilient and Smart Infrastructures. *J Pipeline Syst Eng Pract* 16:03124004
17. Butterfield JD, Meruane V, Collins RP, Meyers G, Beck SB (2018) Prediction of leak flow rate in plastic water distribution pipes using vibro-acoustic measurements. *Struct Health Monit* 17(4):959–970
18. Caputo AC, Pelagagge PM (2002) An inverse approach for piping networks monitoring. *J Loss Prev Process Ind* 15(6):497–505
19. Chan TK, Chin CS, Zhong X (2018) Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *Ieee Access* 6:78846–78867
20. Chandanwala AA, Bhowmik S, Chaudhury P, Rajasekaran U, Nesam JJJ, Kothandaraman M (2025) Water Pipeline Leakage Recognition and Localization Using Machine Learning and Deep Learning Techniques. *J Pipeline Syst Eng Pract* 16:04025035

21. Chen J, Tang P, Rakstad T, Patrick M, Zhou X (2020) Augmenting a deep-learning algorithm with canal inspection knowledge for reliable water leak detection from multispectral satellite images. *Adv Eng Inform* 46:101161
22. Choi J, Im S (2023) Application of CNN models to detect and classify leakages in water pipelines using magnitude spectra of vibration sound. *Appl Sci* 13:2845. <https://doi.org/10.3390/app13052845>
23. Cholewa M, Romaszewski M, Głomb P, Kołodziej K, Gorawski M, Koral J, Koral W, Madej A, Musioł K (2024) Just one more sensor is enough—iterative water leak localization with physical simulation and a small number of pressure sensors. *IEEE Sensors Journal*, 24(15)
24. Choudhary P, Botre BA, Akbar SA (2023) 1-D convolution neural network based leak detection, location and size estimation in smart water grid. *Urban Water J* 20:341–351. <https://doi.org/10.1080/1573062X.2022.2164732>
25. Chuang WY, Tsai YL, Wang LH (2019), March Leak detection in water distribution pipes based on CNN with mel frequency cepstral coefficients. In *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence* (pp. 83–86)
26. Cody R, Narasimhan S, Tolson B (2017) CCWI2017: F47'ONE-CLASS SVM–LEAK DETECTION IN WATER DISTRIBUTION SYSTEMS'
27. Cody RA, Dey P, Narasimhan S (2020a) Linear prediction for leak detection in water distribution networks. *J Pipeline Syst Eng Pract* 11(1):04019043
28. Cody RA, Narasimhan S (2020) A field implementation of linear prediction for leak-monitoring in water distribution networks. *Adv Eng Inform* 45:101103
29. Cody RA, Tolson BA, Orchard J (2020b) Detecting leaks in water distribution pipes using a deep autoencoder and hydroacoustic spectrograms. *J Comput Civil Eng* 34(2):04020001
30. van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84:523–538. <https://doi.org/10.1007/s11192-009-0146-3>
31. El-Abbasy MS, Mosleh F, Senouci A, Zayed T, Al-Derham H (2016) Locating leaks in water mains using noise loggers. *J Infrastruct Syst* 22(3):04016012
32. Eliades DG, Polycarpou MM (2012) Leakage fault detection in district metered areas of water distribution systems. *J Hydroinformatics* 14(4):992–1005
33. Fan X (2024) *Artificial Intelligence Aided Resilient and Sustainable Water Infrastructure Systems* (Doctoral dissertation, Case Western Reserve University)
34. Fan X, Yu X (2022) An innovative machine learning based framework for water distribution network leakage detection and localization. *Struct Health Monit* 21(4):1626–1644
35. Farah E, Shahrour I (2024) Water Leak Detection: A Comprehensive Review of Methods, Challenges, and Future Directions. *Water* 16:2975
36. Farajzadeh N, Sadeghzadeh N, Jokar N (2024) Water distribution pipe lifespans: Predicting when to repair the pipes in municipal water distribution networks using machine learning techniques. *PLOS*

37. Fares A, Tijani IA, Rui Z, Zayed T (2023) Leak detection in real water distribution networks based on acoustic emission and machine learning. *Environ Technol* 44(25):3850–3866
38. Feng J, Zhang H (2006) Algorithm of pipeline leak detection based on discrete incremental clustering method. Springer, Berlin, pp 602–607
39. Fereidooni Z, Tahayori H, Bahadori-Jahromi A (2021) A hybrid model-based method for leak detection in large scale water distribution networks. *J Ambient Intell Humaniz Comput* 12:1613–1629
40. Ferrandez-Gamot L, Busson P, Blesa J, Tornil-Sin S, Puig V, Duviella E, Soldevila A (2015) Leak localization in water distribution networks using pressure residuals and classifiers. *IFAC-PapersOnLine* 48(21):220–225
41. Fontanazza CM, Notaro V, Puleo V, Nicolosi P, Freni G (2015) Contaminant intrusion through leaks in water distribution system: experimental analysis. *Procedia Eng* 119:426–433
42. Garzón-Orduña AJ, Coronado-Hernández OE, Arrieta-Pastrana A, Ramos HM, Pérez-Sánchez M (2026) Physics-informed machine learning for hybrid modelling of water leakages induced by rapid valve manoeuvres in water distribution networks. *Eng Appl Artif Intell* 167:113887. <https://doi.org/10.1016/j.engappai.2026.113887>
43. Geng Z, Hu X, Han Y, Zhong Y (2019) A novel leakage-detection method based on sensitivity matrix of pipe flow: case study of water distribution systems. *J Water Resour Plan Manag* 145(2):04018094
44. Glynis K, Kapelan Z, Bakker M, Taormina R (2023) Leveraging Transfer Learning in LSTM Neural Networks for Data-Efficient Burst Detection in Water Distribution Systems. *Water Resour Manage* 37:5953–5972. <https://doi.org/10.1007/s11269-023-03637-3>
45. Gunatilake A, Miro JV (2024) Multimodel Neural Network for Live Classification of Water Pipe Leaks From Vibro-Acoustic Signals. *IEEE Sens J* 24:14825–14832
46. Guo, G., Yu, X., Liu, S., Xu, X., Ma, Z., Wang, X., ... and Smith, K. (2020). Novel leakage detection and localization method based on line spectrum pair and cubic interpolation search. *Water Resources Management*, 34, 3895–3911.
47. Han Z, Zhao J, Leung H, Ma KF, Wang W (2019) A review of deep learning models for time series prediction. *IEEE Sens J* 21(6):7833–7848
48. Henriques-Silva R, Duchesne S, St-Gelais F, Saran N, N., and, Schmidt AM (2023) On-Line Warning System for Pipe Burst Using Bayesian Dynamic Linear Models. *Water Resour Res* 59:e2021WR031745. <https://doi.org/10.1029/2021WR031745>
49. Hu S, Gao J, Zhong D (2023a) Multi-agent reinforcement learning framework for real-time scheduling of pump and valve in water distribution networks. *Water Supply* 23:2833–2846. <https://doi.org/10.2166/ws.2023.163>
50. Hu S, Gao J, Zhong D, Wu R, Liu L (2023b) Real-Time Scheduling of Pumps in Water Distribution Systems Based on Exploration-Enhanced Deep Reinforcement Learning. *Systems* 11:56.

<https://doi.org/10.3390/systems11020056>

51. Hu X, Han Y, Yu B, Geng Z, Fan J (2021) Novel leakage detection and water loss management of urban water supply network using multiscale neural networks. *J Clean Prod* 278:123611
52. Huang, P., Zhu, N., Hou, D., Chen, J., Xiao, Y., Yu, J., ... and Zhang, H. (2018). Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning. *Water*, 10(12), 1765.
53. Hutton C, Kapelan Z (2015) Real-time burst detection in water distribution systems using a Bayesian demand forecasting methodology. *Procedia Eng* 119:13–18
54. Islam MR, Azam S, Shanmugam B, Mathur D (2022) A review on current technologies and future direction of water leakage detection in water distribution network. *IEEE Access* 10:107177–107201
55. Javed A, Wu W, Sun Q, Dai Z (2025) Leak Management in Water Distribution Networks Through Deep Reinforcement Learning: A Review. *Water*, 17, 1928. <https://doi.org/10.3390/w17131928>
56. Joseph K, Sharma AK, van Staden R, Wasantha PLP, Cotton J, Small S (2023) Application of Software and Hardware-Based Technologies in Leaks and Burst Detection in Water Pipe Networks: A Literature Review. *Water*, 15, 2046. <https://doi.org/10.3390/w15112046>
57. Joseph K, Shetty J, Patnaik R, Matthew NS, Van Staden R, Liyanage WP, Powell G, Bennett N, Sharma AK (2025) Early Leak and Burst Detection in Water Pipeline Networks Using Machine Learning Approaches. *Water* 17:2164. <https://doi.org/10.3390/w17142164>
58. Joseph K, Shetty J, Sharma AK, van Staden R, Wasantha PLP, Small S, Bennett N (2024) Leak and Burst Detection in Water Distribution Network Using Logic- and Machine Learning-Based Approaches. *Water*, 16, 1935
59. Jun S, Lansey KE (2023a) Comparison of AMI and SCADA Systems for Leak Detection and Localization in Water Distribution Networks. *J Water Resour Plan Manag* 149:04023061. <https://doi.org/10.1061/JWRMD5.WRENG-5953>
60. Jun S, Lansey KE (2023b) Convolutional Neural Network for Burst Detection in Smart Water Distribution Systems. *Water Resour Manage* 37:3729–3743. <https://doi.org/10.1007/s11269-023-03524-x>
61. Jung D, Lansey K (2015) Water distribution system burst detection using a nonlinear Kalman filter. *J Water Resour Plan Manag* 141(5):04014070
62. Kammoun M, Kammoun A, Abid M (2022) Leak detection methods in water distribution networks: a comparative survey on artificial intelligence applications. *J Pipeline Syst Eng Pract* 13(3):04022024
63. Kang D, Lansey K (2009) Real-time demand estimation and confidence limit analysis for water distribution systems. *J Hydraul Eng* 135(10):825–837
64. Kang J, Park YJ, Lee J, Wang SH, Eom DS (2017) Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Trans Industr Electron* 65(5):4279–4289

65. Karadirek IE, Kaya-Basar E, Akdeniz T (2024) A study on pipe failure analysis in water distribution systems using logistic regression. *J Water Supply* 24:176–186
66. Karimanzira D (2023) Simultaneous Pipe Leak Detection and Localization Using Attention-Based Deep Learning Autoencoder. *Electronics* 12:4665. <https://doi.org/10.3390/electronics12224665>
67. Karray F, Garcia-Ortiz A, Jmal MW, Obeid AM, Abid M (2016) Earnpipe: A testbed for smart water pipeline monitoring using wireless sensor network. *Procedia Comput Sci* 96:285–294
68. Kemba J, Gideon K, Nyirenda CN (2017), May Leakage detection in Tsumeb east water distribution network using EPANET and support vector regression. In *2017 IST-Africa Week Conference (IST-Africa)* (pp. 1–8). IEEE
69. Komba GM, Mathonsi TE, Owolawi PA (2024) Optimizing Leak Detection and Location in Water Distribution Networks Using SVM-RF Algorithm. In *Proceedings of the 2024 15th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*, Cape Town, South Africa, 17–19 May 2024
70. Laucelli D, Romano M, Savić D, Giustolisi O (2016) Detecting anomalies in water distribution networks using EPR modelling paradigm. *J Hydroinformatics* 18(3):409–427
71. Leite R, Amado C, Azeitona M (2024) Online Burst Detection in Water Distribution Networks Based on Dynamic Shape Similarity Measure. *Expert Syst Appl* 248:123379
72. Liemberger R, Wyatt A (2019) Quantifying the global non-revenue water problem. *Water Sci Technology: Water Supply* 19(3):831–837
73. Lin H, Lin H, Fang X, Wang M, Huang L (2020), August Intelligent pipeline leak detection and analysis system. In *2020 15th International Conference on Computer Science and Education (ICCSE)* (pp. 206–210). IEEE
74. Liu M, Guo G, Xu Y, Yang Y, Liu N (2024) Performance of improved Gaussian extreme learning machine for water pipeline leak recognition. *IEEE Sens J* 24:8474–8483
75. Liu R, Xu Z, Zayed T, Liu H, Zhao Y, Chang Z (2026) Towards smart water infrastructure: Retrieval-augmented large language models for acoustic leak detection. *Water Res X* 100610. <https://doi.org/10.1016/j.wroa.2026.100610>
76. Liu R, Zayed T, Xiao R (2024a) Advanced acoustic leak detection in water distribution networks using integrated generative model. *Water Res* 254:121434
77. Liu R, Zayed T, Xiao R (2024b) Contrastive learning method for leak detection in water distribution networks. *npj Clean Water* 7:118. <https://doi.org/10.1038/s41545-024-00406-6>
78. Liu R, Zayed T, Xiao R (2025) Acoustic leak localization for water distribution network through time-delay-based deep learning approach. *Water Res* 268:122600
79. Loureiro D, Amado C, Martins A, Vitorino D, Mamade A, Coelho ST (2016) Water distribution systems flow monitoring and anomalous event detection: A practical approach. *Urban Water J* 13(3):242–252

80. Mamo TG (2014) *Virtual district metered area municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class support vector machine for risk-based asset management* (Doctoral dissertation, Polytechnic Institute of New York University)
81. Martinez-Ríos EA, Barrientos D, Bustamante R (2024) Water leakage classification with acceleration, pressure, and acoustic data: Leveraging the wavelet scattering transform, unimodal classifiers, and late fusion. *IEEE Access* 12:84923–84951
82. Marvin G, Grbčić L, Družeta S, Kranjčević L (2023) Water distribution network leak localization with histogram-based gradient boosting. *J Hydroinformatics* 25:663–684.
<https://doi.org/10.2166/hydro.2023.102>
83. Mashford J, De Silva D, Burn S, Marney D (2012) Leak detection in simulated water pipe networks using SVM. *Appl Artif Intell* 26(5):429–444
84. Mashford J, De Silva D, Marney D, Burn S (2009), October An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. In *2009 Third International Conference on Network and System Security* (pp. 534–539). IEEE
85. Mashhadi N, Shahrour I, Attoue N, El Khattabi J, Aljer A (2021) Use of machine learning for leak detection and localization in water distribution systems. *Smart Cities* 4(4):1293–1315
86. Mazaev G, Weyns M, Moens P, Haest PJ, Vancoillie F, Vaes G, Debaenst J, Waroux A, Marlein K, Ongenaë F, Van Hoecke S (2023) A microservice architecture for leak localization in water distribution networks using hybrid AI. *J Hydroinformatics* 25:851–866.
<https://doi.org/10.2166/hydro.2023.147>
87. McMillan L, Fayaz J, Varga L (2023) Flow Forecasting for Leakage Burst Prediction in Water Distribution Systems Using Long Short-Term Memory Neural Networks and Kalman Filtering. *Sustainable Cities Soc* 99:104934. <https://doi.org/10.1016/j.scs.2023.104934>
88. McMillan L, Fayaz J, Varga L (2024) Domain-Informed Variational Neural Networks and Support Vector Machines Based Leakage Detection Framework to Augment Self-Healing in Water Distribution Networks. *Water Res* 249:120983
89. Mohan Doss P, Rokstad MM, Tscheikner-Gratl F (2024) The performance of encoder–decoder neural networks for leak detection in water distribution networks. *Water Supply* 24(8):2750–2764
90. Molina JL, Patino-Alonso C, Wan X, Farmani R (2025) StocHastIc Early Leakage Detection System (SHIELDS) for Water Distribution Networks. *Water Resour Manage* 39:4189–4204
91. Mounce SR, Boxall JB, Machell J (2010) Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *J Water Resour Plan Manag* 136(3):309–318
92. Mounce SR, Mounce RB, Boxall JB (2011) Novelty detection for time series data analysis in water distribution systems using support vector machines. *J Hydroinformatics* 13(4):672–686
93. Nagaraj A, Kotamreddy GR, Choudhary P, Katiyar R, Botre BA (2021) Leak detection in smart water grids using EPANET and machine learning techniques. *IETE J Educ* 62(2):71–79

94. Nam K, Ifaei P, Heo S, Rhee G, Lee S, Yoo C (2019) An efficient burst detection and isolation monitoring system for water distribution networks using multivariate statistical techniques. *Sustainability* 11(10):2970
95. Nascimento WMD, Gomes-Jr L (2023) Enabling Low-Cost Automatic Water Leakage Detection: A Semi-Supervised, autoML-Based Approach. *Urban Water J* 20:1471–1481. <https://doi.org/10.1080/1573062X.2022.2056710>
96. Negm A, Ma X, Aggidis G (2023a) Review of leakage detection in water distribution networks. *IOP Conference Series: Earth and Environmental Science*, 1136, 012052. <https://doi.org/10.1088/1755-1315/1136/1/012052>
97. Negm A, Ma X, Aggidis G (2023b) Water Pressure Optimisation for Leakage Management Using Q Learning. In *2023 IEEE Conference on Artificial Intelligence (CAI)*, 270–271. <https://doi.org/10.1109/CAI54212.2023.00120>
98. Negm A, Ma X, Aggidis G (2024) Deep Reinforcement Learning Challenges and Opportunities for Urban Water Systems. *Water Res* 253:121145
99. Nimri W, Wang Y, Zhang Z, Deng C, Sellstrom K (2023) Data-driven approaches and model-based methods for detecting and locating leaks in water distribution systems: a literature review. *Neural Comput Appl* 35:11611–11623. <https://doi.org/10.1007/s00521-023-08497-x>
100. Noori A, Roshani E, Bonakdari H (2026) Aging water distribution networks: a hybrid spatial decision support and machine learning framework for leak detection. *Environ Syst Decisions* 46(3):59
101. Oberascher M, Brentan BM, Menapace A, Herrera M, Fu G, Taormina R, Sitzenfrei R (2026) Assessing supervised machine learning practice in urban water networks: A critical review of methodological transparency, reproducibility and reporting. *Water Res X* 32:100569. <https://doi.org/10.1016/j.wroa.2026.100569>
102. Okeya I, Kapelan Z, Hutton C, Naga D (2014) Online burst detection in a water distribution system using the Kalman filter and hydraulic modelling. *Procedia Eng* 89:418–427
103. Onukwube CU, Aikhuele DO, Sorooshian S (2024) Development of a fault detection and localization model for a water distribution network. *Appl Sci* 14:1620
104. Palau CV, Arregui FJ, Carlos M (2012) Burst detection in water networks using principal component analysis. *J Water Resour Plan Manag* 138(1):47–54
105. Parajuli U, Shin S (2024) Identifying failure types in cyber-physical water distribution networks using machine learning models. *Aqua—Water Infrastructure Ecosyst Soc* 73:504–519
106. Pham Vu Hong S, Nguyen Thanh V (2023) Application of artificial intelligence algorithm to optimize the design of water distribution system. *Int J Constr Manage* 23(16):2830–2840
107. Quiñones-Grueiro M, Verde C, Prieto-Moreno A, Llanes-Santiago O (2018) An unsupervised approach to leak detection and location in water distribution networks. *Int J Appl Math Comput Sci* 28(2):283–295
108. Rahimi M, Alghassi A, Ahsan M, Haider J (2020) Deep learning model for industrial leakage detection using acoustic emission signal. *Informatics*, vol 7. MDPI, p 49. 4

109. Rajabi MM, Komeilian P, Wan X, Farmani R (2023) Leak detection and localization in water distribution networks using conditional deep convolutional generative adversarial networks. *Water Res* 238:120012. <https://doi.org/10.1016/j.watres.2023.120012>
110. Rajan G, Li S (2025) Water Distribution Syst Smart Cities 8:78. <https://doi.org/10.3390/smartcities8030078>. A Systematic Literature Review on Flow Data-Based Techniques for Automated Leak Management
111. Rajasekaran U, Kothandaraman M (2024) A novel custom one-dimensional time-series DenseNet for water pipeline leak detection and localization using acousto-optic sensor. *IEEE Access* 12:7966–7973
112. Rajeswaran A, Narasimhan S, Narasimhan S (2018) A graph partitioning algorithm for leak detection in water distribution networks. *Comput Chem Eng* 108:11–23
113. Ranganathan S, Vinothkumar N, Muthuramalingam E (2026) Benchmarking diverse learning architectures for leak detection and localization in water distribution systems. *J Water Process Eng* 87:110114. <https://doi.org/10.1016/j.jwpe.2026.110114>
114. Rashid S, Akram U, Khan SA (2015) WML: Wireless sensor network based machine learning for leakage detection and size estimation. *Procedia Comput Sci* 63:171–176
115. Ravichandran T, Gavahi K, Ponnambalam K, Burtea V, Mousavi SJ (2021) Ensemble-based machine learning approach for improved leak detection in water mains. *J Hydroinformatics* 23(2):307–323
116. Romano M, Kapelan Z, Savić DA (2014) Automated detection of pipe bursts and other events in water distribution systems. *J Water Resour Plan Manag* 140(4):457–467
117. Romero-Ben L, Alves D, Blesa J, Cembrano G, Puig V, Duviella E (2023) Leak detection and localization in water distribution networks: Review and perspective. *Annu Rev Control* 55:392–419
118. Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* 10(3):e0118432. <https://doi.org/10.1371/journal.pone.0118432>
119. Satterlee N, Zuo X, Lee CW, Park CW, Kang JS (2025) Parallel multi-layer sensor fusion for pipe leak detection using multi-sensors and machine learning. *Eng Appl Artif Intell* 153:110923
120. Serafeim AV, Fourniotis NT, Deidda R, Kokosalakis G, Langousis A (2024) Leakages in Water Distribution Networks: Estimation Methods, Influential Factors, and Mitigation Strategies—A. *Compr Rev Water* 16:1534
121. Shekofteh M, Jalili Ghazizadeh M, Yazdi J (2020) A methodology for leak detection in water distribution networks using graph theory and artificial neural network. *Urban Water J* 17(6):525–533
122. Siddique MF, Ahmad Z, Ullah N, Ullah S, Kim JM (2024) Pipeline leak detection: A comprehensive deep learning model using CWT image analysis and an optimized DBN-GA-LSSVM framework. *Sensors* 24:4009
123. Sitaropoulos K, Salamone S, Sela L (2023) Frequency-based leak signature investigation using acoustic sensors in urban water distribution networks. *Adv Eng Inform* 55:101905. <https://doi.org/10.1016/j.aei.2023.101905>

124. Soldevila A, Blesa J, Fernandez-Canti RM, Tornil-Sin S, Puig V (2019) Data-driven approach for leak localization in water distribution networks using pressure sensors and spatial interpolation. *Water* 11(7):1500
125. Soldevila A, Blesa J, Tornil-Sin S, Duviella E, Fernandez-Canti RM, Puig V (2016) Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Eng Pract* 55:162–173
126. Soldevila A, Fernández-Canti RM, Blesa J, Tornil-Sin S, Puig V (2017) Leak localization in water distribution networks using Bayesian classifiers. *J Process Control* 55:1–9
127. Sourabh N, Timbadiya PV, Patel PL (2023) Leak detection in water distribution network using machine learning techniques. *ISH J Hydraulic Eng* 29(Suppl 1):177–195.
<https://doi.org/10.1080/09715010.2023.2198988>
128. Su HN, Lee PC (2010) Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *scientometrics*, 85(1), 65–79
129. Sun C, Parellada B, Puig V, Cembrano G (2019) Leak localization in water distribution networks using pressure and data-driven classifier approach. *Water* 12(1):54
130. Tao T, Huang H, Li F, Xin K (2014) Burst detection using an artificial immune network in water-distribution systems. *J Water Resour Plan Manag* 140(10):04014027
131. Wan X, Farmani R, Keedwell E (2023) Online Leakage Detection System Based on EWMA-Enhanced Tukey Method for Water Distribution Systems. *J Hydroinformatics* 25:51–69.
<https://doi.org/10.2166/hydro.2022.079>
132. Wang J, Fu G, Savic D (2026) EPANET-Agentica: A multi-agent system for natural language-controlled simulations of water distribution networks. *Water Res* 293:125433.
<https://doi.org/10.1016/j.watres.2026.125433>
133. Wang X, Guo G, Liu S, Wu Y, Xu X, Smith K (2020) Burst detection in district metering areas using deep learning method. *J Water Resour Plan Manag* 146(6):04020031
134. World Health Organization (WHO) and United Nations Children’s Fund (UNICEF) (2026) State of systems for drinking-water, sanitation and hygiene: Global update 2025. World Health Organization. ISBN 978-92-4-011898-0
135. Wu J, Ma D, Wang W (2022) Leakage identification in water distribution networks based on XGBoost algorithm. *J Water Resour Plan Manag* 148(3):04021107
136. Wu Y, Liu S (2017) A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J* 14(9):972–983
137. Wu Y, Liu S (2020) Burst detection by analyzing shape similarity of time series subsequences in district metering areas. *J Water Resour Plan Manag* 146(1):04019068
138. Wu Y, Liu S, Wu X, Liu Y, Guan Y (2016) Burst detection in district metering areas using a data driven clustering algorithm. *Water Res* 100:28–37

139. Wu Y, Ma X, Guo G, Huang Y, Liu M, Liu S, Zhang J, Fan J (2023a) Hybrid method for enhancing acoustic leak detection in water distribution systems: Integration of handcrafted features and deep learning approaches. *Process Saf Environ Prot* 177:1366–1376.
<https://doi.org/10.1016/j.psep.2023.08.011>
140. Wu Y, Ma X, Guo G, Jia T, Huang Y, Liu S, Fan J, Wu X (2024) Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective. *Water Res* 261:121999
141. Wu ZY, Chew A, Meng X, Cai J, Pok J, Kalfarisi R, Lai KC, Hew SF, Wong JJ (2023b) High Fidelity Digital Twin-Based Anomaly Detection and Localization for Smart Water Grid Operation Management. *Sustainable Cities Soc* 91:104446. <https://doi.org/10.1016/j.scs.2023.104446>
142. Xie Y, Gao M, Luo F, Zhou A, Yang Y, Hu J, Jiang W, Ye Y (2025) Django-based framework database for leakage detection using machine learning for water distribution networks. *Eng Appl Artif Intell* 149:110525
143. Xu T, Zeng Z, Huang X, Li J, Feng H (2021) Pipeline leak detection based on variational mode decomposition and support vector machine using an interior spherical detector. *Process Saf Environ Prot* 153:167–177
144. Xu Z, Liu H, Fu G, Zeng Y, Li Y (2024) Feature selection of acoustic signals for leak detection in water pipelines. *Tunn Undergr Space Technol* 152:105945
145. Xu Z, Liu H, Fu G, Zheng R, Zayed T, Liu S (2025) Interpretable deep learning for acoustic leak detection in water distribution systems. *Water Res* 273:123076.
<https://doi.org/10.1016/j.watres.2024.123076>
146. Ye G, Fenner RA (2011) Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *J Pipeline Syst Eng Pract* 2(1):14–22
147. Young P, Tych W (1996) Operational control of water systems: Structures, algorithms and applications. By MA Brdys and B. Ulanicki
148. Yu J et al (2021) An integrated bottom-up approach for leak detection in water distribution networks based on assessing parameters of water balance model. *Water*, 13(6)
149. Yussif A-M, Sadeghi H, Zayed T (2023) Application of Machine Learning for Leak Localization in Water Supply Networks. *Buildings* 13:849. <https://doi.org/10.3390/buildings13040849>
150. Zaman M, Tantawy A, Abdelwahed S (2023) Optimizing Smart City Water Distribution Systems Using Deep Reinforcement Learning. In 2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), 228–233.
<https://doi.org/10.1109/HONET59747.2023.10374894>
151. Zhang Q, Wu ZY, Zhao M, Qi J, Huang Y, Zhao H (2016) Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *J Water Resour Plan Manag* 142(11):04016042
152. Zhou B, Lau V, Wang X (2019a) Machine-learning-based leakage-event identification for smart water supply systems. *IEEE Internet Things J* 7(3):2277–2292

153. Zhou M, Pan Z, Liu Y, Zhang Q, Cai Y, Pan H (2019b) Leak detection and location based on ISLMD and CNN in a pipeline. IEEE Access 7:30457–30464

154. Zhou M, Zhang Q, Liu Y, Sun X, Cai Y, Pan H (2019c) An integration method using kernel principal component analysis and cascade support vector data description for pipeline leak detection with multiple operating modes. Processes 7(10):648

155. Zuñiga-Uribe M, Rojas-Galván R, Álvarez-Alvarado JM, Aviles M, Pérez-Soto GI, Pérez-Moreno V (2026) Artificial Intelligence in Water Distribution Networks: A Systematic Review of Models, Input Variables, Databases, and Output Strategies for Leak Detection. Smart Cities 9:45.
<https://doi.org/10.3390/smartcities9030045>

Figures

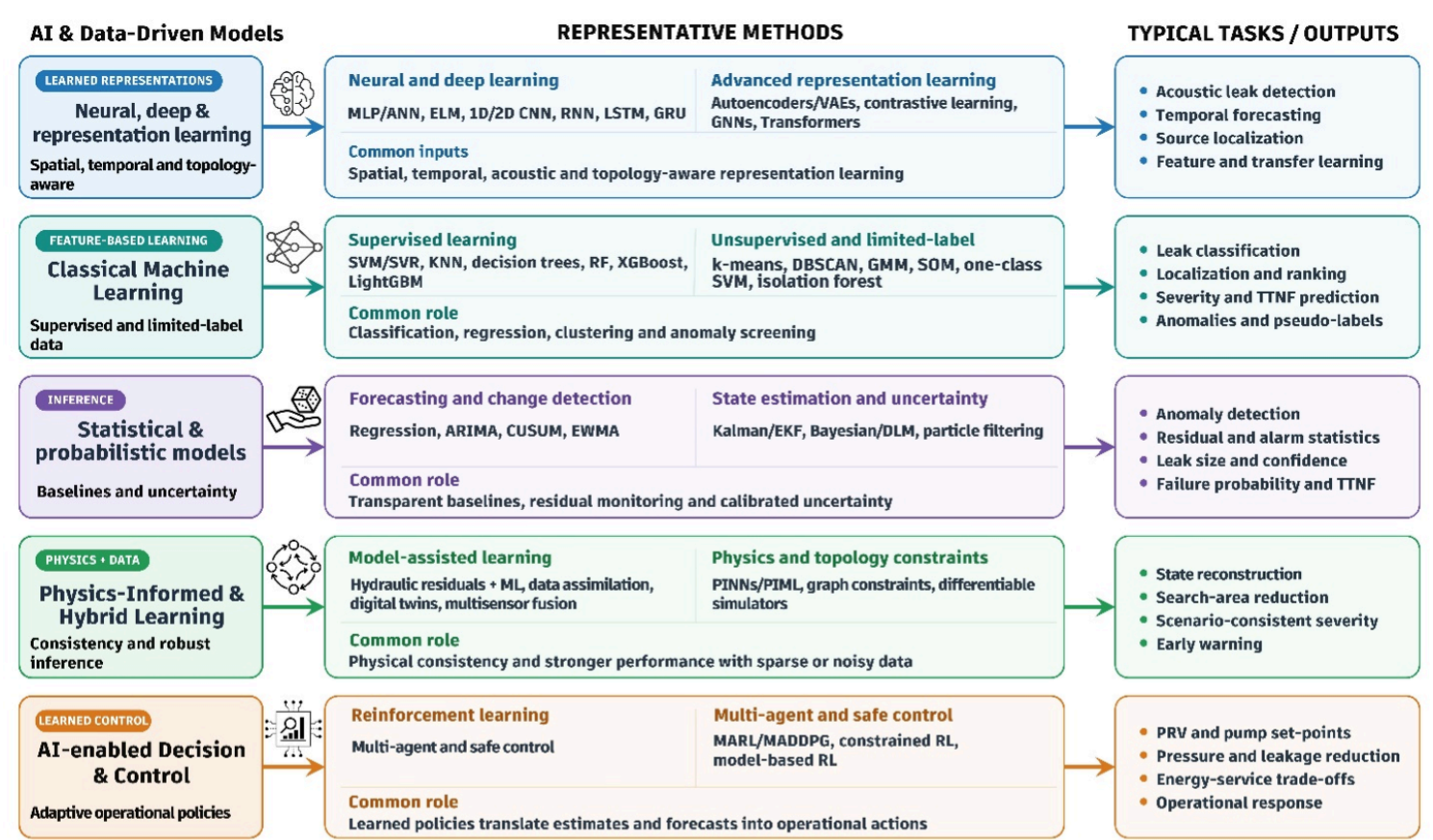


Figure 1

Taxonomy of AI and data-driven approaches for leakage management in WDSs.

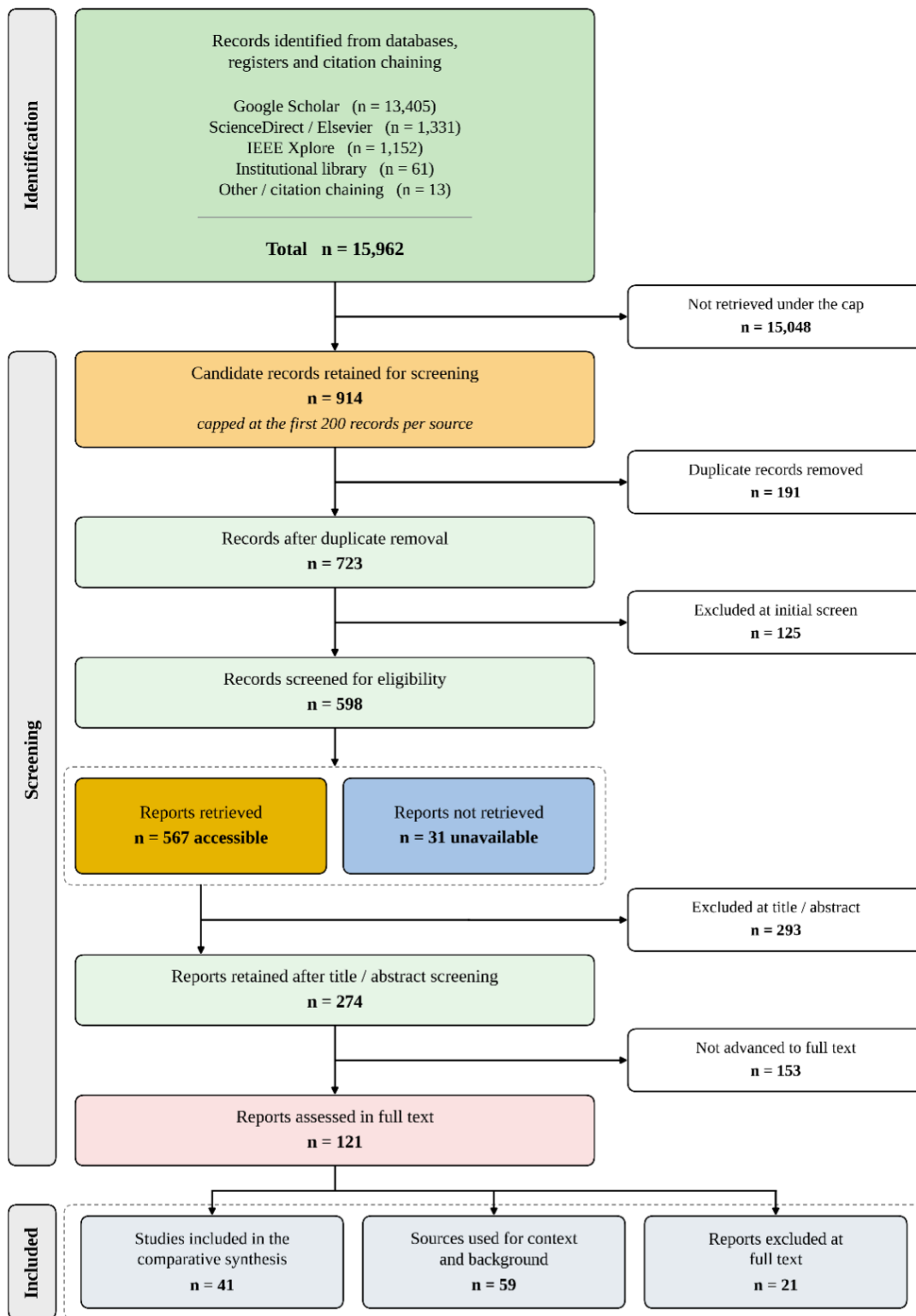


Figure 2

Search and screening procedure used to identify leakage-management studies.

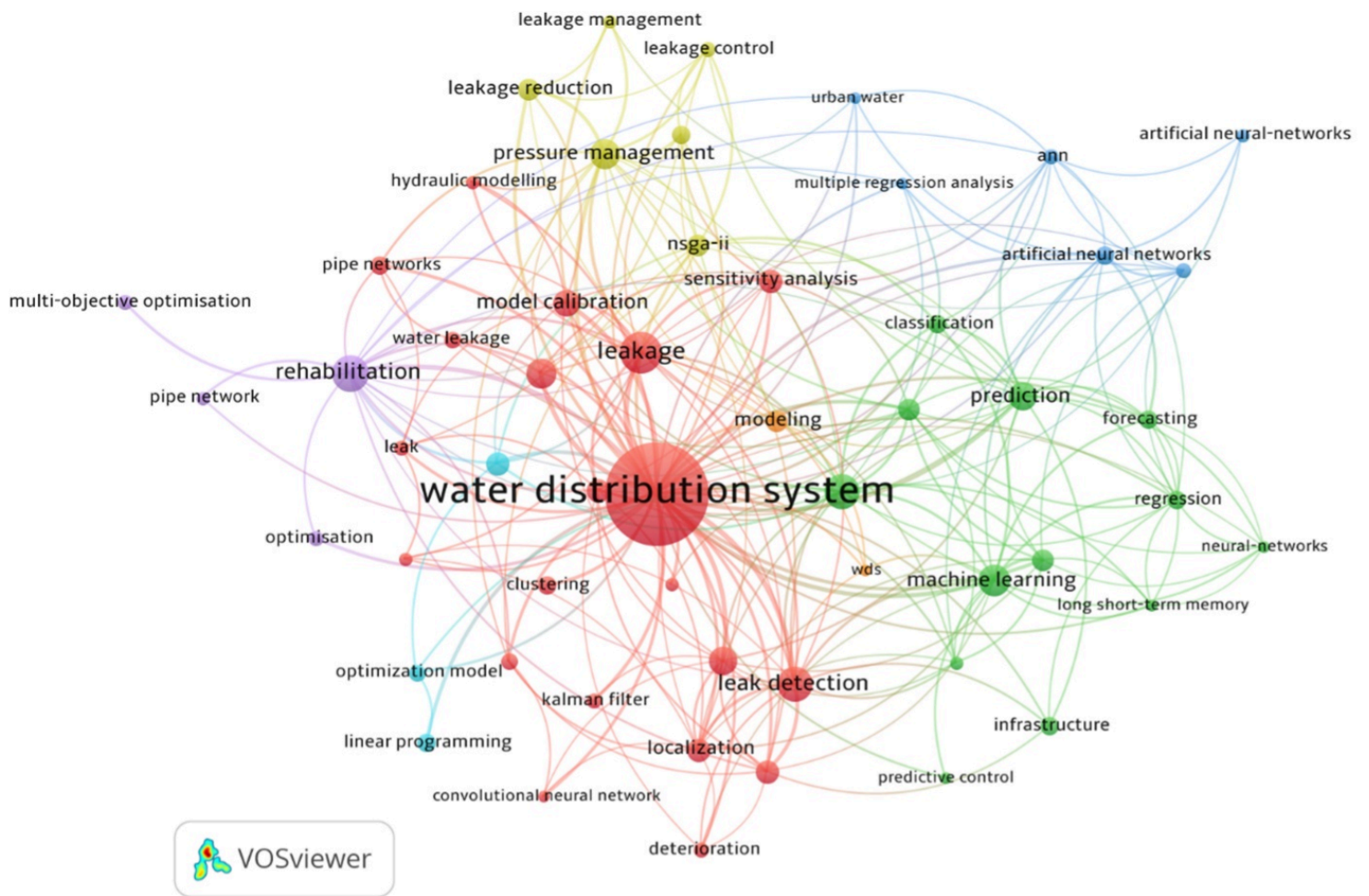


Figure 3

Literature clusters identified through keyword co-occurrence analysis using VOSviewer (version 1.6.17).

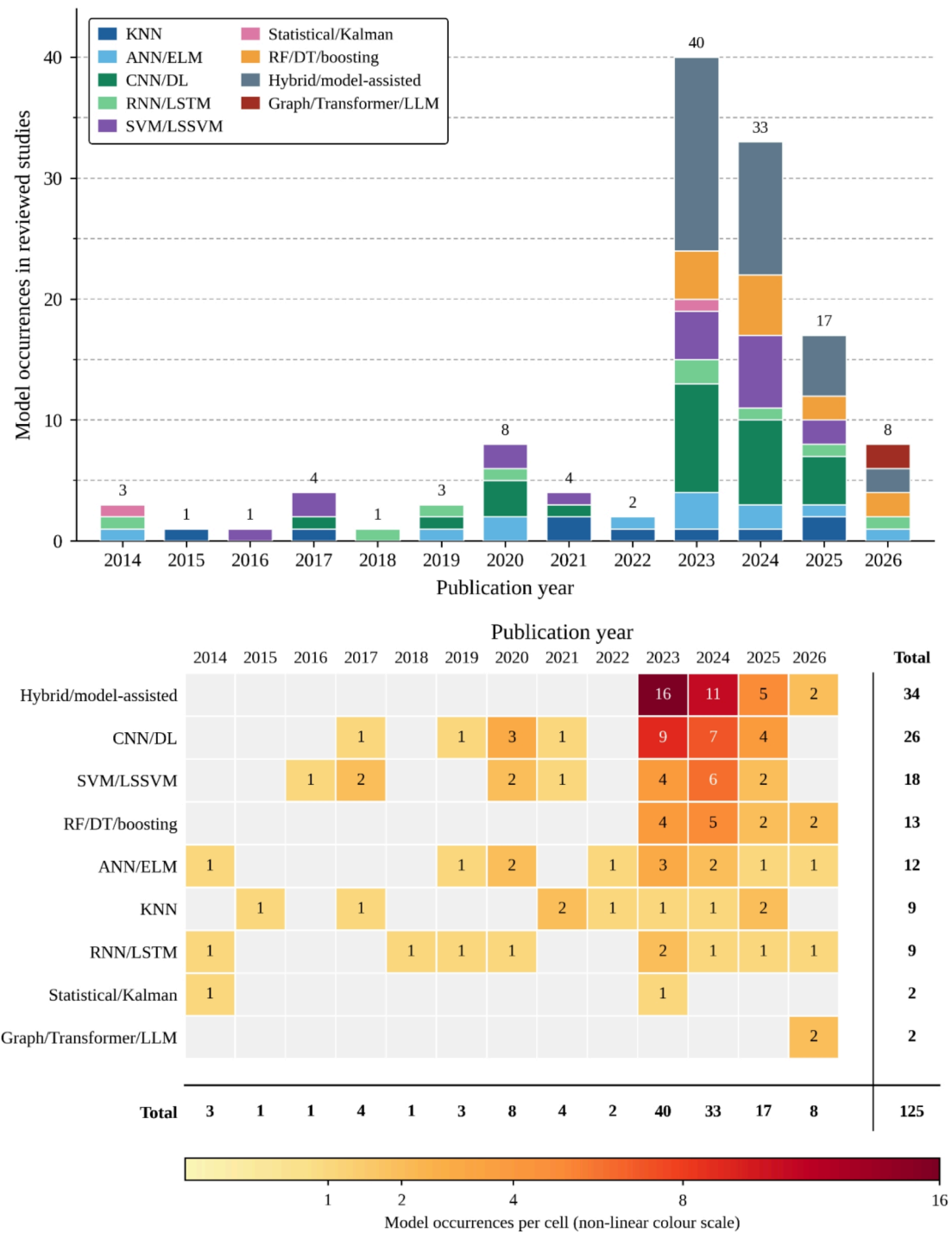


Figure 4

Occurrences of AI and data-driven model from 2014–2026.

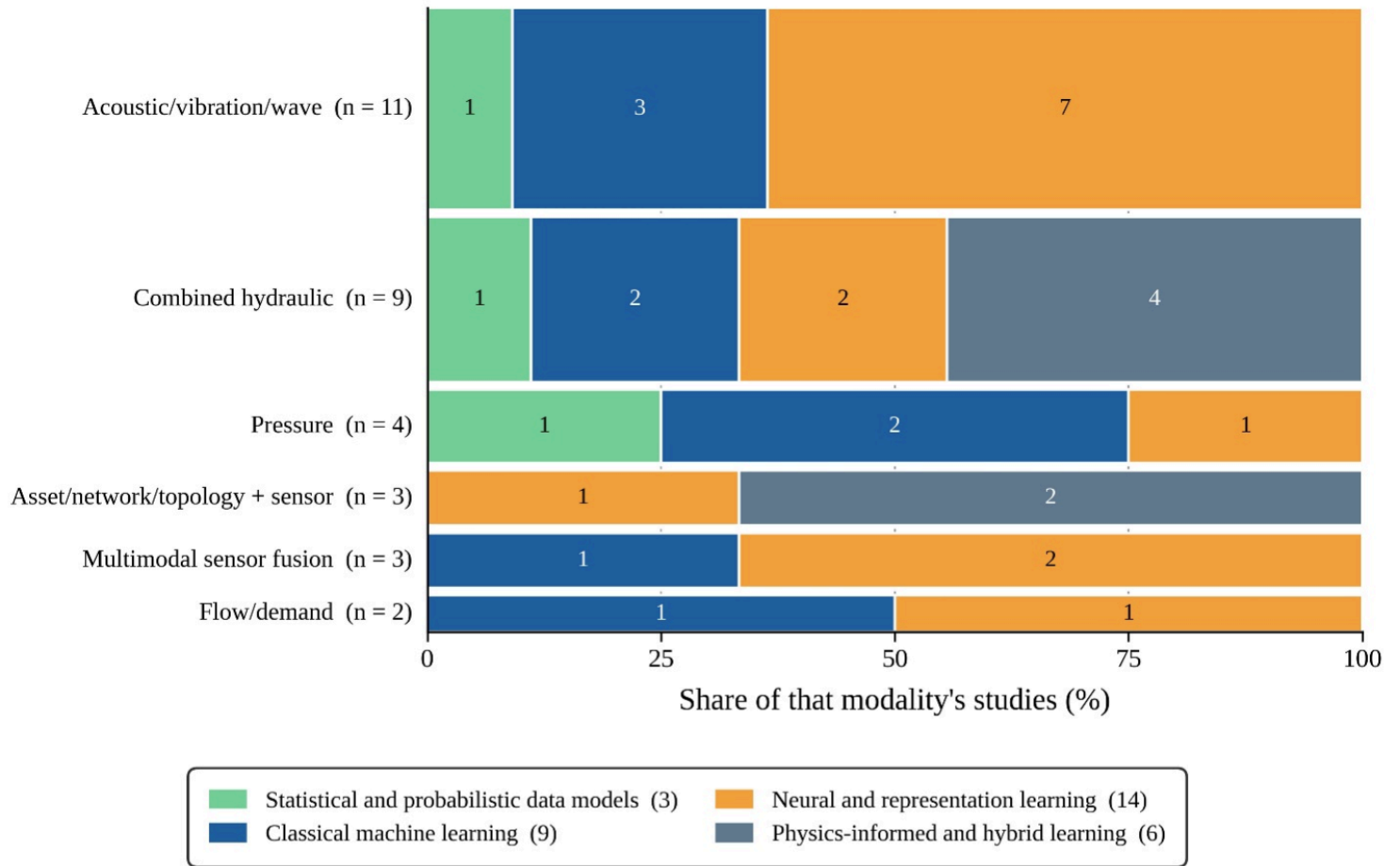


Figure 5

Input modalities and analytical approaches among primary studies.