REVIEW

Data-driven approaches and model-based methods for detecting and locating leaks in water distribution systems: a literature review

Waid Nimri[1] · Yong Wang[1] · Ziang Zhang[2] · Chengbin Deng[3] · Kristofor Sellstrom[4]

Abstract

Water distribution systems are made up of interconnected components that should allow water systems to meet demand, but leaks can waste enough water to limit supply. To limit financial losses, water utilities must quickly determine that a leak is occurring and where it is referred to as the localization of the leak. Over the years, there have been various methods proposed to detect and locate leaks. This literature review summarizes many of the methodologies introduced, categorizes them into data-driven approaches and model-based methods, and reviews their performance. Data-driven approaches demand efficient exploitation and use of available data from pressure and flow devices, and model-based methods require finely calibrated hydraulic models to reach a verdict. Data-driven approaches can manage uncertainty better than model-based methods.

Keywords Leak detection · Water distribution systems · Data-driven approaches · Deep learning · Model-based methods · Performance measures

Abbreviations

ACO	Ant colony optimization
ANN	Artificial neural network
AWWA	American Water Works Association
BLIFF	Burst location identification framework by fully linear DenseNet
BN	Bayesian network
BPNN	Back-propagation neural network
Cas-SVDD	Cascade support vector data description
CNN	Convolutional neural network
CtL-SSL	Clustering-then-localization semi-supervised learning
CUSUM	Cumulative sum
DenseNet	Densely connected convolutional networks
DL	Deep learning
DMA	District metering areas
EM	Expectation maximization
EWMA	Exponential weighted moving average
ExSem	Expert structural expectation–maximization
FLS	Fuzzy logic system
FNR	False-negative rate
FPR	False-positive rate
IDW	Inverse distance weighted
ISLMD	Improved spline-local mean decomposition
IWA	International water association
KPCA	Kernel principal component analysis
LDA	Linear discriminant analysis
LKF	Linear Kalman filter
LMD	Local mean decomposition
LP	Local polynomial
LS	Least squares
LSS	Leak signature space
LSTM	Long short-term memory
MDN	Mixture density network
MLPNN	Multilayer perceptron neural network
MDN	Mixture density network
MNF	Minimum night flow
NKF	Nonlinear Kalman filter
NRW	Non-revenue water
OC	Ordinary cokriging
OK	Ordinary kriging

✉ Yong Wang
  yongwang@binghamton.edu

[1] Department of Systems Science and Industrial Engineering, Binghamton University, Binghamton, NY 13902, USA

[2] Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902, USA

[3] Department of Geography, Binghamton University, Binghamton, NY 13902, USA

[4] Jamestown Board of Public Utilities, Jamestown, NY 14701, USA

PCA	Principal component analysis
PKF	Predictive Kalman filter
RMSE	Root-mean-square error
RNN	Recurrent neural network
SCADA	Supervisory control and data acquisition
SIP	Standardized innovation process
SOM	Self-organizing maps
SPC	Statistical process control
SVR	Support vector machine
TDOA	Time difference of arrival
TFR	Transient frequency response
TNR	True-negative rate
TPR	True-positive rate
WDS	Water distribution system

1 Introduction

Water loss has received worldwide attention because of the continuing concern of the water industry regarding rapidly increasing population demands.

The IWA states that NRW was 48.6 billion cubic meters per year divided between residential and commercial losses [1] and that number grew to 126 billion cubic meters according to a 2019 study [2]. While there are many components of NRW, the biggest losses come from the pipelines, reaching 30% of the water pumped by the WDS [3]. About 10% of the supplied water is wasted in residential areas [4]. Therefore, it is important to detect and manage leaks as soon as possible to either repair or replace pipelines as a preventive measure [5]. The negative effects of leaks cover many aspects such as public health, in which water contamination is likely to happen through pathogen intrusion [6], the economy, in which utilities lose considerable revenue [7], dissatisfaction among the consumers from supply interruption [8], and serious environmental damage [9].

Avoiding water leaks in WDSs is challenging; most distribution systems in the USA are older than 50 years, which is over the typical design period. However, minimizing damages caused by leaks and improving the performance of the water supply system are doable [9] by detecting leaks as timely as possible and reducing NRW. Lambert categorizes water leaks into three classifications: reported, unreported, and background [10]. Reported leaks are easily repaired because consumers report water that rises to the surface of the ground. The situation differs for reported and background leaks because the utility, or a third-party company, needs to continually look for anomalies. There are situations where some small leaks are unidentified and may run unnoticed for months.

One of the most commonly used methods by water utilities is MNF [11], which compares flow rates with flow thresholds. Unusual jumps that signify leaks trigger an alarm as the flow goes beyond the threshold; flows should be minimal at night, which makes the process of identification smoother and valid. This cumulative pressure on water availability is a motive to consider the optimal management of clean water resources [12].

The IWA launched DMA to help with constant supervision of water consumption [13]. In collaboration with the AWWA, they initiated an effort to develop an effective water balance-based method under the name "*The Water Audit Method.*" The method helps utilities manage supply efficiently under the ongoing pressure of a growing population, urbanization, industrialization, and other global pressures [14] without having to be weighed down by financial liabilities produced by physical losses. Since *The Water Audit Method* launched in the latter years of the past century, many methods have been developed that consider how the technical infrastructure is constantly evolving and how utilities collect massive amounts of data from the SCADA or other telemetry systems.

Many hardware- and software-based methods have been developed to detect bursts and leaks [15]. Hardware methods depend on sensors and devices [16] and are split between acoustic detection methods (e.g., leak correlator, leak noise loggers, and listening rods) and non-acoustic detection methods (e.g., ground penetrating radar and gas injection). While hardware-based methods have proven efficient and accurate in locating leaks and bursts, they are time and labor-intensive, in addition, such manual methods lose effectiveness as the depth increases [11, 17]. On the other hand, software-based methods are faster and less expensive [18] and are divided into two categories: numerical or model-based (e.g., time-domain analysis) and non-numerical or data-driven modeling methods (e.g., artificial neural networks and support vector regression).

Much effort has been invested in constructing systems capable of detecting leaks and precisely locating them in their distribution networks through a combination of mathematical models and machine learning algorithms; these techniques and methods fall under either transient-based, model-based, or data-driven approaches. Transient-based approaches find leaks by measuring pressure signals with frequent tests [19] and can locate some leaks. Model-based approaches employ mathematical formulas and first principles modeling to analyze data and discover links between the metrics and process states [16, 20] and are successful in finding leaks. Data-driven approaches do not demand a full understanding of water networks or the deciphering of the micro-interactions that take place in

normal WDSs. Data-driven modeling can deliver results based on signal processing or statistical analysis, sometimes without needing deterministic models [12, 18].

Leak detection in WDSs is challenging given the current technologies. This review presents some techniques and approaches for novelty detection so that the utility acts as quickly as possible to alleviate the adverse impact of leaks and provide uninterrupted supply to consumers. The rest of the review is outlined as follows: Sect. 2 discusses the methodology used. Section 3 goes into depth into data-driven approaches, and Sect. 4 discusses model-based methods. Section 5 assesses the performance of many discussed methods. Section 6 summarizes problems with the approaches, and Sect. 7 concludes the review.

2 Research methodology

The research methodology follows PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21]. This systematic review reflects and considers peer-reviewed research papers from 2002 through 2021. The identification phase was completed using Web of Science and Google Scholar databases from different journals and conferences. The keywords used in this literature research are "water," "leak detection," and "data-driven" combined with "deep learning" and "model-based" in different combinations. The initial identification from different search engines identified 319 possibilities. Twenty-six duplicates were removed, leaving 293. We further excluded 89. Then, we were unable to retrieve 120 studies, leaving 84 for additional screening. Finally, 18 studies were eliminated for some reason, leaving 66 to be reviewed in detail.

Figure 1 shows the exact procedure followed from the identification of relevant papers to the final step of inclusion. Figure 2 shows the distribution of the papers included in this section against the publication year, and Fig. 3 shows that 71.1% are focused on data-driven approaches. The exclusion criteria were as follows:

I. Any article that is not related to the body of the review
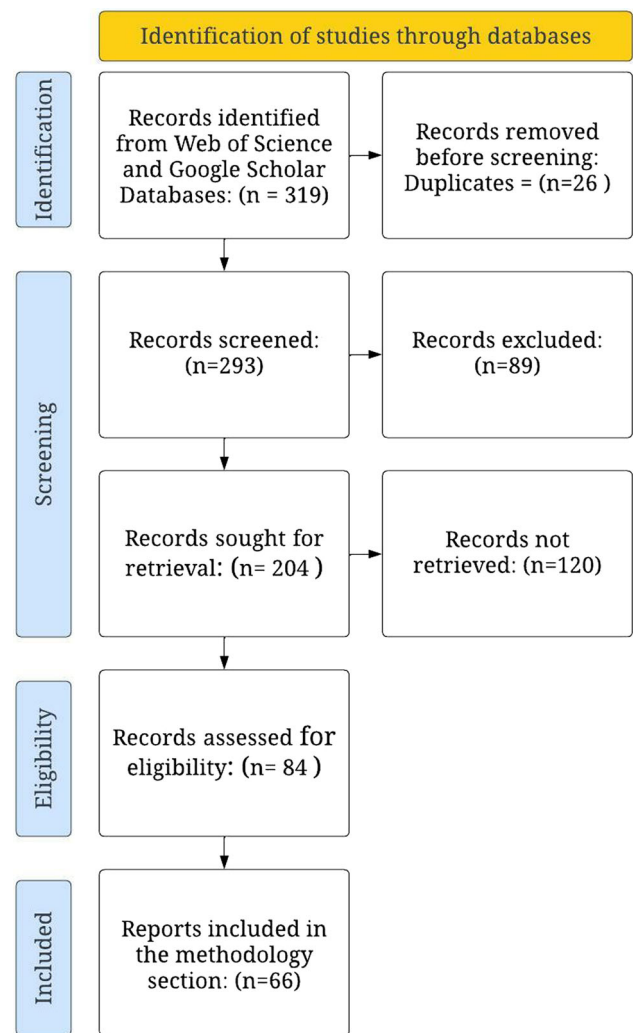II. Papers in languages other than English

Fig. 1 Research methodology diagram based on the PRISMA framework

3 Data-driven methods and techniques used for burst and leak detection

Data-driven approaches are divided into four main categories. Figure 4 illustrates a generic framework for data-driven approaches.

One of the earliest researches this century related to burst and leak detection was by Mounce in [22]. Their study used a neural network to detect bursts in the WDS. Since then, there has been a dramatic growth in the number of studies relating to the topic. Mounce and all used ANNs to identify leaks by building an empirical model that classifies and predicts when such an event occurs from time-series data produced by sensors; this automated computing structure acknowledged the complexity and non-linearity of the water distribution system and solved the problem by decomposing the network and constructing

Fig. 2 Number of publications included in the research methodology section
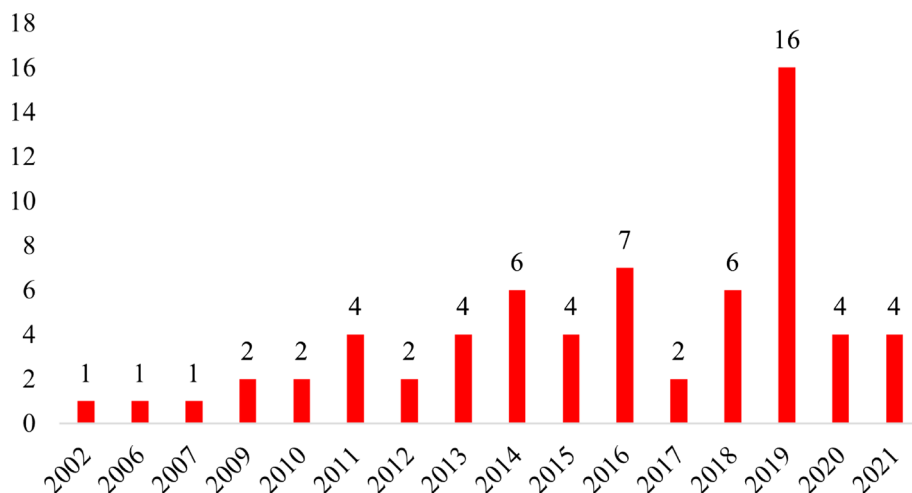
Data-driven approaches vs Model-based methods

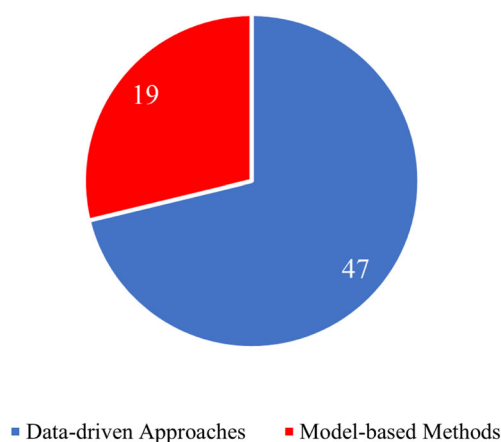- Data-driven Approaches
- Model-based Methods

Fig. 3 Number of data-driven approaches and model-based methods included in the research methodology section

more than one neural network to supervise zones in the DMA [22].

In a study published in 2006, Mounce et al. tested static and time delay architectures of neural networks to classify bursts that were simulated through hydrant flushing. The results showed that time delay NNs are more accurate than static neural networks. Time delay NNs feature a memory element and can learn patterns through time. That feature is lacking in static NNs. The outperformance was solidified by the results that showed that time delay NNs identified 75% of bursts, while the static architecture could detect only 4% of the simulated events [23].

The theoretical basis of applying FLS and engaging ANNs in the form of MDNs is to draw conclusions about abnormalities from a data stream. This implementation of FLS requests flow data, and it results in an error of 10%, which is smaller than the error resulting from MNF [24].

SOMs employed a leak function to help ease detection [25]. Mounce and others presented an online application of artificial intelligence to detect leaks and other abnormal events, and they verified the ability to detect small and medium leaks with an exceptionally small number of ghosts [26].

Mounce et al. [27] used SVM, a supervised statistical learning machine learning model, to model and predict from time-series data. When comparing the SVM model to ANNs, SVMs appear to have less sensitivity to noise, making it more practical according to [28].

Romano et al. [84, 85] used artificial neural networks, statistical process control, and a Bayesian network to forecast pressure and flow values, compare the actual and expected values, and finally classify the signals as normal or leak. Two years later, Romano et al. [29] improved their methodology to approximately locate bursts by using geostatistical techniques and a multivariate Gaussian graphical model and overcame the shortcoming of the probability that accompanies the increasing number of accessible data. In that paper, they used four geostatistical techniques, OC, OK, LP, and IDW, and of the four techniques, OC provided the smallest root-mean-square errors in all the five scenarios studied.

The LKF method developed by Ye and Fanner is based on regression and Kalman filtering of flow and pressure measurements, and the results obtained show that flow data are more sensitive to bursts [30]. Researchers argue that LKF has the advantage of efficient computations that make it an effective and practical method to expose leaks [14].

Palau et al. [31] performed PCA to detect severe outliers. Pattern matching methods and auto-associative neural networks for novelty detection showed promising results in exploring enormous time-series datasets very quickly despite the limitations in pattern libraries [12]. Fourier series and CUSUM are used for leak detection. The former

Fig. 4 Basic framework of data-driven approaches
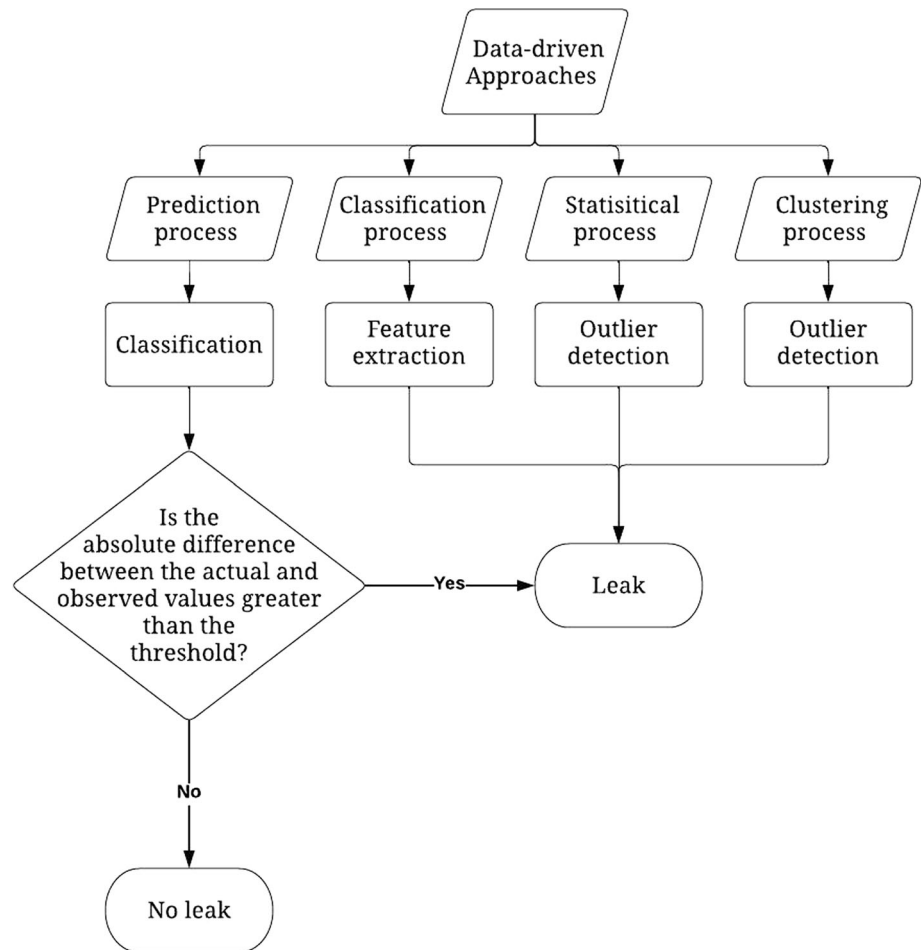
learns the flow dynamics, and the latter is used for detection. Eliades et al. [32] presented their solution in addition to the night flow analysis for comparison, Fourier series and CUSUM outperformed the night flow analysis and scored a TPR of 94.5%, whereas the highest TPR the NF analysis achieved is 79.5%. The combined Fourier series and CUSUM method may miss small leaks unless a proper detection threshold is chosen to weed out background noises. The method, however, does not require sensors that sample pressures frequently, which most systems do not have. It also avoids misclassifying existing small leaks when the analysis starts.

Weighted least squares with the EM algorithm is an unbiased unsupervised procedure, and it detects a leak based on the difference between the measured flow and the expected flow [33]. Bekker et al. [34] developed an investigative approach where the setup starts with producing the expected values of demand and pressure, and then compares the generated values with the measured values. The deviations indicate a burst or leak, and to continuously supervise the network, a threshold value is established based on the previous year's data. This method

was applied on three networks in three different areas, and the performance varied based on area. As the area covered by a flow meter gets larger, the size of the leak must grow to be detected. For example, the Rhine area in the Netherlands was studied in the paper, and only leaks that exceeded 150 m^3/h were detected. To move past such an obstacle, investment in more flow meters to divide into smaller areas is recommended.

The probabilistic demand forecasting model distinguishes abnormalities, e.g., bursts, based on previously documented data under the assumption of being normal (no bursts). The methodology identifies a burst based on the forecasted and actual values, and when the probability of the actual demand is larger than that of the forecasted one by a probability distribution function (PDF) then a burst exists [8, 35]. The methodology worked better at night with bursts greater than 5% of the mean flow.

NKF overcame the drawback of LKF by addressing the current system's condition. Jung and Lansey [36] compared their methodology to two statistical process control methods, i.e., CUSUM and Hotelling T^2. NKF, which uses SIP and nodal group demand estimates, had a detection

probability close to SPC-Hotelling T^2 in identifying large bursts, but the former proved to be better when it came to small leaks and instant detection. Anomalous event detection produced by [37] is a four-step statistical methodology, which relies on the desired TPR and FPR values and the window size. This modified SPC has proven its flexibility and is used by many water utilities. Karray et al. [38] proposed a hybrid method to detect leaks in long pipes, and it showed promising results: fast detection speed, acceptable reliability, full-time availability, good sensitivity, and medium complexity.

Evolutionary polynomial regression has been proven to be an effective leak detection method with no prerequisites, such as replacing the metering devices in the DMA [39]. BNL is an ANN approach that overcame the problem of assuming mutual relationships between the factors that cause leaks and the problem of prediction and uncertainty. It used the algorithm ExSEM and achieved an accuracy of almost 85% when applied to a water system. However, the aforementioned technique cannot manage continuous data, and discretization must take place for it to operate, such limitations must be overcome to ensure a better performance for future uses [40].

BPNN, one of the most used NNs, contains at least three layers (input, output, and one hidden layer) and is integrated with five hoop strain variations that can recognize patterns and predict the leak point when choosing the right hidden layers. Its RMSE, which can be as small as 1%, proves the success of the method [41]. As opposed to the model-based genetic algorithm used by [42], the dynamic time wrapping and supervised learning algorithm proposed by [43] performs better at night with great accuracy and sensitivity and a low number of false alarms.

MLPNN was investigated by [44]. Another data-driven approach that employs two different classifiers individually, namely, LDA and neural network, finds leaks based on pressure data. This AI-based technique achieves 80% accuracy; the approach suggests that increasing the number of sensors [45], Rayaroth Sivaradje [46]) and considering the optimal positioning in the WDS improve the performance of the classifiers [45].

A hybrid method that incorporates KPCA and Cas-SVDD detected small leaks with an accuracy of 90% by using historical or real-time data depending on if it is offline modeling or online detection, then denoising and signal reconstruction based on LMD take place before announcing a verdict [47].

Research conducted by [48], which located leaks by using Kriging spatial interpolation, shows that many parameters factor into the outcome, which is the same as concluded by [45]. Bayesian reasoning is then employed to better the accuracy [48]. Like many techniques that rely on pressure data to manage leaks effectively, disturbance extraction and isolation forest techniques by [49] showed great promise for application to large WDSs by detecting leaks from noisy pressure measurements without the introduction of extra flow sensors. The performance evaluation measures are summarized in Table 1.

3.1 Clustering strategies

In 2016, [50] came up with a new strategy to manage uncertainty: the clustering algorithm. That algorithm eliminates the uncertainty that results from the prediction and classification stages.

There are two steps before reaching a verdict; the first is a clustering algorithm for outlier detection, and the second identifies the leak. The method is promising given the low FPR compared to other studies. Differentiating from the huge amount of historical data needed for classification and prediction methods, [51] created a method that needs only "1 day's worth of time-series data" but can consider the variations throughout the year. They employed the temporally fluctuating correlation among flow sensors to evaluate the discrepancies between flow data coming from cosine distances to lower the false-positive rate brought on by cyclical changes in water demand, weather, and season. However, this did not result in better TPR results when compared to the clustering algorithm.

Monitoring support by Geleen et al. [52] is a proactive leakage control method that employs feature-based clustering that looks for cluster patterns to identify a leak based on pressure anomalies. This application is expanded to include oil, electricity, or gas leaks.

Lu and Sela [53] proposed a method that offers a quick and effective solution to find hidden information in WDSs by analyzing high-frequency pressure signals from scattered sensors [53].

Quiñones-Grueiro et al. [54] provided a method to manage large WDSs. They used topological clustering for network partitioning to locate leaks. CtL-SSL is a machine learning-based framework for WDS leak detection and localization. This framework lessens the data requirement, shows very effective results, and is a tool for water management in any WDS [55].

3.2 Methods implementing deep learning

Deep learning offers a framework to solve detection problems and a wider range in overcoming/issues that basic traditional techniques find challenging and has the potential to separate the variabilities more effectively [56]. Quiñones-Grueiro et al. [57] used a model-based approach to link a deep neural network, a data-driven approach, and Gaussian process regression to detect and locate leaks, a flowchart of the proposed approach is shown in Fig. 5.

Table 1 Data-driven approaches

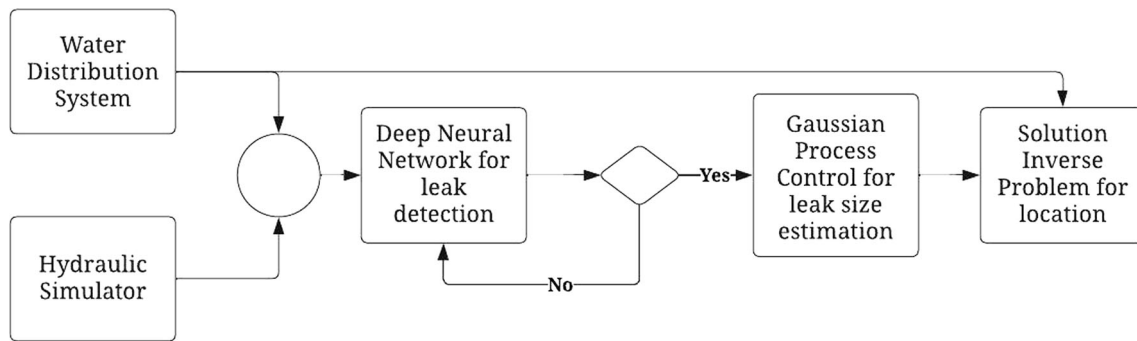| Technique	Category	Data requested	Efficiency						Reference
			TPR	FPR	ACC	Remarks			
Time delay NNs	Classification	Flow	75%	0	–	Error or DP		DT	[23]
ANNs and FLS	Prediction-classification	Flow	–	–	–	10% error in identifying flushing events		Several hours	[26]
SVR	Prediction-classification	Flow/pressure	–	22%	–	–		–	[27]
Fourier series and CUSUM	Prediction-classification	Flow	94.50%	0	–	–		9.8 days	[32]
Weighted least squares and EM algorithm	Prediction-classification	Flow	–	15%	–	–		–	[33]
Deviation analysis	Prediction-classification	Demand/pressure	90%	2.10%	–	–		20 min	[34]
Adaptive forecasting model	Prediction-classification	Demand/pressure	–	3.30%	–	Detection probability of 88.3%		18.7 min	[92]
NKF	Prediction-classification	Nodal group demand	–	0	–	Detection probability of 87%		–	[36]
BLN	Prediction-classification	Pressure	–	–	84.60%	–		–	[40]
Anomalous event detection	Statistical method	Flow	80%	10%	–	–		> = 1/2 h	[37]
Clustering algorithm	Clustering	Flow	71.43%	0.78%	–	–		–	[50]
Dynamic time wrapping and supervised learning algorithm	Prediction-classification	Flow/demand	98.30%	6.70%	–	–		–	[43]
Back-propagation NN	Prediction-classification	Pressure	–	–	–	1.01% RMSE		–	[41]
Clustering and cosine distances	Clustering	Flow	71.43%	0.40%	–	–		–	[51]
Feature-based clustering	Clustering	Pressure	–	–	–	Average accuracy F1-score of 93.5%		–	[52]
LDA, NN	Classification	Pressure	–	–	80%	–		–	[45]
KPCA, Cas-SVDD	Classification	Pressure	–	–	90% (small leaks)	The accuracy gets higher as the leaks get larger		–	[47]
Kriging spatial interpolation, Bayesian reasoning	Classification	Pressure	–	–	–	–		–	[48]
Disturbance extraction and isolation forest	Classification	Pressure	–	As low as 3% in some cases	–	–		DP as high as 90%	[49]
CtL-SSL	Clustering	Pressure	–	–	95%	–		83% leakage localization accuracy	[55]

Fig. 5 Flowchart of the methodology combining a data-driven approach with a model-based method

Table 2 Data-driven approaches that implement DL

Technique	Data requested	Efficiency			Reference
		TPR	FPR	ACC	
CNN and SVM	Normal and abnormal signal pairs	99.20%	–	99.30%	[59]
BLIFF	Pressure	–		98.38%	[47, 60, 62]
CNN and ISLMD	Pressure	–	79.15–92.5%		[62]
LSTM-RNN	Flow	94.82%	0.21%	88.80%	[58]
CNN and variational autoencoder	Acoustic signals	–	–	97.20%	[63]
LSTM-RNN	Flow	–	–	90%	[5]
DNN and Gaussian process regression	Flow/pressure	–	–	87%	[57]

Table 3 Model-based methods

Category	Methods	Reference
Sensitivity analysis	Flow sensitivity matrices, pressure sensitivity matrices	[67, 68, 70–73]
Optimization-calibration	Genetic algorithm, multi-objective ACO, LS	[42, 74, 76]
Error-domain model falsification	Leak detection based on leak-scenario falsification, time-series leakage detection, and Bayesian model updating approach	[77–82]

Wang et al. [58] succeeded in employing an LSTM neural network, which is an advanced version of the RNN structure, in leak detection because the method showed high accuracy and sensitivity.

Kang et al. [59] combined CNNs with support vector machines. That architecture uses a one-dimensional CNN for local feature detection and the latter classifies leaks, they proposed an algorithm to approximately localize the leaks. A novel framework developed by [60] locates the leak by narrowing down the locations to specific pipes and has proven applicability in practice. BLIFF uses pressure data in a mixed data-driven and model-based approach, a method that has not received much attention in other studies, to locate leaks. The framework proved useful because the locations were not as generalized and narrowed down to pipes. To escape the disorder of pressure signals, linear connections were used instead of the usual convolutional layers in DenseNet [61], which also eased calculations. With all the advantages that the method presents, there is a constraint, which requires pressure meters instead of flow meters.

A study by [62] employed ISLMD and CNN to detect and locate leaks. ISLMD eliminates noise from nonlinear and nonstationary signals. Their sensitivity results varied

depending on whether the AlexNet or LSSVM model was used. AlexNet detected 79.15% of the leaks, while LSSVM found that 92.5% CNN was integrated with a variational autoencoder in a paper by [63], and the combination achieved great accuracy and could detect leaks as small as 0.25 L/s, which proves its success.

Liao et al. [64] used DL to identify leaks in the water pipeline system. The accuracy of their model relies on calibrated parameters (e.g., pipe roughness), and it could reach 99.7%. However, their method does not cover the grounds where several leaks occur simultaneously. Lee et al. [5] introduced an LSTM-RNN, which accounts for numeral conditions in the system and provides a summary of the network's performance. Table 2 summarizes the performance of some of the methods mentioned in the literature. It shows the techniques used, the type of data needed for the models, and some metrics found.

4 Model-based methods and techniques used for burst and leak detection

Many researchers have preferred model-based methods over the years before they fully realized their full potential or limitations. These approaches are represented by mathematical formulas. Table 3 classifies the methods under three main categories.

Pudar and Liggett et al. [65] introduced a sensitivity matrix to guide measurement locations, but they suggested the method could not replace traditional leak surveys. That was the case until research conducted by (Perez et al. [66]) took place and adopted a mathematical model. Consequently, from 2011 to 2019, much research was invested into introducing leak detection and localization methods based on pressure sensitivity analysis [67–71].

Perez et al. [67] focused on determining a reasonable number of sensors to obtain more accurate locations based on an optimal sensor placement methodology that gave successful results under ideal conditions, but its accuracy was adversely affected by noise. Casillas et al. [70] located leaks using pressure measurements. Their LSS method assigns a key signature to every leak and then applies a linear model to approximate the location. In a case study that employed 15 pressure sensors and five flowmeters, the success rate reached 87%, better than the other six cases studied [71]. The change that resulted from replacing pressure data with flow measurements for leak detection in pipelines was covered by [72, 73].

A series of model-based methods such as genetic algorithms, least squares, and multi-objective ant colony optimization fall under the optimization-calibration category [42, 74–76]. Wu et al. [42] found leak hotspots by optimizing their model before calibration. They used genetic
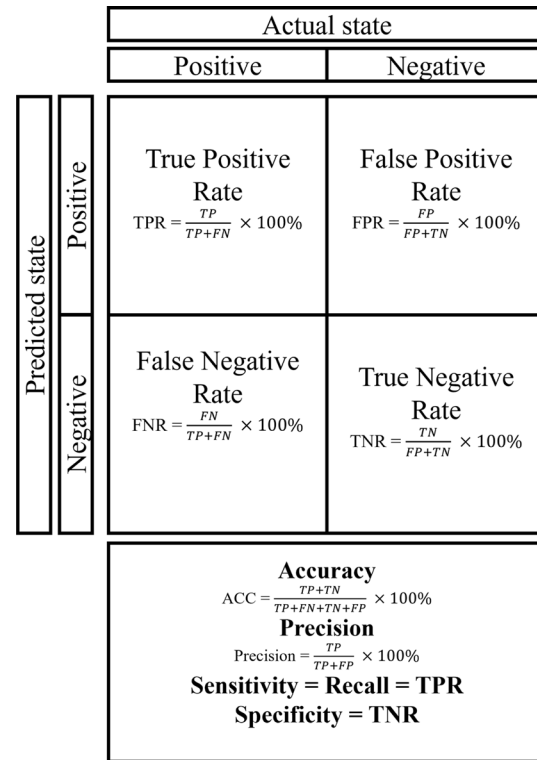

Fig. 6 Confusion matrix

algorithms to detect hard-to-find leaks. However, Hu et al. [20] revealed Wu's method congregates to the local optima due to the huge solution space.

Sanz et al. [76] established a methodology that requires calibration to solve the long-term issue of the leak location. Multi-objective ACO succeeded in detecting leaks by using two stages: dividing the WDS into pressure zones and then calibrating the nodal demands to quantify and locate the leaks [74]. Many methodologies build on error-domain model falsification where a model is falsified based on the thresholds, and the remaining models describe the leak region. The falsification method proposed by [77] paved the way for work by [78–82].

5 Performance evaluation

Most authors assessed their methodologies using different measures. A confusion matrix is a machine learning concept [83], which represents visually the performance of an algorithm. A confusion matrix shows important measures: TPR, TNR, FPR, and FNR. Such measures are mentioned in various papers [23, 51, 59, 84, 85]. A true positive refers to the occurrence of an actual leak in the WDS, a false positive is an alarm for a leak when there is none. Other common measures for evaluation are accuracy, precision,

and *F*-score. Figure 6 shows a confusion matrix with measures and their respective mathematical equations.

6 Results and Discussion

Data-driven approaches can be exploited to detect and locate leaks without deep knowledge of every component or relationship in the WDS. However, the approaches come with the price of obtaining high FPR. Wu et al. [8] summarized the most frequent limitations of classification, prediction-classification, and statistical methods.

The requirement of data and the absence of hydraulic data labels are two limitations in classification methods. Mounce et al. [23] used the same simulated events in training and testing the method. MLPNN by [44] is immune to noise, shows a parallel structure to the real-life WDS, and has speedy processing and sorting abilities. However, the performance evaluation rates were not explicitly mentioned as was the case with many other papers [30, 35, 48, 84, 85].

Prediction-classification methods could provide misleading results at times because of the uncertainty that results from measurement inaccuracies, absent data, or variations over the year. Many methods may not be able to identify small leaks [33, 36–38, 50, 51, 84–89] or be based on impractical assumptions that may not reflect reality [36, 90, 91].

Background noises often cause an error that leads to inaccuracy in conclusions [32, 88].

Many methodologies require the optimal placement of sensors or the investment in more sensors [45, 46, 84–88]. The presence of large leaks overshadows smaller leaks in the EPR modeling paradigm [39].

Model-based methods are widely applied and studied for leak detection but similar to data-driven approaches, there are some limitations and weak spots that must be addressed. Generally, their effectiveness is questioned because of uncertainties generated by full-scale water distribution systems. Many papers address this in the literature. Hu et al. [20] summarized the weaknesses, the hydraulic models must be well-calibrated and reality simulating for model-based approaches to be applicable, which requires a huge amount of data and investment. Well-calibrated hydraulic models are not common in water utilities.

Perez et al. [68] stated that the pressure sensitivity matrix is very applicable and provides adequate results. However, that method delivers satisfactory outcomes when the sensors are optimally placed. This methodology by (Perez et al. [68]) assumes that leaks occur at nodes and neglects the leaks that happen at the pipeline level. Consequently, the model locates leaks at the nearest node.

Casillas et al. [70] continued and said that pressure sensitivity matrices' accuracy is affected by noise inference in the system. The genetic algorithm used by [42] operates well when working on large WDSs in the mornings, but it finds the local optima. Multi-objective ACO can be used with either pressure or demand, but the performance relies on the number of pressure measurements [74]. Error-domain falsification models show potential and have been tested on a real full-scale WDS [82], but the method of [77] cannot detect small leaks.

7 Conclusions and future research

Water loss has received worldwide attention over the years. Loss from leaks could be controllable and should be minimized. In this review, many methodologies are discussed under two distinct categories: data-driven and model-based approaches. Classification, prediction-classification, statistical methods, and clustering algorithms fall under the former while sensitivity matrices, optimization-calibration, and error-domain falsification-based methods fall under the latter. There is an overlap between the two categories with methods that mix the two [16, 20, 46, 89–91]. The performance of the proposed methodologies is summarized as well as their weaknesses. Many studies provided effective or promising detection techniques, and many papers provided approaches to locate leaks. Future work should focus on minimizing false-positive rates and maximizing true-positive rates in methodologies and techniques that water utilities can easily access. Some techniques can be used in other fields such as oil or electricity. Also, some techniques with high detection rates should be tested in real-life WDSs and compared with the theoretical results.

Data availability Data sharing does not apply to this article as this is a literature review.

Declarations

Conflict of interest There is no conflict of interest associated with this paper.

References

1. Liemberger R, Marin P (2006) The challenge of reducing non-revenue water in developing countries—how the private sector can help: a look at performance-based service contracting
2. Liemberger R, Wyatt A (2019) Quantifying the global non-revenue water problem. Water Supply 19(3):831–837

3. Xu J, Chai KTC, Wu G, Han B, Wai ELC, Li W, Gu Y (2018) Low-cost, tiny-sized MEMS hydrophone sensor for water pipeline leak detection. IEEE Trans Ind Electron 66(8):6374–6382

4. Britton TC, Stewart RA, O'Halloran KR (2013) Smart metering: enabler for rapid and effective post meter leakage identification and water loss management. J Clean Prod 54:166–176

5. Lee CW, Yoo DG (2021) Development of leakage detection model and its application for water distribution networks using RNN-LSTM. Sustainability 13(16):9262

6. Karim MR, Abbaszadegan M, LeChevallier M (2003) Potential for pathogen intrusion during pressure transients. J Am Water Works Ass 95(5):134–146

7. Fox S, Shepherd W, Collins R, Boxall J (2016) Experimental quantification of contaminant ingress into a buried leaking pipe during transient events. J Hydraul Eng 142(1):04015036

8. 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

9. Wang XJ, Simpson AR, Lambert MF, Vítkovský JP (2001) Leak detection in pipeline systems using hydraulic methods: a review. In: Conference on hydraulics in civil engineering, the institution of engineers, Australia, Hobart (pp. 23–30)

10. Lambert A (1994) Accounting for losses: the bursts and background concept. Water Environ J 8(2):205–214

11. Farah E, Shahrour I (2017) Leakage detection using smart water system: combination of water balance and automated minimum night flow. Water Resour Manage 31(15):4821–4833

12. Mounce SR, Mounce RB, Jackson T, Austin J, Boxall JB (2014) Pattern matching and associative artificial neural networks for water distribution system time series data analysis. J Hydroinf 16(3):617–632

13. Loveday M, Dixon J (2005) DMA sustainability in developing countries. In: Proceedings. IWA Specialized Conference: Leakage

14. Mutikanga HE, Sharma SK, Vairavamoorthy K (2013) Methods and tools for managing losses in water distribution systems. J Water Resour Plan Manag 139(2):166–174

15. Zaman D, Tiwari MK, Gupta AK, Sen D (2020) A review of leakage detection strategies for pressurized pipeline in steady-state. Eng Fail Anal 109:104264

16. Xie J, Xu X, Dubljevic S (2019) Long range pipeline leak detection and localization using discrete observer and support vector machine. AIChE J 65(7):e16532

17. Xue Z, Tao L, Fuchun J, Riehle E, Xiang H, Bowen N, Singh RP (2020) Application of acoustic intelligent leak detection in an urban water supply pipe network. J Water Supply Res Technol AQUA 69(5):512–520

18. Li R, Huang H, Xin K, Tao T (2015) A review of methods for burst/leakage detection and location in water distribution systems. Water Sci Technol Water Supply 15(3):429–441

19. Colombo AF, Lee P, Karney BW (2009) A selective literature review of transient-based leak detection methods. J Hydro Environ Res 2(4):212–227

20. Hu Z, Chen B, Chen W, Tan D, Shen D (2021) Review of model-based and data-driven approaches for leak detection and location in water distribution systems. Water Supply 21(7):3282–3306

21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group* (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151(4):264-269

22. Mounce SR, Day AJ, Wood AS, Khan A, Widdop PD, Machell J (2002) A neural network approach to burst detection. Water Sci Technol 45(4–5):237–246

23. Mounce SR, Machell J (2006) Burst detection using hydraulic data from water distribution systems with artificial neural networks. Urban Water Journal 3(1):21–31

24. Mounce SR, Boxall JB, Machell J (2007) An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation. Water management challenges in global change, pp 313–320

25. Aksela K, Aksela M, Vahala R (2009) Leakage detection in a real distribution network using a SOM. Urban Water J 6(4):279–289

26. 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

27. Mounce SR, Mounce RB, Boxall JB (2011) Novelty detection for time series data analysis in water distribution systems using support vector machines. J Hydroinf 13(4):672–686

28. Nasir MT, Mysorewala M, Cheded L, Siddiqui B, Sabih M (2014) Measurement error sensitivity analysis for detecting and locating leak in pipeline using ANN and SVM. In: 2014 IEEE 11th international multi-conference on systems, signals and devices (SSD14), IEEE, (pp. 1–4)

29. Romano M, Kapelan Z, Savić DA (2013) Geostatistical techniques for approximate location of pipe burst events in water distribution systems. J Hydroinf 15(3):634–651

30. 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

31. Palau Estevan CV, Arregui de la Cruz F, Carlos Alberola MDM (2012) Burst detection in water networks using principal component analysis. J Water Resour Plan Manag 138(1):47–54

32. Eliades DG, Polycarpou MM (2012) Leakage fault detection in district metered areas of water distribution systems. J Hydroinf 14(4):992–1005

33. Ye G, Fenner RA (2014) Weighted least squares with expectation-maximization algorithm for burst detection in UK water distribution systems. J Water Resour Plan Manag 140(4):417–424

34. Bakker M, Vreeburg JHG, Van De Roer M, Rietveld LC (2014) Heuristic burst detection method using flow and pressure measurements. J Hydroinf 16(5):1194–1209

35. Hutton C, Kapelan Z (2015) Real-time burst detection in water distribution systems using a Bayesian demand forecasting methodology. Procedia Eng 119:13–18

36. Jung D, Lansey K (2015) Water distribution system burst detection using a nonlinear Kalman filter. J Water Resour Plan Manag 141(5):04014070

37. 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

38. 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

39. Laucelli D, Romano M, Savić D, Giustolisi O (2016) Detecting anomalies in water distribution networks using EPR modelling paradigm. J Hydroinf 18(3):409–427

40. Leu SS, Bui QN (2016) Leak prediction model for water distribution networks created using a Bayesian network learning approach. Water Resour Manage 30(8):2719–2733

41. Jia Z, Ren L, Li H, Sun W (2018) Pipeline leak localization based on FBG hoop strain sensors combined with BP neural network. Appl Sci 8(2):146

42. Wu ZY, Sage P, Turtle D (2010) Pressure-dependent leak detection model and its application to a district water system. J Water Resour Plan Manag 136(1):116–128

43. Huang P, Zhu N, Hou D, Chen J, Xiao Y, Yu J, 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

44. Gómez-Camperos JA, Espinel-Blanco EE, Regino-Ubarnes FJ (2019) Diagnosis of horizontal pipe leaks using neural networks. In: Journal of physics: conference series (Vol. 1388, No. 1, p. 012032). IOP Publishing

45. 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

46. Rayaroth R, S G (2019) Random bagging classifier and shuffled frog leaping based optimal sensor placement for leakage detection in WDS. Water Resour Manage 33(9):3111–3125

47. Zhou M, Zhang Q, Liu Y, Sun X, Cai Y, Pan H (2019) 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

48. 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

49. Weirong X, Zhou X, Xin K, Boxall J, Yan H, Tao T (2020) Disturbance extraction for burst detection in water distribution networks using pressure measurements. Water Resour Res. https://doi.org/10.1029/2019WR025526

50. 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

51. Wu Y, Liu S, Smith K, Wang X (2018) Using correlation between data from multiple monitoring sensors to detect bursts in water distribution systems. J Water Resour Plan Manag 144(2):04017084

52. Geelen CV, Yntema DR, Molenaar J, Keesman KJ (2019) Monitoring support for water distribution systems based on pressure sensor data. Water Resour Manage 33(10):3339–3353

53. Xing L, Sela L (2019) Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems. Water Res 158:291–300

54. Quiñones-Grueiro M, Verde C, Llanes-Santiago O (2019) Novel leak location approach in water distribution networks with zone clustering and classification. In: Carrasco-Ochoa JA, Martínez-Trinidad JF, Olvera-López JA, Salas J (eds) Pattern recognition: 11th Mexican conference, MCPR 2019, Querétaro, Mexico, June 26–29, 2019, proceedings. Springer International Publishing, Cham, pp 37–46. https://doi.org/10.1007/978-3-030-21077-9_4

55. Fan X, Yu X (2021) An innovative machine learning based framework for water distribution network leakage detection and localization. Struct Health Monit 21(4):1626–1644

56. Yuan Q, Shen H, Li T, Li Z, Li S, Jiang Y, Zhang L (2020) Deep learning in environmental remote sensing: achievements and challenges. Remote Sens Environ 241:111716

57. Quiñones-Grueiro M, Milián MA, Rivero MS, Neto AJS, Llanes-Santiago O (2021) Robust leak localization in water distribution networks using computational intelligence. Neurocomputing 438:195–208

58. 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

59. 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

60. Zhou X, Tang Z, Xu W, Meng F, Chu X, Xin K, Fu G (2019) Deep learning identifies accurate burst locations in water distribution networks. Water Res 166:115058

61. Zhang J, Lu C, Li X, Kim HJ, Wang J (2019) A full convolutional network based on DenseNet for remote sensing scene classification. Math Biosci Eng 16(5):3345–3367

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

63. Cody RA, Tolson BA, Orchard J (2020) Detecting leaks in water distribution pipes using a deep autoencoder and hydroacoustic spectrograms. J Comput Civ Eng 34(2):04020001

64. Liao Z, Yan H, Tang Z, Chu X, Tao T (2021) Deep learning identifies leak in water pipeline system using transient frequency response. Process Saf Environ Prot 155:355–365

65. Pudar RS, Liggett JA (1992) Leaks in pipe networks. J Hydraul Eng 118(7):1031–1046

66. Pérez R, Puig V, Pascual J, Peralta A, Landeros E, Jordanas L (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Sci Technol Water Supply 9(6):715–721

67. Pérez R, Puig V, Pascual J, Quevedo J, Landeros E, Peralta A (2011) Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Eng Pract 19(10):1157–1167

68. Perez R, Sanz G, Puig V, Quevedo J, Escofet MAC, Nejjari F, Sarrate R (2014a) Leak localization in water networks: a model-based methodology using pressure sensors applied to a real network in Barcelona [applications of control]. IEEE Control Syst Mag 34(4):24–36

69. Pérez R, Cugueró MA, Cugueró J, Sanz G (2014b) Accuracy assessment of leak localisation method depending on available measurements. Procedia Eng 70:1304–1313

70. Casillas MV, Garza-Castañón LE, Puig V, Vargas-Martinez A (2015) Leak signature space: an original representation for robust leak location in water distribution networks. Water 7(3):1129–1148

71. Salguero FJ, Cobacho R, Pardo MA (2019) Unreported leaks location using pressure and flow sensitivity in water distribution networks. Water Supply 19(1):11–18

72. 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

73. Jiménez-Cabas J, Romero-Fandiño E, Torres L, Sanjuan M, López-Estrada FR (2018) Localization of leaks in water distribution networks using flow readings. IFAC-PapersOnLine 51(24):922–928

74. Hajibandeh E, Nazif S (2018) Pressure zoning approach for leak detection in water distribution systems based on a multi objective ant colony optimization. Water Resour Manage 32(7):2287–2300

75. Nasirian A, Maghrebi MF, Yazdani S (2013) Leakage detection in water distribution network based on a new heuristic genetic algorithm model. J Water Resour Prot 05(03):294–303

76. Sanz G, Pérez R, Kapelan Z, Savic D (2016) Leak detection and localization through demand components calibration. J Water Resour Plan Manag 142(2):04015057

77. Goulet JA, Coutu S, Smith IF (2013) Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Adv Eng Inform 27(2):261–269

78. Moser G, Paal SG, Smith IF (2015) Performance comparison of reduced models for leak detection in water distribution networks. Adv Eng Inform 29(3):714–726

79. Moser G, Paal SG, Jlelaty D, Smith IF (2016) An electrical network for evaluating monitoring strategies intended for hydraulic pressurized networks. Adv Eng Inform 30(4):672–686

80. Jensen HA, Jerez DJ (2019) A Bayesian model updating approach for detection-related problems in water distribution networks. Reliab Eng Syst Saf 185:100–112

81. Moser G, Paal SG, Smith IF (2018) Leak detection of water supply networks using error-domain model falsification. J Comput Civ Eng 32(2):04017077

82. Shao Y, Li X, Zhang T, Chu S, Liu X (2019) Time-series-based leakage detection using multiple pressure sensors in water distribution systems. Sensors 19(14):3070

83. Deng X, Liu Q, Deng Y, Mahadevan S (2016) An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inf Sci 340:250–261

84. 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

85. Romano M, Kapelan Z, Savić DA (2014) Evolutionary algorithm and expectation maximization strategies for improved detection of pipe bursts and other events in water distribution systems. J Water Resour Plan Manag 140(5):572–584

86. Srirangarajan S, Allen M, Preis A, Iqbal M, Lim HB, Whittle AJ (2013) Wavelet-based burst event detection and localization in water distribution systems. J Sign Process Syst 72(1):1–16

87. 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

88. Zan TTT, Lim HB, Wong KJ, Whittle AJ, Lee BS (2014) Event detection and localization in urban water distribution network. IEEE Sens J 14(12):4134–4142

89. 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

90. 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

91. Soldevila A, Fernandez-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

92. Bakker M, Trietsch EA, Vreeburg JHG, Rietveld LC (2014b) Analysis of historic bursts and burst detection in water supply areas of different size. Water Sci Technol Water Supply 14(6):1035–1044