IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 20, 2018, accepted December 2, 2018, date of publication December 6, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2885444

Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection

T. K. CHAN[1,2], CHENG SIONG CHIN[1], (Senior Member, IEEE), AND XIONGHU ZHONG[2], (Member, IEEE)
[1]Faculty of Science, Agriculture, and Engineering, Newcastle University, Singapore 599493
[2]Visenti Pte. Ltd., Singapore 609935

Corresponding author: Cheng Siong Chin (cheng.chin@ncl.ac.uk)

This work was supported by the Economic Development Board-Industrial Postgraduate Programme (EDB-IPP) of Singapore under Grant BH180750 with Visenti Pte. Ltd.

ABSTRACT Water is a precious resource that should be managed carefully. However, due to leakages in water distributed networks (WDNs), a large amount of water is lost each year that suggests the need for reliable and robust leak detection and localization system. This paper attempts to review the current technologies for leakage detection in WDN as well as several proposed intelligent methodologies (such as support vector machine, neural network, and convolution neural network) over the past few years. The current methodologies and their limitations are discussed. Uncertainties involved in the implementation of WDN leakage detection are also discussed, and several suggestions to overcome such uncertainties are provided for future implementations.

INDEX TERMS Water distribution networks, leakage, localization, review.

I. INTRODUCTION

Water is a critical and essential resource supporting our daily activities for sustaining our ecosystem. In any city, water is distributed through a highly sophisticated network commonly known as the Water Distribution Networks (WDNs). These networks may experience deterioration that causes leakage due to several factors. They can be categorized into internal or external factors. Internal factors are damages on the pipe itself such as pipe corrosion, pipe age, pipe defects, and poor workmanship. The external factors are damaged due to the external or third party such as mechanical damage caused by excessive pipe load (traffic above the road), excavation, ground movement and climate conditions [1]–[4].

Due to such damages, water loss through leakage had been reported to be approximately 20% to 30% of the total water supplied in different countries [5]–[11]. This is regarded as a costly problem due to the wastage of natural resources [12]. In some cases, total damage cost had been reported to reach several million US dollars [13] as water lost through the pipe can damage the environment including nearby infrastructure, causing service disruption and increase unnecessary energy cost and carbon footprint [14]. Therefore, effective leak detection and localization system have the potential to save a large quantity of water as well as money.

Leakages can be classified into reported, unreported leakage or background leakage [12], [15]. Reported burst event is usually visible on the ground as they can be easily detected by maintenance personnel or the public. Unreported burst event exhibits the same type of leakage as the reported burst event without surfacing to the ground. On the other hand, the background type leakages are small leakages that are difficult or cannot be detected through normal methods such as leakage through creeping joints [16]. Some literature may refer to burst leakage as burst event and background leakage as leakage [9], [17], [18]. This type of small leakage may often go unnoticed resulting in significant losses [18], [19]. Among the leak detection system, it can be classified into passive and active systems. The former requires direct visual inspection or monitoring of sites. The latter comprises of an analysis of signals such as acoustic signals, vibration, flow, and pressure measurement [4], [20]. An active system can be further classified into mainly transient-based approaches, model-based approaches, and data-driven approaches [17].

In regards to the WDN leakage detection and localization technologies, several authors had published reviews and their insights. Adedeji *et al.* [12] published their review of leakage detection methods classified into external (acoustic emission, gas injection, fiber optics sensing, magnetic induction and

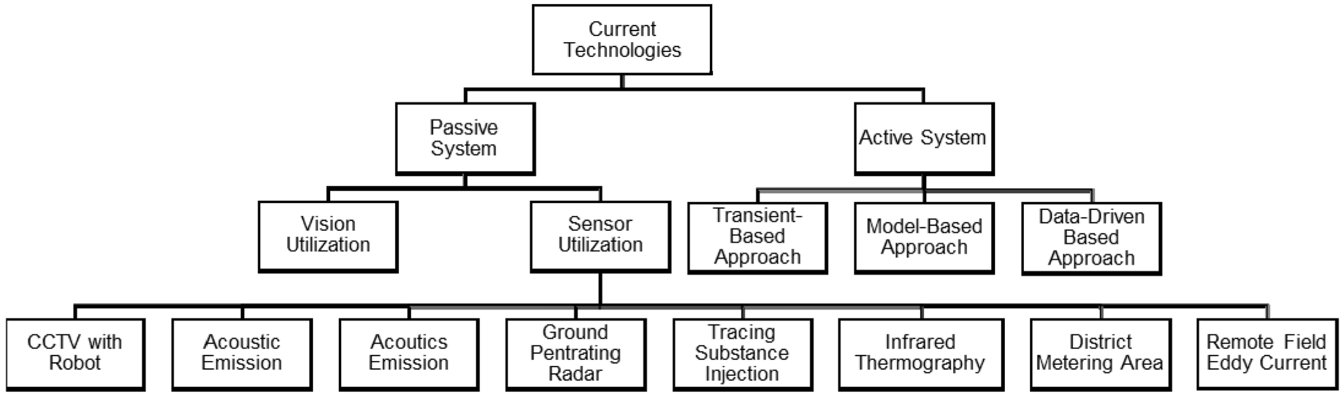

FIGURE 1. Overview of Current Leak Detection Technologie.

ground-penetrating radar methods) or internal (balancing, model-based, pressure or flow monitoring, signal processing approach and statistical analysis) based methods targeted towards pressurized piping system. Li *et al.* [21] also did a rather similar review as Adedeji *et al.* [12] but categorized into non-numerical, numerical and time domain, frequency domain analysis.

Wu and Liu [22] presented their review on data-driven approaches for burst detection in the water distribution system that include the performance capabilities and limitations. Datta and Sarkar [23] presented their review which included methodology on blockage and leakage detection. Their review focuses on the process and measurement principles that are used for the pipeline faults detection. Colombo *et al.* [24] presented their findings on transient based methods that presented a summary of current and past work. They also described state of the art in the area. Liu and Kleiner [25] did a review on the inspection techniques and technologies towards condition assessment of water distribution and transmission mains. Finally, Obeid *et al.* [26] published their comprehensive survey on software and hardware solutions proposed for water pipeline infrastructure monitoring.

There is a lot of reviews written in the area but remains incomplete.

The researchers [12] focused on the summary of different categories of leak detection which can be said for Datta and Sarkar [23]. Some methodologies reviewed by Li *et al.* [21] was more than a decade old with some new methods during the time of publication. Similarly, methodologies reviewed by Colombo *et al.* [24] was around a decade ago even though it was comprehensive. Since Wu and Liu [22] focused their review on data-driven approaches, the paper did not include recent progression and methodology in other categories. Liu and Kleiner [25] focused on the current technologies for detection and did not include any review or survey for proposed methodology by other authors. Similarly, Obeid *et al.* [26] provided a summary of currently available technologies and focused on the development of a wireless sensor node platform for detection.

Therefore, the focus of this paper is to provide the latest review on published methodologies. This paper aims to provide a noteworthy contribution to the literature by

1) providing an extensive survey on the newer published methodologies targeted towards WDNs (up to last five years) which include where, how, the detection rate of methodologies proposed;
2) identifying and discussing the limitations of proposed methodologies;
3) proposing solutions or recommendations for the reader to avoid/overcome the limitations.

The paper starts with the introduction of the current leak detection technologies. Followed by the survey on different published methodologies and their detection rate. After that, the paper presents a discussion and suggestions for common limitations present in literature. Finally, the paper ends with a conclusion.

II. CURRENT LEAK DETECTION TECHNOLOGIES

The review of current leak detection technologies will begin with the passive system followed by the active system. An overview of the different types of technologies available is given in Figure. 1.

A. VISION UTILIZATION

The oldest and perhaps unsystematic passive leak detection system is to observe any indication of ponding at ground surface or anomalous vegetation growth that suggest a possible leaking pipe [24].

B. SENSOR UTILIZATION

As technological advances, more accurate leak detection is carried by using manual sticks or portable measurement devices. These devices can detect sound or vibration produced by water leaking from pressurized pipes [27]. Subsequently, with the availability of remote-controlled robots, detection of pipe leakage can be carried out using a Closed Circuit Television (CCTV). The CCTV system comprised of a remote-controlled pan, and a camera is mounted on a robot traveling between two manholes inside the pipeline

FIGURE 2. CCTV Inspection (Image from M. TUCKER & SONS).

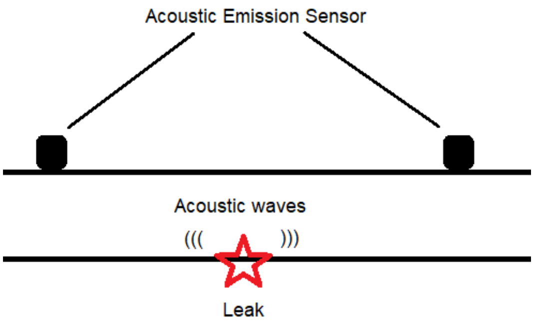

FIGURE 3. Acoustic Emission Detection Method for Pipe Leakage.

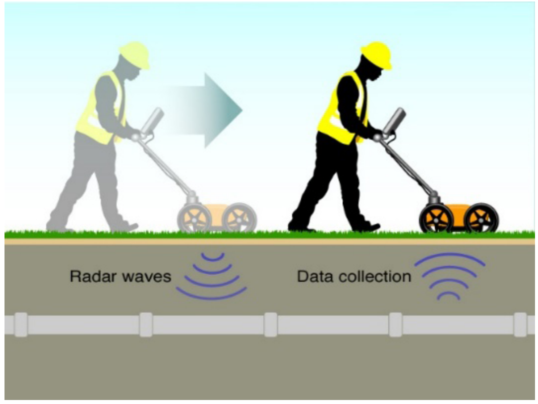

FIGURE 4. Ground Penetrating Radar Operating Principle (Image from SiteScan).

controlled by certified operators [25], [28]. However, such passive systems have several drawbacks as such practices are time-consuming, labor intensive and have low reliability in detecting leaks as accuracy is dependent on the user's experience [12], [27].

Other methods include acoustic emission technique, ground penetrating radar, tracing substance injection, the use of IR thermography, monitoring of a District Metering Area (DMA) through flow and pressure sensors, fiber optics sensors and Remote Field Eddy Current (RFEC) technique.

Acoustics emission technique relies on propagating elastic waves released from an active source [29]. Since escaping liquid creates an acoustic signal as it passes through a defect, this phenomenon makes the acoustics emission technique a suitable candidate. When a leak occurs, the acoustic signal propagates along the pipeline which is picked up by the acoustic sensors installed along the pipeline, if the received signal by one sensor is stronger and has a higher magnitude then the location of the leak can easily be identified [12].

This technique has the advantage of detecting a leak in real time [29], high detection and localization accuracy [12] and is generally faster than other methods such as tracer gas, infrared (IR) thermography, ultrasonic and electromagnetic scanning [30]. However, signals are influenced by the type of pipe materials, changing sound propagation conditions from one pipeline section to another [8]. External sources such as background noise also affect the signal [4]. It is also not applicable for long pipelines as it requires a great number of sensors which increase the cost significantly [12]. Furthermore, many acoustics techniques are insensitive to large leaks as they do not generate enough vibrations in the high frequencies [24].

The operating principle of a ground penetrating radar is to capture an image of the pipe underground where the electromagnetic signatures of leak regions would manifest themselves in the captured images. The source of a leak can then be located through a direct interpretation of the images [31].

However, this detection method is influenced by the type of soil where the pipes are buried in. The results showed that reflections beneath the leak regions are comparatively weaker than surrounding soil medium for most homogeneous soil and may not be possible to observe void phenomenon for most inhomogeneous soil. Furthermore, the applicability of this method for deeply buried pipes is limited, as either the moisture of soil or the inhomogeneity within the soil may not provide sufficient signal power levels above the noise floor [31]. In addition, it is difficult to interpret the results [12].

For tracing substances injection, it is an effective and proven method and can detect even the smallest of leaks with low false alarm. Another pipeline with in-built sensors is usually inserted along with the pipeline that requires monitoring so when leakage occurs; the sensors can alert the engineers immediately [32].

However, there may be a need to filter or cleanse the water before use which complicates the pipeline distribution network. In addition, it may risk environmental contamination in the presence of leak [8]. Furthermore, for large low-pressure applications where tracer gas is used, the high volume of gas required makes this method impractical [12]. Finally, installation of in-built sensors together with existing underground distribution system requires much manual labor to excavate the ground that is quite impractical.

As for infrared thermography, the method relies on the measurement of IR radiation across spatial surfaces and energy transfer theory to identify thermal anomalies for

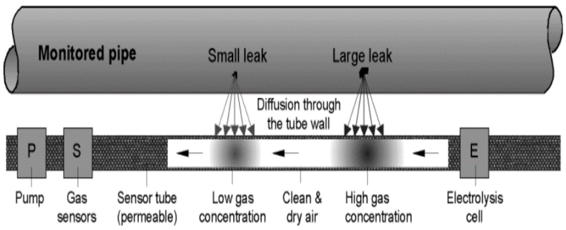

FIGURE 5. Tracing Substance Injection [32].

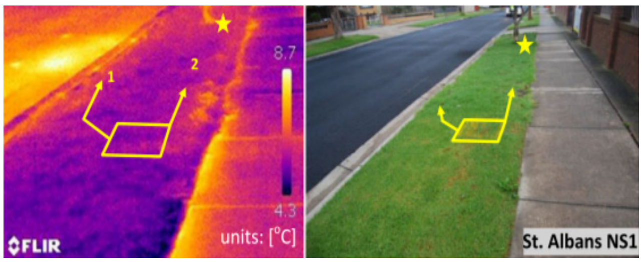

FIGURE 6. IR Thermography on Water Leakage (Box indicate leak location, arrow indicate leak flow, star represent the same landmark in different pictures [35].

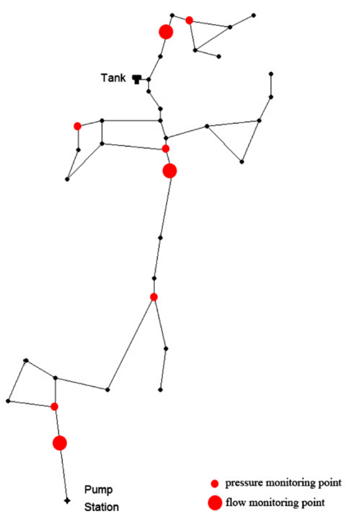

FIGURE 7. Example of a DMA [18].

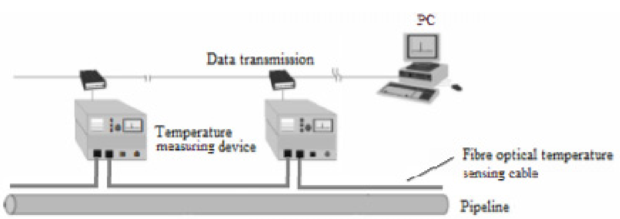

FIGURE 8. Fiber Optical Sensing Cable [48].

different application [33]. Due to the fact that water leakage causes temperature differences in the vicinity of the leak, IR camera can detect leaks by capturing the thermal profile of the surface above the pipeline [34]. Bach and Kodikara [35] provided an in-depth explanation of how water leakage can be detected via the temperature differences of the water leakage and the surrounding soil.

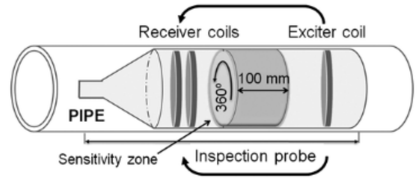

FIGURE 9. Example of an RFEC Probe In a Pipe [46].

Although detection of leaks using IR thermography has a low success rate, the results presented by Bach and Kodikara [35] on leakage detection was not impressive. The leak detection rate was only 59% (16 out of 27). As explained [21], IR thermography is affected by many factors such as weather conditions, soil and pavement surface conditions. To increase the likelihood of detection, thermal contrast should be at the maximum between leakage and the surrounding environment [35]. Thus, detection can only be carried out during a pre-sunrise condition when the influence of solar radiation is negligible. Moreover, the reliability of detection depends on the thermal sensitivity of the device [35]. The user's experience also plays an essential role as an IR camera is susceptible to noise such as the reflection of light and the angle of capturing the image. A certain level of expertise is also required to analyze the image.

A DMA is a subsection of a large complex distribution network. This sectorization is done through the closure of valves or disconnection of network pipes with the inlet and outlet flow metered [36]–[39]. It allows continuous monitoring of inlet and outlet flow measurements. The subsequent analysis then calculates the level of leakage within a DMA [39]. If a leak occurs, flow rate increases while a transient pressure drop can be observed [40]. It allows real-time alert of any leakage present in DMA.

The optimal approach in the analysis of DMA flows is when the flow is at a minimum. This window occurs at night time when consumer demand is low. Hence, the leak magnitude over the total DMA flow is at the highest rate [41], [42]. However, the boundary and size of a DMA depend on the topographic conditions and the number of water users. It cannot be too large as it will be hard to locate the leakage [40]. Therefore, the partitioning of the WDN into DMAs is crucial for identifying the most vulnerable areas to schedule leak detection activities [36].

Distributed fiber optics sensors are often adopted for the detection of leakage. It has superior immunity to electrical noise, long-term measurement stability, corrosion resistant properties and provides long-distance sensing capabilities with many measurement points through one optical fiber line. These make it suitable for pipeline monitoring [44], [45]. As leakages from pipeline cause local temperature anomalies in the vicinity, fiber optics installed which take temperature measurements over the entire pipeline can detect this leakage in a short time [46], [47], [51].

Despite the success of fiber optics, this methodology can only be deployed for monitoring linear pipelines. Moreover, the cost of implementing such a system is high [12].

RFEC technique utilizes low-frequency alternating current and through the wall transmission to inspect pipes and is done by passing a probe through the pipe wall. This technique allows the detection of both external and internal defects not visible to the naked eye caused by corrosion [48], [49].

Although RFEC provides reliable information on corrosion defects, it cannot detect leakage or small corrosion pits that can result in a larger area of corrosion and leakage. Leaking through corrosion pits are better inspected by other leak detection methods [49]. In addition, this method is only applicable to conducting materials where the network is accessible to the RFEC probe.

C. TRANSIENT-BASED APPROACH

A leak is a hydraulic phenomenon. Thus, a transient pressure wave is an ideal parameter for such a task [24]. A transient pressure wave refers to a pressure wave that is short-lived [50]. In the presence of a leak opening; there is a drop in pressure in the surrounding medium [52] that produces such transient pressure wave phenomenon. The transient based method usually extract information about the presence of leak from measured or modeled transient pressure traces within a pipeline network [12].

As mentioned by researchers [8], these methods are not straightforward in their application beyond pipeline networks such as the inverse transient analysis [53] that analyze the pressure data collected during the occurrence of transitory events via minimization of the difference between the observed and the calculated parameters [54].

Detecting negative wave pressure relies on regular data collection and has high transmission cost. Such a method may be influenced by background noise or other events in a complex network [22] Negative pressure waves generated can be influenced by length, the diameter of pipelines and fluid properties. It has low detection accuracy in case of a microporous leak, or for liquid with high elastic coefficients, densities, or viscosities [56].

Few researchers [22] and [24] emphasized the fact that earlier literature on transient based approaches has not been validated in the real system. It was also reiterated by Adedeji *et al.* [12], that most of those transient methods could not be used for the real-time application. Moreover, such methods also rely heavily on complex simulation models or investment in a large number of sensors [57].

D. MODEL-BASED APPROACH

The model-based approach usually involves the use of mathematical functions or formulas to represent or replicate the operation of a pipe network. It can determine the approximate leakage location [17] by comparing pressure measurement with their estimation obtained using the hydraulic network model [40]. The condition is that the model should be a good representation of the network. Perez *et al.* [38], Adachi *et al.* [58], Meseguer *et al.* [59] based on some of the mathematical functions for their implementation of the model-based approach.

The law of conservation of mass states that under steady state condition without the presence of a leak, the inflow of water must be equal to the outflow of water. In pressurized water distributed pipelines, there is no storage of water [32], [38]. This condition can be written as

$$f(x) = m_i - m_o \tag{1}$$

where m_i is the mass of water flowing in and m_o is the mass of water flowing out. Theoretically, for $f(x)$ equals zero then the pipe has no leakage and vice versa.

In the conservation of energy, it states that the total energy of an isolated system remains constant since energy cannot be created and destroyed. Therefore, the energy difference between two points in a pipeline is the difference between the energy added to the flow and frictional and heat losses written as follows [38].

$$\Delta e = e_i - e_{fh} \tag{2}$$

where e_i is the energy added by the pump, e_{fh} is the energy loss through friction and heat and Δe is the difference in energy. The relationship between pipe flow and energy loss caused by friction can be represented by

$$e_{fh} = Kq^r \tag{3}$$

where K is the pipe coefficient, q is the pipe flow, and r is an exponent of value 2 [60].

The nodal demand at a junction is modeled as

$$d_n(t) = bd_n p_{a,n}(t)D(t) \tag{4}$$

where $d_i(t)$ is the demand at a node n at a time t, bd_n is the base demand of node n, $p_{a,n}(t)$ is the value of pattern a associated to the node n at a time t and $D(t)$ is the sum of supplied water to the system measured at the network inputs and storage units at the time t [38].

Leak size can then be modeled as

$$L = C_e p^\gamma \tag{5}$$

where L is the leak size, C_e is the emitter coefficient, p is the pressure at the node and γ is an exponent of value 0.5 for detectable leaks and burst on metallic pipes [61].

Although the conservation of mass theory is simple, cost-effective and very sensitive, the changes in the flow pressure can suffer from false detections [26]. It is unable to localize leaks [62] and needs more sophisticated equations. However, the drawback of using more sophisticated equations to reflect real-time condition is the need for humongous data for calibration and is rather computationally expensive [38]. Furthermore, the network topology changes with the addition or elimination of any element (pipes, nodes or tanks). Consumers' demand is also hard to determine and be considered [63], [64]. As a result, constructing and maintaining well calibrated hydraulic models are quite challenging for water companies and can only be handled by experts [12], [17], [22].

Another major drawback in a model-based approach is the uncertainty of model parameters such as the condition of the pipe. Leakage detection methods often based on the assumption that pipe condition remains unchanged. However, as pipe ages, roughness coefficient increases and results in a decrease in pipe diameter. Therefore, for a more realistic representation, any leakage detection model should include this factor [12].

Furthermore, some of the published methods were only tested on simple networks using synthetic data [22]. They could not apply directly to WDN as the use of nonlinear equations do not describe the actual behavior of the hydraulic systems [36].

E. DATA-DRIVEN APPROACH

The data-driven approach relies on the collection of data to perform signal processing and statistical analysis for leak detection. The advantage of this approach is as follows. It does not require any specific in-depth knowledge about the system. It only needs to learn from the historical data collected coupled with any statistical or pattern recognition tools [57].

The main disadvantage of this method is a large amount of data is needed to develop a classification or predictive model [22]. Moreover, anomaly data caused by burst or leaks may be scarce [65] that result in minority data. In addition to the problem of minority data, modeling of classification or predictive model may be further hampered by missing data, anomaly data from sensors, and communications and data noise [22], [65]. Hence, these suggest the need to have some hydraulic system knowledge in order to validate the data collected which must be reliable and conceivable.

As burst induced data differs significantly from data generated from water consumption, burst induced data are usually deemed as outliers [17]. The classification models are usually trained with this condition. However, due to seasonal change or festive season, water consumption will fluctuate. Water consumption peaks in summer due to the seasonal increase of residential population as well as an increase in the per capita water consumption due to higher temperature [66]. Thus, causing the classification model to have a lot of false alarms if the seasonal fluctuation is not taken into consideration [17], [22], [57]. However, as observed in [66], the night inflow follows a consistent trend and do not varies significantly during seasonal change. Thus, night inflow should be taken into consideration to improve the detection rate and accuracy.

III. PROPOSED METHODOLOGIES IN LITERATURES

The proposed methodologies in literature can be classified into the following categories; Prediction, Classification, Clustering, Model-Based, Statistical and Transient Signal Analysis. Among, the categories, prediction, classification, clustering and statistical methods can be grouped into the data-driven approach.

A brief overview of the methodologies surveyed is presented in Table 1. It gives the information of the proposed algorithm and the DMA used for testing. As seen in Table 1, methodologies proposed are tested on different DMA using different sets of data and do not necessarily include both leakage detection and localization.

Reviews of the surveyed methodologies are given in the next few subsections and summaries of their limitations are shown in Table 2 and 3.

A. PREDICTION

The idea of such a category is to perform a series of water demand/consumption/flow rate prediction ahead of time. The series of prediction is then analyzed in conjunction with actual readings to detect any discrepancies caused by abnormal flow [67].

Jung and Lansey [68] employ a Nonlinear Kalman Filter (NKF) to identify the system operational condition, estimate nodal group demands, and detect bursts. Jung and Lansey [68] seek to overcome solutions that are limited to networks supplied by gravity or under consistent operating conditions.

Their methodology begins with field measurements collection (pipe flows and pressure heads are generated from a hydraulic model of a real system for conditions with and without bursts) followed by using nonlinear Kalman filter to estimate the group nodal demand. There are two methods that Jung and Lansey [68] suggested to use in conjunction with NKF for anomaly detection. They are namely: Cumulative Sum (CUSUM) and Hotelling T^2. The former does not require pipe flow and pressure head estimates. But it will provide a long-term impact after a burst. The latter requires estimation of pipe flow and pressure head, but it provides a short-term impact after a burst.

Although the proposed methodology produced promising results, it was only tested on consistent operating condition with a simple hydraulic change. Additional investigation is needed to assess other types of operational changes, such as complex combinations of tank open/closure and pump operation. Jung and Lansey [68] also assumed hydraulic model is correctly modeled. The system parameters such as pipe roughness coefficients are known with certainty, and high demand events have not occurred. Such assumptions may not be realistic and valid for the real systems. Firstly, having a hydraulic model that represent the real system perfectly is impossible as there are several parameters that are hard to determine. Secondly, roughness coefficient changes as pipe age. Such a parameter cannot be known with certainty.

Karray *et al*. [69] also utilize Kalman Filter (KF) in their work to provide a reliable solution for inspecting

TABLE 1. Proposed methodologies over the past five years.

References	Proposed Methodology	Event Detection	Event Localization	Case Study	SD	ET	HD
[68]	Nonlinear Kalman Filter	✓		Modified Austin Network From Brion and Mays [114]	✓		
[69]	Predictive Kalman Filter, Time Difference of Arrival	✓	✓	Laboratory Testbed	✓		
[70]	EPR paradigm	✓		Unspecified DMA		✓	
[71]	Weighted Least Squares With Expectation-Maximization Algorithm	✓		1 DMA in North of England		✓	
				8 DMAs in North of England			✓
[57]	ANN, SPC, Bayesian Inference System	✓		1 DMA in UK		✓	
				5 DMAs in UK			✓
[27]	Ensemble CNN-SVM, Graph-Based Localization	✓	✓	WDN in Seoul		✓	
[73]	Pattern Matching, Associative Artificial Neural Networks	✓		WDN in UK			✓
[18]	Artificial Immune Network	✓	✓	2 case studies	✓		
				1 case study			✓
[75]	Clustering Algorithm by Rodriguez and Laio (Euclidean Distance)	✓		DMA in South China		✓	
[17]	Clustering Algorithm by Rodriguez and Laio (Cosine Similarity)	✓		DMA in South China		✓	
[77]	Multiclass SVM	✓	✓	Case study (Bentley Systems 2013)	✓		
				WDN in China			✓
[40]	K-nearest Neighbours		✓	DMA in Hanoi	✓		
				DMA in Limassol	✓		
				DMA in Nova Icaria		✓	
[54]	Bayesian Classifier		✓	DMA in Hanoi	✓		
				DMA in Nova Icaria		✓	
[78]	Graph Partitioning Algorithm with Flow Balancing Equation	✓	✓	EXNET Network	✓		
				Richmond Network	✓		
				DTown Network	✓		
				Colorado Springs Network	✓		
				WDN in Bangalore	✓		
[38]	Model-Based Methodology	✓	✓	DMA in Nova Icaria		✓	
[16]	Leakage Model	✓		Two different case studies from literature	✓		
[79]	Four Steps Statistical Method	✓		4 DMAs in North of Portugal			✓
				1 DMA in Greater Lisbon			✓
[4]	Joint Time-Frequency Analysis, Energy Based Localization	✓	✓	WDN in Singapore		✓	
[50]	Wavelet-Based Leakage Detection, Graph-Based Search Algorithm	✓	✓	WDN in Singapore		✓	
[80]	Interval Estimation	✓	✓	WDN in Yeongwol, South Korea	✓		
				WDN in Yangsan, South Korea	✓		
[81]	Wavelet Change-Point Detection	✓		Synthetic Data	✓		

SD refers to simulated data – data that are simulated by simulation. ET refers to engineered test – data collected with simulated events (e.g., opening a fire hydrant). HD refers to historical data – data collected based on real events.

pipe infrastructure. They proposed a Predictive Kalman Filter (PKF) that handles data compression to identify leakage. Karray *et al.* [69] explained that their PKF is a predictor combined with KF to reduce the communication cost of the wireless sensor network. The KF adopted estimates the pressure variation caused by leaks. If the difference exceeds

TABLE 2. Summary of proposed methodologies' limitations (part 1 of 2).

Authors	Limitations of Methodology
[68]	1. SPC techniques were better than proposed methodology for a system with consistent operating conditions. 2. Additional investigation is needed to assess other types of operational changes. 3. Several assumptions may not be realistic and may not hold in real systems. 4. Methodology not validated with a real system data. 5. May not be able to detect a small leak.
[69]	1. Force sensitive sensors used were not accurate enough. 2. Applicable only to above ground pipelines. 3. Only tested on a small testbed. 4. May not be able to detect a small leak. 5. Localization of burst is also highly dependent on the correct estimation of wave speed. 6. Equipment used is not ruggedized, may not be suitable to apply in the real-world condition
[70]	1. Presence of a significant leak can mask the occurrence of smaller leaks and can only be detected after the significant leak is resolved. 2. Degree of exceedance above the threshold is dependent on the requirements and experience of water utility operators as well as network consumption history. 3. Dependent on the data used which means that if an unexpected water demand occurs, it would be classified as an anomaly.
[71]	1. Did not take into account unexpected water demand. 2. Small leakages will not raise the alarm if the burst size is smaller than the normalized standard deviation. 3. Tuning may take some time before the model is reset according to an unplanned network configuration.
[57]	1. Extensive use of sensors. 2. Detection of the leak will be delayed if the pressure sensors are located far away from the leak. 3. ANN has to be constantly retrained and updated to ensure prediction accuracy. 4. Relies on a large amount of historical data which may be unavailable in a new network system. 5. Do not take into account unexpected water demand. 6. The system will rely on an operator's decision, confidence and expertise to define the suitable probability threshold value. 7. May not be able to detect small leaks.
[27]	1. Proposed CNN method to extract features could not remove noise well. 2. Requires a considerable computational time. 3. Localization algorithm also did not take into account external noise. 4. Methodology may not be applicable outside the time window (midnight). 5. Localization accuracy may be affected by inappropriate estimated wave speed.
[73]	1. Requires a vast amount of historical anomaly data to build up the knowledge/library. 2. Requires a certain level of expertise to extract profiles that are indicative of different events. 3. Overall performance is found not to be as good as using outlier detection based methods for WDN time series data.
[18]	1. Methodology requires a good calibrated hydraulic model. 2. It does not perform well with a real world burst event data. 3. The lack of historical burst records hampered accuracy. 4. Methodology does not seem to be able to detect small leak events. 5. Localization accuracy is influenced by the placement of sensors.
[75]	1. Needs a large amount of historical data to generate a daily series. 2. Data were taken at the specific time which makes their method only applicable in the time frame. 3. Such a method is unreliable when daily flow changes significantly due to factors such as seasonal change. 4. Comparison of *x-mean* is also not exhaustive enough. 5. Do not take into account unexpected water demand. 6. The true positive rate was only 71%. 7. May not be able to detect small leaks.
[17]	1. Unable to address the issue caused by unexpected water demand. 2. Ability to detect small leaks remains unknown. 3. The true positive rate was only 71%.
[77]	1. Selecting the optimal number of clusters which has a big impact on clustering classification. 2. By having more clusters, it would also indicate that user has to train more SVM models which result in a more computational cost. 3. This method only identifies possible leak zone. 4. May not be able to detect small leaks.
[40]	1. Based on several assumptions which cannot hold in a real system.
[54]	1. Based on several assumptions which cannot hold in a real system.

the permissible threshold, an alarm is raised. To locate the position of the leak, Karray *et al*. [69] suggested a method hybridizing the physical principle of leak wave propagation and the time difference of arrival, as water leaks, it creates a pressure wave along the pipeline. Karray *et al*. [69] will choose the two sensors to detect the leak based on the

TABLE 3. Summary of proposed methodologies' limitations (part 2 of 2).

Authors	Limitations of Methodology
[78]	1. Unrealistic assumption that all flow measurements are without noise and the flow balancing equation will result in zero if there is no presence of leakage. 2. There is a need to establish a statistical threshold for sensors' error. However, such implementation will give rise to a false alarm and miss detection. 3. Methodology is a very iterative process and does not seem to be feasible for large water distribution network. 4. Network will require a large number of flow sensors otherwise much manual work is needed to check the flow rate.
[38]	1. Unrealistic assumption that leaks only occur at nodes. 2. Can only pinpoint the node nearest to leak rather than the leak location itself. 3. Methodology is hard to model and require a lot of simulations and calibration. 4. May not be able to detect small leaks.
[16]	1. Hydraulic model, as well as the leakage model, have to be extremely accurate before this methodology can be adopted. 2. Only tested using simulated data.
[79]	1. Did not take into account large consumers. 2. Dependent on past observation, true positive and false positive rate. 3. May not be able to detect small leaks.
[4]	1. Optimal window time is critical for this system which affects the detection capability and localization accuracy. 2. Method can only locate the possible leak zone if measurements points are far away. 3. May not be able to detect a small leak. 4. By applying a wavelet transform there is a possibility that an insignificant pressure wave signal is emphasized. 5. Signals may be influenced by background noise. 6. High investment cost for sensors.
[50]	1. Only tested on data collected over 2 hours with nine controlled events. 2. May not be able to detect a small leak. 3. Burst induced pressure signals may be influenced by background noise or other events happening in the nearby vicinity. 4. High investment cost for sensors. 5. By applying a wavelet transform there is a possibility that an insignificant pressure wave signal is emphasized. 6. Average localized error was not accurate enough which will require other techniques to identify the exact location of burst. 7. Localization of burst is also highly dependent on the correct estimation of wave speed.
[80]	1. Did not take into account unexpected water demand. 2. Localization technique requires the system to have at least four sensors. 3. Only tested using modified data.
[81]	1. Did not take into account unexpected demand. 2. By applying a wavelet transform there is a possibility that an insignificant signal is emphasized. 3. May not be able to detect small leaks.

previous KF. Then making use of the following equation to estimate the leak location.

$$x = \frac{L - C\Delta t}{2} \tag{6}$$

where x is the distance from the nearest sensor node, L is the distance between 2 sensors, C is the wave propagation speed, and Δt is the time difference of pressure signal arriving from the nodes.

The main issue with this methodology is that it was only tested on a small testbed. Furthermore, it is only applicable to above ground pipelines with limited use for underground pipelines. In addition, caution has to be taken when using estimated wave speed to calculate TDoA. A study carried out by Srirangarajan *et al*. [50] shown that a 10% error in wave speed estimation can result in a maximum 20% error in localization.

Laucelli *et al*. [70] proposed an Evolutionary Polynomial Regression (EPR) paradigm for event detection that aims to reproduce the behavior of a WDN using online data recorded by low-cost pressure and flow devices. As explained by Laucelli *et al*. [70], the behavior of a WDN is represented by the average flow rate. If the observed values of water consumption are higher than the maximum flow rate predicted, then such value is an anomaly. In the case study, Laucelli *et al*. [70] assumed a probability density function to represent the possible behavior of the network statistically. To give more weights to predicted values, predictions are characterized by a normal distribution with average predicted flow rate and a standard deviation. The observed flow rate is then used to calculate the cumulative probability. If the value exceeds the threshold (value is based on all cumulative probability of maximum predicted flow rate for each time step), then an anomaly is detected.

However, as explained by Laucelli *et al*. [70], the presence of a significant leak can mask the occurrence of small leaks. Thus, small leaks can only be detected after the significant leak is resolved. As methodology relies on predicted and observed value to detect an anomaly, the degree of exceedance is dependent on the requirements and experience of water utility operators and network consumption history. Thus, if an unexpected water demand occurs, it would be classified as an anomaly since the degree of exceedance depends on network consumption history.

Ye and Fenner [71] suggested a weighted least squares with Expectation-Maximization (EM) algorithm for burst detection in UK water distribution system and investigated the use of inlet flow measurement for burst detection at the DMA level.

To model the dynamics of flow pattern by a simpler and effective function, Ye and Fenner [71] proposed to resample the flow measurement into two-dimensional parallel flow sets to check the flow rate at the different time of the days. In the paper, Ye and Fenner [71] decomposed the parallel flow set into a weekly mode to minimize the standard deviation. By using the EM algorithm, unlabeled flow data are manually separated into normal or burst data. The EM algorithm uses a set of weights to represent the importance of each data points to the model and model parameters are estimated based on weighted least squares. The algorithm starts with equal weights to estimate the model parameters. After the estimation, weights are updated by examining the model fitting error. After weights are updated, model parameters are re-estimated again. This iterative process only stopped after the model parameters converge or a defined minimum error is reached. Burst can then be detected if the difference between actual and prediction passes a threshold.

However, Ye and Fenner [71] did not take into account unexpected water demand that may cause false alarms. Moreover, small leakages will not raise the alarm if the burst size is smaller than the normalized standard deviation. Although, Ye and Fenner [71] claimed their method could adapt to unplanned network configuration, the tuning time is subjected to the length of data points fitted to the model. Thus, adjustment usually takes some time before the model is reset. Lastly, the methodology is highly sensitive towards sensors' noise and faults.

Romano *et al*. [57] presented their methodology that can detect pipe burst or events that possess the similar abnormal pressure or flow variation at a DMA level. The objective of [57] was to develop a faster and higher reliability detection system. Their methodology makes synergistic use of artificial intelligence techniques, statistical data analysis tools, Statistical Process Control (SPC) techniques and Bayesian inference systems for inferring the probability of a pipe burst or other events. Their methodology consists of three subsystems which inspects the data at a different level. The first subsystem (Discrepancies Based Subsystem) measures the difference between the actual value of pressure/flow signals values and predicted pressure/flow signals values by ANN. A statistical limit is set to detect an occurring event. The second subsystem (Boundary Based Subsystem) also focused on identifying a pressure or flow variation. But it can only identify medium or large deviations. In this subsystem, check are done for incoming data that exceed statistical boundaries. The third subsystem (Trend Based Subsystem) focus on the identification of pressure or flow deviation caused by developing events. A control chart is used to monitor how the mean of pressure or flow measurements relative to the particular hour window changes over time and determine any out of control situation. These three subsystems will generate their analysis results and used as input for the Bayesian inference system.

The main drawback of [57] is the extensive use of sensors. Firstly, this methodology is highly sensitive towards sensors' noise or faults; unnecessary noisy data can result in a false alarm. Secondly, detection of the leak will be delayed if the pressure sensor is located far away from the leak as demonstrated in [57]. Such a system also relies on a large amount of historical data that may be unavailable in a new network system. Also, to account for the fluctuation of water demand, ANN has to be constantly retrained and updated to maintain the level of prediction accuracy.

Moreover, it does not take unexpected water demand into account [17]. Such a scenario will result in a false alarm [68] since the system condition differs from the state where the control chart is developed. Finally, such a system will ultimately rely on the operator's decision, confidence and expertise to define the suitable probability threshold value for leakage detection (trade-off between true and false alarm).

B. CLASSIFICATION

In a classification problem, a classifier is trained based on a feature set that best identifies the unique characteristics of different event types to categorize future event into normal or abnormal event [72]. Kang *et al*. [27] proposed an ensemble Convolution Neural Network-Support Vector Machine (CNN-SVM) and graph-based localization for leakage detection in the water distribution system. Their motivation was to improve the detection and localizing accuracy. In their method, features extracted from CNN are used as inputs to a Multi-Layer Perceptron (MLP) and SVM. The outputs from SVM was converted to a probability using Platt's trick. Subsequently, the results from the two classifiers are combined using a method which multiplies the probability of each model with their corresponding weighting factors [27]. In localization, the graph-based algorithm from Srirangarajan *et al*. [50] was used with two proposed changes. First, a method to search for the nearest node to the leak location which reduces computational time. Second, imposing the search range limitation of the virtual node. By using a generalized cross-correlation with a maximum likelihood weighting function, Time Difference of Arrival (TDoA) was calculated. Subsequently, using the manually estimated wave speed and length of pipe, the location of the leak can be estimated using the same formula as [68] (refer to equation 6).

The drawback of the proposed CNN method is feature extraction could not remove noise well. In addition, the proposed method required considerable computational time. Localization algorithm did not take into account external noise as burst were emulated in midnight where they expected noise to be minimal. Thus, this methodology may not be applicable outside the time window (midnight) [92]. Finally, incorrect estimation of wave speed to calculate TDoA will affect the localizing accuracy.

Mounce *et al.* [73] proposed pattern matching and associative artificial neural networks for water distribution time series data to identify anomaly and classify its event type. The primary goal was to test the applicability of pattern matching in the domain of leak detection. Data collected from different sensors are first processed so that they can be compared. Key variables for different event type need to be identified, and profiles from past events must be placed in the library. For every new incoming data, a similarity-based search will be performed so that similar-shaped profiles of different amplitudes are matched. If the matched is over a given threshold, the user will be alerted.

However, such a methodology requires a vast amount of historical anomaly data to build the knowledge or library. It also requires a certain level of expertise to extract profiles that are indicative of different events. Most importantly, Mounce *et al.* [73] mentioned that overall performance was not as good as using outlier detection based methods for WDN time series data.

Tao *et al.* [18] suggested the idea of using an artificial immune network to detect burst. The main drive was to test the applicability of artificial immune network that has been successfully applied in other fields. In this work, multi-level ANN was used. The first level was to identify the occurrence of a leak and the second level was to estimate the magnitude and location of leaks. In the paper, the methodology was described in four steps, data collection, data processing, features extraction and training of the classification model. After the artificial immune system network is established, it requires further calibration through other burst data until accuracy meets a satisfactory level. Once the network is done, the location of the burst is identified through the principle of nearest neighbors.

The prerequisites of such methodology are a well calibrated hydraulic model and a vast amount of historical data. The lack of such prerequisites will significantly hamper the accuracy of this methodology. In addition, the methodology cannot detect small leak events as demonstrated in the first case study [18]. The localization accuracy is also influenced by the placement of sensors. The result can only suggest the possible leak zone. Therefore, requiring a further inspection to locate the exact leak point.

C. CLUSTERING

This type of approach belongs to the unsupervised methodology as clustering does not require knowledge of all leaks for a practical application of algorithm [74]. The idea of this category is to divide the WDN or data into different clusters. It is followed by adopting another strategy to identify the possible leak zone or leakage. Wu *et al.* [75] motivation were to develop an unsupervised clustering based method that detects bursts inside a DMA with multiple inlets and outlets. Their leak detection methodology consists of two steps. The first step was to identify outliers from all data. The second step was to differentiate the leak from other types of outliers. Data from a specific time in a consecutive time series was extracted to form new data series and transformed to reduce variation. Clustering of vectors was then based on evaluating similarities; a vector is considered an outlier if it does not belong to any cluster. In their paper, the clustering algorithm adopted by Rodriguez and Laio [76] is clustered by fast search and density peaks. Once outliers are identified, burst identification is carried out by comparing outliers with mean of vectors.

$$mean = (m_1, m_2, m_3, \ldots, m_n) \tag{7}$$

where n is the number of the flow meter in a DMA, m_i is the mean value of column i in a detection matrix where an outlier is detected, and the outlier is defined as

$$x = (x_1, x_2, x_3, \ldots, x_n) \tag{8}$$

Wu *et al.* [75] suggested if *x-mean* was more than 0 with n number of anomalously large element then it was due to warm/sunny weather or festive season. If *x-mean* was less than 0 with 0 number of anomalously large element then it was due to sudden cold/wet weather. If outlier belongs to neither of the categories, then it was considered as a leakage.

Although clustering excludes the prediction process, it still needs a large amount of historical data to generate a daily series. Secondly, Wu *et al.* [75] extracted data at a specific time where variation in flow is at minimal (6 am in the morning). It makes their method only applicable in the morning [92]. In addition, such a method is unreliable when daily flow changes significantly due to factors such as seasonal change [17]. The comparison of *x-mean* is also not exhaustive enough. Mean vector calculated cannot reflect the overall variation of flow measurement with insufficient data. It also does not take into account unexpected water demand. As demonstrated [75], large consumers increased the demand unexpectedly and were classified as a leakage event.

Wu *et al.* [17] then extended his work to address issues in his proposed methodology [75], namely, to reduce demand of historical data to omit data cleaning process and to account for weather changes, festivals and periodic changes in water demand. The essence of the methodology remained the same that utilized the clustering algorithm by Rodriguez and Laio [76], But the distance measurement was changed to cosine similarity instead of Euclidean distance. As explained by Wu *et al.* [17], when events such as weather changes, festivals are reflected as normalized vectors, they usually have an almost similar angle as vectors from similar time frame from different seasons with higher amplitude. As such, cosine similarity will be able to classify them as the same cluster, thus reduces the number of false alarms.

Although Wu *et al.* [17] reduced false alarm caused by weather changes, festivals and periodic changes in water demand, they are still unable to address the issue caused by unexpected water demand. In addition, whether this method can detect small burst remains unknown. As compared to the previous methodology [75], the true positive rate did not have any improvement while the false positive rate has a slight improvement.

Zhang *et al*. [77] proposed to use multiclass SVM to improve the effectiveness and efficiency of leakage detection in large-scale WDN. Their methodology began with dividing the network into several leakage zones using k-means clustering. Each zone was labeled as a multilevel category label for multiclass SVM. Monte Carlo method was then used to generate leakage events that were analyzed by hydraulic model simulation. The results of the leakage event were then used as training samples. The trained multiclass SVM can identify likely leakage zone according to field observed flows and pressures. As SVM is a binary classification problem, Zhang *et al*. [77] used one against one method for multiclass classification that involves training of multiple SVM with training samples from different categories. Although Zhang *et al*. [77] rely on a supervised method to detect the leak zone, it first clusters the WDN into different clusters thus it belongs to the clustering cateory.

The drawback of this method is the number of clusters to divide the network. As the choice of initial cluster is random and if relatively isolated points are chosen, it will cause a big impact on the clustering classification. Moreover, there is a user's dilemma in choosing the number of clusters. More clusters would mean higher localization accuracy but would have the possibility of more suspected leakage zones. By having more clusters, the user has to train more SVM models causing higher computational cost. Furthermore, since this method only identifies the possible leak zone, other methods will need to be used to identify the exact leak location.

Soldevila *et al*. [40] proposed a mixed model based and data-driven approach for leak localization in water distribution networks to avoid complications faced when using model-based leak detection method. The hydraulic model using EPANET hydraulic simulator was modeled [40]. The model was assumed to represent the WDN after the calibration. The model was then used to generate data in the residual space for each possible fault and different operating and uncertainty condition. Nodes that leak had a similar effect on the pressure sensors were grouped as the same class. Then, the data were used as training data for the classifier which is the K-nearest neighbors.

Soldevila *et al*. [54] then proposed a Bayesian classifier for leak localization in a water distribution network. This methodology was built on the earlier methodology whereby a hydraulic model has to be built. This data was then used to calculate probability density functions. In the online stage, Bayesian classifier provides the time-dependent posterior probability of every possible leak. The only difference with the earlier methodology is that a Bayesian classifier was used. The results showed that the Bayesian classifier is more accurate than the previously used K-nearest neighbors.

The issue with both methodologies [40], [54] is they are based on too many assumptions that may not hold for real systems. Soldevila *et al*. [40], [54] assumed that pressure sensors are installed in inner nodes of the network, the hydraulic model is perfectly tuned, and leaks only appear on the network nodes. Firstly, the methodology cannot be used without pressure sensors. Secondly, the hydraulic model cannot be modeled perfectly due to unknown parameters. Thirdly, leaks can occur on pipes instead at the nodes. Such a method can only suggest a possible leak zone near the identified leaking node and other methods will have to be used for pinpointing the exact leak location.

Rajeswaran *et al*. [78] proposed a graph partitioning algorithm for leak detection in water distribution network capable of localizing the leak. Their motivation was to improve leak detection rate in a vast network. The idea [78] is to divide the water network into smaller networks using a graph partitioning algorithm. Subsequently, the flow rate (inflow and outflow) of each subnetwork was measured and compared using the flow balancing law. If inflow equals to outflow, then there is no leak. Otherwise, there is a leak.

Rajeswaran *et al*. [78] assumed that all flow measurements do not have noise and the flow balancing equation will result in zero without leakage. It is not true as sensors' data will have the presence of noise. Thus, there is a need to establish a statistical threshold for sensors' error. However, such implementation will give rise to a false alarm and miss detection. In addition, this methodology is a very iterative process and does not seem to be feasible for a large water distribution network even though the motivation is to improve leak detection rate in a vast network. Furthermore, since the methodology relies on flow balancing law, it suggests a need for a large number of flow sensors. Most importantly, Rajeswaran *et al*. [78] did not validate their methodology in a real system.

D. MODEL-BASED

The model-based method involves the use of mathematical functions or formulas to represent or replicate the operation of a pipe network. It can determine the approximate leakage location by comparing pressure measurement with their estimation obtained using the hydraulic network model.

Perez *et al*. [38] presented a model-based methodology using pressure sensors. The methodology was built on the principles of model-based diagnosis and attempted to enhance fault isolation by using fault sensitivity analysis. The methodology started with a simulation of possible faults using a hydraulic model. Perez *et al*. [38] then computed the sensitivity-to-leak matrix that contained the theoretical fault signatures used by the leak localization methodology. Pressure measurements were collected to generate residuals to compare against the theoretical fault signatures of all potential leaks. Finally, the most probable node that leaks will be identified. Detailed equations and formulations can be found in their paper.

Similar to the papers [40] and [54], Perez *et al*. [38] also assume leaks to occur at nodes although leaks can also occur at pipes. The methodology can only pinpoint the node nearest to leak rather than the leak location itself. Further inspection will conduct to locate the exact leak location. In addition, this method is hard to model and require a lot of simulations and calibrations.

Adedeji *et al*. [16] presented a leak detection and estimation algorithm for loss reduction in water network with the focus of detecting background leakage. The algorithm incorporates a leakage model into a water distribution network hydraulic simulation model to estimate the network flows. It included leakage outflow at each node and each pipe. The algorithm began with the loading and reading of the supplied water distribution network data. Followed by a hydraulic analysis through the modeling of water network topology and solving the resulting model using an iterative Newton-based methodology. The nodal leakage outflow was then computed and compared with the defined threshold. If the threshold was not exceeded, then no leakage was reported. Otherwise, leakage was reported. All leakage flow at each of the pipes connected to the node was then computed. If the estimated leakage flow is high, then the pipe is deemed to be a critical pipe that needs to be repaired.

The main drawback of this methodology is the hydraulic model, as well as the leakage model, have to be very accurate before it can be adopted. Moreover, this methodology was only tested using simulated data.

E. STATISTICAL

This method relies on statistical theory to analyze collected data to identify a leak [12], [22]. Loureiro *et al*. [79] presented a four steps statistical method to detect a leak. The main motivation was to develop a detection module that is easy to implement without using a large dataset. The first step was the collection of instantaneous or high-frequency flow data with a time step of fewer than 15 minutes from the existing Supervisory Control and Data Acquisition (SCADA) or telemetry system. In the second step, data that involves the detection and correction of anomalies were validated, cleansed and normalized to a regular time step. In a DMA with multiple flow meter for inflow and outflow, normalized data were combined to estimate the network consumption and continuously large consumers were filtered to allow the focus of other outliers. In step 3, the methods to detect an anomaly in the flow time series were employed. Before applying a statistical method to detect an anomaly, cluster analysis was done where data were separated into work days and weekend days and each day data was then split into 96 15 mins intervals. Loureiro *et al*. [79] highlighted that splitting data into different categories would allow more efficient detection of an anomaly as evidence shown that most DMA displays a periodic behavior during the day. Since outlier detection requires a threshold, a ROC curve was adopted to determine the appropriate threshold. In step 4, methods were applied to detect different types of anomalous events.

In the study, Loureiro *et al*. [79] did not take into account large consumers or unexpected water demand. In such events, this methodology may classify them as anomaly and results in a false alarm. The concept as explained by the Loureiro *et al*. [79] is dependent on past observation, true positive and false positive rate. Thus, similar to [57], it will ultimately rely on the operator's decision, confidence and expertise to define the suitable threshold for leakage alarm.

F. TRANSIENT SIGNAL ANALYSIS

The transient signal analysis builds on top of the detection of the transient pressure wave. Zan *et al*. [4] motivation was to develop a cost-effective wireless sensor network for real-time monitoring, analyzing and modeling of urban water distribution systems. Zan *et al*. [4] utilized the joint time-frequency analysis of the pressure transient signals for leak detection. The first step was the use of one-dimensional wavelet transform to remove high-frequency noise on the transient pressure signal acquired. Short-Time Fast Fourier Transform (STFT) was then applied to extract the leak induced features. To minimize the effect of spectral leakage, Blackman Window was adopted to choose the optimal window for the STFT. The spectrogram was then computed which shows the signal energy and Gabor transform was used to remove an unnecessary portion of the spectrogram while the same Blackman Window was used to denoise. After which, Zan *et al*. [4] selected the optimal frequency range for event detection based on the interpretation of spectrogram. For the localization of leak, Zan *et al*. [4] proposed an energy-based localization approach based on estimates of distances between known measurement points and unknown leak. Zan *et al*. [4] calculated the mean intensity, variance and standard deviation of given distances and linear regression was applied to model the relationship. The intensity values from real pipeline leakage were applied to approximate the distance between sensor nodes and leak and to validate the relationship formulation. Once the distance matrix was obtained, sensors were ranked according to scores. The nearest sensor to the leak has the highest score.

The limitation of this methodology is the detection capability of the system is dependent on the spectral resolution of STFT while localization accuracy is dependent on temporal resolution. This suggests that optimal window time is critical for this system. Moreover, the burst induced pressure signals can be influenced by background noise or other events happening in nearby vicinity for a complex network [22]. In addition, high investment cost is required to install a large number of sensors with high sampling frequency to collect transient signals [17]. Although negative wave pressure is useful in the detection of pipe bursts, it cannot detect a slow/small leak as it does not generate a distinct pressure reduction signal [12], [55].

Furthermore, by applying a wavelet transform, there is a possibility that insignificant pressure wave signal can cause false alarm [55]. Since localization approach is based on the distance between measurement points and leak, it implies that method can only locate the possible leak zone if the measurements points are far away. Thus, requiring a further inspection to locate the exact leak location.

Srirangarajan *et al*. [50] proposed a wavelet-based leakage detection and localization method which aim to detect and localize events in a WDNs based on pressure traces gathered by a dense wireless sensor network. Raw signals were first de-noised using wavelet de-noising. A four-level decomposition was then performed on the denoised signal to identify the transient event. Subsequently, identified transient events are further classified into true positive and false positive by using wavelet coefficients and Lipschitz exponents. For the localization of leak, a graph-based search algorithm was proposed that utilized the TDoA of burst transient at measurement points. The first step assumed burst occurs at one of the nodes and a global search was done to find the nearest node where the leak occurred. The second step performed a local search around the nearest node to estimate the most probably leak location.

Although this methodology is quite promising, it was only tested on data collected over two hours with nine controlled events. In addition, bursts were emulated above ground and did not emulate underground pipe burst. Thus, the applicability of this methodology on underground pipe burst remains unknown. As this methodology relies on capturing transient signal and wavelet transform, it has the same limitation as the methodology presented in [4] discussed earlier. Localization of burst is also highly dependent on the correct estimation of wave speed. From their results, the average localized error was not accurate enough to determine the exact location of burst but only the section of the pipe thus requiring other techniques to identify the exact location of burst.

Kim *et al*. [80] aim to overcome the three problems in detection of small leakage: lack of robustness under the noisy conditions, low accuracy caused by loss of time information, and absence of confidence bound. Kim *et al.* [80] proposed a leak detection and localization using interval estimation for the water distribution network. Kim *et al.* [80] only adopted the use of pressure values and aims to detect small leaks by simulating small leak data based on burst data. First, measured pressure values were filtered by the KF and were shifted by subtracting the average pressure followed by the cumulative integral of the shifted one. As explained by Kim *et al*. [80], it helped them to visualize the trend of pressure drop after a leak had occurred. Floor function with three parameters and curvature function are then applied. After the occurrence time was obtained, statistical techniques were then applied to find the segment containing the leak point with confidence bound.

Although Kim *et al*. [80] had a low rate of false alarm, they did not take unexpected water demand into account, and such occurrence will result in a false alarm. Subsequently, localization techniques require the system to have at least four sensors.

The methodology cannot work if only one sensor is available. If only two sensors are available, no points exist for verification. If three sensors are available, the confidence of interval estimation is very low. Again, this method faces the same limitation as [4] and [50] when it comes to using pressure drop to detect leaks. Finally, the methodology was only validated using modified data.

Christodoulou *et al*. [81] aimed to address the automatic detection of water losses in WDN through dynamic analysis of time series. Christodoulou *et al*. [81] proposed using wavelet change point detection for the detection of anomalies in WDN. The water consumption time series was first processed macroscopically to identify the time periods of concern and then microscopically to zoom in on possible consumption anomalies. At the macroscopic level, the original signal was transformed using a continuous wavelet transform with Daubechies wavelets. Change-point detection was then applied, and detected anomalies were then converted into change point scores.

Change-point detection refers to the identification of whether a change (or several changes) has occurred in a time series and the time when such change occurs. Change-point detection methods may relate to changes in the mean, variance, correlation, density, or slope of the signal [82]. The change point detection method was not discussed in detail in [81]. At the microscopic level, the identified major change points were then examined to confirm the findings. Such methodology also did not take into account unexpected water demand and had not been tested on a real system. As explained earlier, the use of the wavelet transform may amplify an insignificant signal affecting the subsequent analysis [55].

IV. ACCURACY ASSESSMENT

This section will provide a discussion on the accuracy assessment that the authors used to access their algorithms. It is important to state that not all authors provide a measurement to assess their methodologies. Other accuracy assessments which are useful will also be discussed. From the surveyed literature, the most commonly used accuracy assessments were the True Positive Rate (TPR), False Positive Rate (FPR) and Receiver Operating Characteristics (ROC) curve [17], [27], [57], [75], [79].

In the case of leakage in a WDN, a True Positive (TP) can be referred to as detection of an actual event taking place. On the other hand, a True Negative (TN) is referred to correct identification of no actual event taking place. Therefore, a False Postive (FP) is a false alarm. Whereas, a False Negative (FN) is a missed detection. TPR can be referred to as the percentage of how well can a model identify a true event. On the other hand, Tue Negative Rate (TNR) may be referred to as the percentage of how well can a model identify a negative.

Positive Predictive Value (PPV) is the percentage of correct identification of true event over all event alarms raised. Negative Predictive Value (NPV) is the percentage of correct identification of a true negative over all negative alarms raised.

TABLE 4. Accuracy assessment used by different authors.

Authors	Detection Accuracy Formulation	False Alarm Formulation
[68]	TPR	Average false alarm per day
[57]	TPR, Accuracy	FPR
[27]	Accuracy, TPR, Area Under Curve (AUC)	FPR
[75]	TPR	FPR
[17]	TPR	FPR
[79]	ROC	

TABLE 5. Example of a confusion matrix.

	Actual		
	Leak	No Leak	
Predicted	TP	FN	TPR/Recall
	FP	TN	TNR/Specificity
	PPV/Precision	NPV	**Accuracy**

TABLE 6. Cohen's kappa interpretation.

Cohen's Kappa Value	Interpretation
≤ 0 to 0.20	No Agreement
0.21 to 0.39	Minimal
0.40 to 0.59	Weak
0.60 to 0.79	Moderate
0.80 to 0.90	Strong
0.91 to <1.00	Almost Perfect
1.00	Perfect

A visual representation of such terminologies can be seen in a confusion matrix given in Table 5.

The equations to calculate each of the terminologies is given as follow.

$$TPR/Recall = \frac{TP}{TP + FN} \tag{9}$$

$$TNR/Specificity = \frac{TN}{TN + FP} \tag{10}$$

$$FPR = 1 - TNR = \frac{FP}{FP + TN} \tag{11}$$

$$PPV/Precision = \frac{TP}{TP + FP} \tag{12}$$

$$NPV = \frac{TN}{TN + FN} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{14}$$

However, accuracy may not be an optimal indicator as it can be biased when one class's proportion severely outweighs the other class's proportion. For example, dataset contains data points for both leak and no leak. The leak only has 10 data points, but no leak has 1000 data points. If a model identifies all data points as no leak, it still have a high classification accuracy of 99%. As such, there are other indicators that eliminate such biases such as F1 score, Cohen's Kappa, g-means, Matthew Correlation Coefficient (MCC), and ROC.

The F1 score is simply the harmonic mean of precision and recall [83], [84] and is given as

$$F1 - score = \frac{2(Precision^{*}Recall)}{Precision + Recall} \tag{15}$$

Cohen's kappa is a statistical measure on the degree of agreement between two raters taking into account of chance [83], [85], [86]. In this case, it can be referred to how well a model agrees with the ground truth. It ranges from -1 to 1 where the higher the value, the higher the agreement between 2 raters. The interpretation is given as follow [87].

$$Cohen'sKappa = \frac{Accuracy - P_e}{1 - P_e} \tag{16}$$

where P_e can be calculated based on the confusion matrix. Since the confusion matrix is a square matrix thus P_e can be computed as follow.

$$P_e = \frac{\sum_{i=1}^{n} [(\sum_{j=1}^{n} e_{ij})(\sum_{j=1}^{n} e_{ji})]}{(\sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij})^2} \tag{17}$$

where e is the element in the matrix and indices represent the position of the element in the matrix. The first alphabet of the indices (i) represent the row of the matrix and the second alphabet (j) represent the column of the matrix. Since it is a square matrix, n can either be the number of rows or the number of columns. Using the confusion matrix, the equation can be simplified into

$$P_e = \frac{(TP + FN)^{*}(TP + FP) + (TN + FP)^{*}(TN + FN)}{(TP + TN + FP + FN)^2} \tag{18}$$

In the context of binary classification, G-means is the geometric mean of recall and specificity given as

$$g - mean = \sqrt{Recall^{*}Specificity} \tag{19}$$

MCC is a correlation coefficient between observed and predicted classification [88]. It returns a value of -1 to 1. A MCC of 1 indicates perfect agreement, MCC of 0 is expected for a prediction no better than random, and MCC of −1 indicates total disagreement between prediction and observation [89].

$$MCC = \frac{(TP^{*}TN) - (FP^{*}FN)}{\sqrt{(TP+FN)(TN+FP)(TP+FP)(TN+FN)}} \tag{20}$$

A ROC curve is a two-dimensional graph with TPR as the y-axis and the FPR as the x-axis that gives a visual overview of how well a classifier perform [90]. The closer the line to the top left border, the better the accuracy with lower false positive rate. If the plotted curve is a diagonal line across the graph, it infers that the classifier has no better odds of detecting something than a random flip of a coin. In the case of WDN, some authors established a minimum leakage threshold by finding the threshold along the ROC curve that maximizes the TPR at the highest acceptable FPR [57], [79].

To compare with different classifiers or different parameters used in the same classifier, one can merely plot the ROC curve and calculate the AUC. The larger the area, the better the classifier. The AUC can be approximated by the use of trapezoidal rule [91] given as

$$AUC = \frac{\Delta x}{2}[f(x_0)+2f(x_1)+\ldots+2f(x_{n-1})+f(x_n)] \tag{21}$$

where Δx is the size of each partition.

V. RESULTS BY PROPOSED METHODOLOGIES

This section presents the results tabulated by different methodologies. The results include the accuracy, the time taken to detect the event and the rate of false alarm. As methodologies were applied using different case studies, a direct comparison between these methodologies is illogical. Therefore, the discussion on each of their results will be performed individually.

As discussed earlier, water is a precious resource which should be conserved. Therefore, a good detection algorithm should have, 1) a high detection rate, 2) short detection time and 3) low false positive.

Jung and Lansey [68] achieve high accuracy with low false alarm but detection time for most events were more than two hours. Localization accuracy achieved by Karray *et al.* [69] is unreliable as the methodology was tested on a small testbed. Romano *et al.* [57] obtained around 80% accuracy with and up to 10% FPR. Two of the events were detected after the thirteenth hour. Kang *et al.* [27] had an extremely good detection and localization accuracy. But results were validated on data collection at midnight. Tao *et al.* [18] only achieved an accuracy of 48.3% when the proposed methodology was applied to historical data. Wu *et al.* [17], [75] can only achieve an accuracy of 71.43% and identifying the possible leak zone. Zhang *et al.* [77] had a high detection rate but did not include any localization accuracy. On the other hand, Perez *et al.* [38] had an approximate 300m localization error. Loureiro *et al.* [79] used different outlier detection method on different time window with results presented in a range. From their results, using a larger time window generally achieved a better accuracy and lowered false positive rate. However, the false positive rate remains rather high.

Zan *et al.* [4] and Srirangarajan *et al.* [50] achieved a 100% accuracy rate with rather good localization accuracy. Although Srirangarajan *et al.* [50] had a 10% FPR, it can be argued that an FP is always better than an FN. However, they [4], [50] tested their methodology through engineered events and may have different results during a real-life event.

VI. DISCUSSION

As discussed in Section III, most of the proposed methodologies have the following limitations.

1) Not validated in real system data or require more investigation for different types of operating conditions;
2) Unrealistic assumptions (such as pipe roughness coefficients and other system parameters are known with certainty);
3) Unexpected water demand;
4) Emulated burst instead of a small leak;
5) A large number of sensors required;
6) Localization of leak is not accurate enough.

Trying to overcome these limitations is no easy task, and the uncertainties and complications that come with it will be described as follow.

Limitation #1 appears easy to solve since it only revolves around the different type of fault scenario with different operating conditions to be decided by the researchers. Although water utility management personnel can allow a researcher to create a test event, they will not create a real leak in an existing pipeline due to the risk and cost involved. Furthermore, they will not allow any experiment to disrupt the entire water pipelines. However one should always attempt to validate their algorithms against engineered test events and historical dataset. By validating an algorithm against engineered test events, it proves the validity and practicality of the proposed algorithm. By validating against the historical dataset. it proves the reliability of their methodology if they can detect historical events.

The main issue with Limitation #2 is the inability to perform modeling if no assumptions can be used in model-based methodologies. To avoid making unrealistic assumptions, one can attempt to use data-driven methods. The most attractive characteristic of the data-driven approach is it does not need any specific in-depth knowledge about the system. Although the data-driven approach has several disadvantages, they can be solved or minimized. Methods such as cost-sensitive methods [93], [94] or data sampling methods [95]–[97] have been employed in other domains to solve minority data issues. Thus it is worthwhile to apply such methods to solve or minimize the issue of minority data in WDN.

If data-driven methods are used, it is of utmost importance that data integrity is high [98], [99]. Unfortunately, during the data collection phase, one will always encounter missing or anomaly data caused by noise or sensor's fault [100], [101].

TABLE 7. Detection time by different methodologies.

<table>
<tr><th>Authors</th><th>Case Study</th><th colspan="7">Detection Time</th></tr>
<tr><td rowspan="6">[68]</td><td rowspan="6">Simulated Data</td><td></td><td colspan="3">Consistent Condition</td><td colspan="3">Varying Condition</td></tr>
<tr><td rowspan="2">Proposed Methods</td><td colspan="3">No. Events</td><td colspan="3">No. Events</td></tr>
<tr><td><1hr</td><td>1-2hr</td><td>>2hr</td><td><1hr</td><td>1-2hr</td><td>>2hr</td></tr>
<tr><td>NKF + Hotelling T^2 (1)</td><td>32</td><td>0</td><td>0</td><td>28</td><td>0</td><td>1</td></tr>
<tr><td>NFK + CUSUM (2)</td><td>13</td><td>20</td><td>54</td><td>13</td><td>16</td><td>58</td></tr>
<tr><td>(1)+(2)</td><td>33</td><td>7</td><td>47</td><td>29</td><td>8</td><td>50</td></tr>
<tr><td rowspan="3">[71]</td><td rowspan="3">Engineered test</td><td colspan="7">No. Events</td></tr>
<tr><td colspan="3"><1hr</td><td colspan="2">1-2hr</td><td colspan="2">>2hr</td></tr>
<tr><td colspan="3">3</td><td colspan="2">1</td><td colspan="2">1</td></tr>
<tr><td rowspan="3">[57]</td><td rowspan="3">Engineered test</td><td colspan="7">No. Events</td></tr>
<tr><td colspan="4"><1hr</td><td colspan="3">>12hr</td></tr>
<tr><td colspan="4">7</td><td colspan="3">2</td></tr>
</table>

TABLE 8. Accuracy of detection by different methodologies.

<table>
<tr><th>Authors</th><th>Case Study</th><th colspan="5">Accuracy of Event Detection (TPR)</th><th>Accuracy of Event Localization</th></tr>
<tr><td rowspan="4">[68]</td><td rowspan="4">Simulated Data</td><td>Proposed Methods</td><td colspan="2">Consistent Condition</td><td colspan="2">Varying Condition</td><td rowspan="4">N.A</td></tr>
<tr><td>NKF + Hotelling T^2 (1)</td><td colspan="2">32%</td><td colspan="2">29%</td></tr>
<tr><td>NFK + CUSUM (2)</td><td colspan="2">87%</td><td colspan="2">87%</td></tr>
<tr><td>(1)+(2)</td><td colspan="2">87%</td><td colspan="2">87%</td></tr>
<tr><td>[69]</td><td>Simulated Data</td><td colspan="5">N.A</td><td>1.86cm–2.19cm</td></tr>
<tr><td rowspan="2">[57]</td><td>Engineered Test</td><td colspan="5">80% (Based on ROC curve)</td><td rowspan="2">N.A</td></tr>
<tr><td>Historical Data</td><td colspan="5">76% (29 out of 38 alarms for real data)</td></tr>
<tr><td>[27]</td><td>Engineered Test</td><td colspan="5">98.2%</td><td><3m-58m</td></tr>
<tr><td>[18]</td><td>Historical Data</td><td colspan="5">48.3%</td><td>N.A</td></tr>
<tr><td>[75]</td><td>Engineered Test</td><td colspan="5">71.43%</td><td>N.A</td></tr>
<tr><td>[17]</td><td>Engineered Test</td><td colspan="5">71.43%</td><td>N.A</td></tr>
<tr><td rowspan="2">[77]</td><td>Simulated Data</td><td colspan="5">99.5%</td><td rowspan="2">N.A</td></tr>
<tr><td>Historical Data</td><td colspan="5">98.75%</td></tr>
<tr><td>[38]</td><td>Engineered Test</td><td colspan="5">N.A</td><td>~300m</td></tr>
<tr><td rowspan="4">[79]</td><td rowspan="4">Historical Data</td><td rowspan="2">20 Days Window</td><td>DMA1</td><td>DMA 2</td><td>DMA 3</td><td>DMA 4</td><td rowspan="4">N.A</td></tr>
<tr><td>75%-96%</td><td>58%-90%</td><td>83%-95%</td><td>62%-93%</td></tr>
<tr><td rowspan="2">90 Days Window</td><td>DMA 1</td><td>DMA 2</td><td>DMA 3</td><td>DMA 4</td></tr>
<tr><td>85%-98%</td><td>67%-88%</td><td>67%-78%</td><td>87%-92%</td></tr>
<tr><td>[4]</td><td>Engineered Test</td><td colspan="5">100%</td><td>19.65m-45.72m</td></tr>
<tr><td>[50]</td><td>Engineered Test</td><td colspan="5">100%</td><td>2.72m-61.82m</td></tr>
</table>

Missing data can be estimated using mathematical and statistical methods. For example, the arithmetic mean, inverse distance weighting, regression-based analysis methods, kriging estimation method, and gamma distribution function [102]. One can also delete all records with a missing value [103]. However, by removing the entire record, there lies the possibility of removing useful training data.

The noisy data can be handled in three different ways [104]. They are namely: 1) robust algorithm insensitive to noise; 2) removing noise using filtering methods or using outlier detection techniques [105], and 3) correcting noisy instances. In the detection of sensors' fault, there are mainly three approaches. They are namely: knowledge-based, measurement aberration detection and model-based. The knowledge-based approach relies on qualitative model handling using heuristic reasoning. The measurement aberration detection method examines the output of a single sensor for indications of faults. Model-based fault detection relies on analytical redundancy in the form of dedicated observers [106].

However, to achieve an accurate classification or predictive model, a large amount of data is unavoidable [18]. Moving towards Industry 4.0 or an Industrial Internet of Things era, the amount of data may not be a concern in the near future.

TABLE 9. Rate of false alarm by different methodologies.

<table>
<tr><th>Authors</th><th>Case Study</th><th colspan="5">False Alarm (FPR)</th></tr>
<tr><td rowspan="4">[68]</td><td rowspan="4">Simulated Data</td><td>Proposed Methods</td><td colspan="2">Consistent Condition (Per Day)</td><td colspan="2">Varying Condition (Per Day)</td></tr>
<tr><td>NKF + Hotelling T^2 (1)</td><td colspan="2">0.01</td><td colspan="2">0.015</td></tr>
<tr><td>NFK + CUSUM (2)</td><td colspan="2">0</td><td colspan="2">0.005</td></tr>
<tr><td>(1)+(2)</td><td colspan="2">0.01</td><td colspan="2">0.015</td></tr>
<tr><td>[71]</td><td>Engineered Test</td><td colspan="5">4%</td></tr>
<tr><td rowspan="2">[57]</td><td>Engineered Test</td><td colspan="5">10%</td></tr>
<tr><td>Historical Data</td><td colspan="5">8%</td></tr>
<tr><td>[27]</td><td>Engineered Test</td><td colspan="5">0.2%</td></tr>
<tr><td>[75]</td><td>Engineered Test</td><td colspan="5">0.76%</td></tr>
<tr><td>[17]</td><td>Engineered Test</td><td colspan="5">0.4%</td></tr>
<tr><td rowspan="4">[79]</td><td rowspan="4">Historical Data</td><td rowspan="2">20 Days Window</td><td>DMA 1</td><td>DMA 2</td><td>DMA 3</td><td>DMA 4</td></tr>
<tr><td>10%-13%</td><td>7%-24%</td><td>13%-22%</td><td>10%-19%</td></tr>
<tr><td rowspan="2">90 Days Window</td><td>DMA 1</td><td>DMA 2</td><td>DMA 3</td><td>DMA 4</td></tr>
<tr><td>9%-15%</td><td>6%-18%</td><td>8%-18%</td><td>6%-8%</td></tr>
<tr><td>[4]</td><td>Engineered Test</td><td colspan="5">0%</td></tr>
<tr><td>[50]</td><td>Engineered Test</td><td colspan="5">10%</td></tr>
</table>

Therefore, data-driven methods can be seen as the optimal solution in the near future.

Total water demand can be explained as a sum of water demand from different aspects [107] given as

$$D^T = D^{Con} + D^{Vol} + D^{Unc} + D^L \tag{22}$$

where D^T is the total water demand. D^{Con} is consumers' water demand from controlled orifices (opening degree is controlled by customers depending on a maximum required demand over time and pressures in system, example showers, washbasin). Here, D^{Vol} is the demand of volume controlled orifices (orifices on/off status depends on water storage filling levels, example bathtub, industrial process tank), D^{Unc} is the demand of fire hydrants, sprinklers (maximum opening degree is set during an event to have the demand allowed by pressures in the system) and D^L is the demand of uncontrolled orifices attribute to leaks [108].

Assuming unexpected water demand comes only from industrial process plant when they need a large amount of water due to a large order from a customer and during an emergency such as firefighting. Then the unexpected water demand can happen in D^{Vol} and D^{Unc} . If a city is partitioned into an industrial estate and housing estate then flow meters installed in the industrial estate can help to identify unexpected water demand from a process plant. However, this method will require a lot more flow meters to be installed along pipes that can be too expensive to identify a false alarm due to unexpected water demand. To differentiate between a firefighting incidence and a burst from a fire hydrant, one can establish a time window in which the maximum flow is allowed. However, if a burst occurs, it will only raise the alarm after the time window has passed which may not be efficient. These are the complications from just considering two unexpected water demand. Although researchers [22] mentioned that a low FPR might be more important than low detectable burst size, a false alarm is still better than a missed detection. As a result, it may be negligible to include unexpected water demand into the equation when developing an algorithm when the rate of unexpected water demand is low.

Background leaks are defined as losses below around 500 liters per hour [14], [121]. Thus, the cause of limitation #4 may be due to the fact that proposed methodologies are not reliable enough to detect small leaks.

If one assumes a uniform distribution of leakage along a pipe, then background leak can be model as [16], [109]–[112]

$$L_B = \begin{cases} B_k l_k (P_k)^{ak}, & P_k > 0 \\ 0, & otherwise \end{cases} \tag{23}$$

where L_B is the background leak. B_k is the pipe parameter which is dependent on pipe characteristics and influenced by external factors.

Giustolisi *et al.* [111] tested the range of B_k from 1×10^{-8} to 2×10^{-9} and 1×10^{-6} to 2×10^{-7} for two different networks. In another study, Giustolisi *et al.* [112] adopted a value from 1×10^{-8} to 4×10^{-8} for four different DMA. Laucelli *et al.* [113] adopted B_k of 1.073×10^{-7} for a smaller network in Italy, B_k of 7.285×10^{-9} for newer part of a bigger network in Italy and $\beta_k$of 1.457×10^{-8} for the older part of the same network. Adedeji *et al.* [16] assumed the value of 2×10^{-8} for in their study. Here, l_k is the length of pipe, P_k is the average pressure in the pipe computed as the mean of the pressure values at the end nodes of the kth pipe. a_k are the leakage model parameter which ranges from 0.5 to 2.5 used by different authors [114]–[117], but Lambert [115] and May [116] suggested that a_k should be around 1.5 for background type leakage. If the background leak is modeled as a leakage model, the user will face the uncertainty of B_k. Moreover, small leaks can be easily masked by sensors' noise. Thus, planning medium-long term interventions for asset rehabilitation and pressure management can be an alternative to reduce background leakages [123].

To address Limitation #5, sensors should be installed at locations that maximize localization accuracy [118]–[120].

The performance of leak localization is highly dependent on sensor number and placement within a DMA [41]. Thus, sensors' placement becomes an optimization problem to be solved at the sensor placement stage where there is a trade-off between a number of sensors and subsequent cost [41]. Limitation #6 can worsen since localization accuracy is highly dependent on the sensor number and placement within the DMA. With an optimized minimum number of sensors, users can identify a smaller possible leak zone instead of the exact leak location.

VII. CONCLUSION

Water is a precious resource that needs to conserve in the ecosystem. Any wastage through pipe leakage should be minimized. In this paper, different technologies for water pipe leakage were discussed. Proposed methodologies over the past five years with their limitations were also reviewed and discussed in details. The uncertainty involved and suggestions were also provided. The main issue in the detection of water leakage is the stochastic nature of water demand that is influenced by many factors such as seasonal change and consumers' habits. The limited number of sensors can further hamper the implementations of detection algorithms. Although results by several different authors display their prowess in leakage detection, achieving a high leakage localization accuracy remains a challenging task.

REFERENCES

[1] A. Lay-Ekuakille, G. Vendramin, and A. Trotta, "Robust spectral leak detection of complex pipelines using filter diagonalization method," *IEEE Sensors J.*, vol. 9, no. 11, pp. 1605–1614, Nov. 2009.

[2] S. K. Mandal, F. T. S. Chan, and M. K. Tiwari, "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3071–3080, Feb. 2012.

[3] R. Kiliç, "Effective management of leakage in drinking water network," *Acta Phys. Polonica, A*, vol. 130, no. 1, pp. 479–483, Jul. 2016.

[4] T. T. T. Zan, H. B. Lim, K.-J. Wong, A. J. Whittle, and B.-S. Lee, "Event detection and localization in urban water distribution network," *IEEE Sensors J.*, vol. 14, no. 12, pp. 4134–4142, Dec. 2014.

[5] M. Farley and S. Trow, *Losses in Water Distribution Networks: A Practitioners' Guide to Assessment, Monitoring and Control*. London, U.K.: IWA Publishing, 2003.

[6] L. Ferrandez-Gamot *et al.*, "Leak localization in water distribution networks using pressure residuals and classifiers," *IFAC-PapersOnline*, vol. 48, no. 21, pp. 220–225, Sep. 2015.

[7] W. Li *et al.*, "Development of systems for detection, early warning, and control of pipeline leakage in drinking water distribution: A case study," *J. Environ. Sci.*, vol. 23, no. 11, pp. 1816–1822, Feb. 2011.

[8] J. Mashford, D. De Silva, S. Burn, and D. Marney, "Leak detection in simulated water pipe networks using SVM," *Appl. Artif. Intell.*, vol. 26, no. 5, pp. 429–444, May 2012.

[9] I. Okeya, C. Hutton, and Z. Kapelan, "Locating pipe bursts in a district metered area via online hydraulic modelling," in *Proc. 13th Comput. Control Water Ind. Conf. (CCWI)*, Leicester, U.K., Sep. 2015, pp. 101–110, 2015.

[10] D. Sala and P. Kołakowski, "Detection of leaks in a small-scale water distribution network based on pressure data—Experimental verification," in *Proc. 12th Int. Conf. Comput. Control Water Ind. (CCWI)*, Perugia, Italy, Sep. 2013, pp. 1460–1469.

[11] J. Thornton, R. Sturm, and G. Kunkel, *Water Loss Control*. 2nd ed. New York, NY, USA: McGraw-Hill, 2008.

[12] K. B. Adedeji, Y. Hamam, B. T. Abe, and A. M. Abu-Mahfouz, "Towards achieving a reliable leakage detection and localization algorithm for application in water piping networks: An overview," *IEEE Access*, vol. 5, pp. 20272–20285, Sep. 2017.

[13] Y. Ishido and S. Takahashi, "A new indicator for real-time leak detection in water distribution networks: Design and simulation validation," in *Proc. 16th Conf. Water Distrib. Syst. Anal. (WDSA)*, Bari, Italy, Jul. 2014, pp. 411–417.

[14] I. Okeya, Z. Kapelan, C. Hutton, and D. Naga, "Online burst detection in a water distribution system using the Kalman filter and hydraulic modelling," in *Proc. 16th Conf. Water Distrib. Syst. Anal. (WDSA)*, Bari, Italy, Jul. 2014, pp. 418–427.

[15] A. O. Lambert, "Accounting for losses: The bursts and background concept," *Water Environ. J.*, vol. 8, no. 2, pp. 205–214, 1994.

[16] K. B. Adedeji, Y. Hamam, B. T. Abe, and A. M. Abu-Mahfouz, "Leakage detection and estimation algorithm for loss reduction in water piping networks," *Water*, vol. 9, no. 10, p. 773, Oct. 2017.

[17] Y. Wu, S. Liu, K. Smith, and X. Wang, "Using correlation between data from multiple monitoring sensors to detect bursts in water distribution systems," *J. Water Resour. Planning Manage.*, vol. 144, no. 2, p. 04017084, Feb. 2018.

[18] T. Tao, H. Huang, F. Li, and K. Xin, "Burst detection using an artificial immune network in water-distribution systems," *J. Water Resour. Planning Manage.*, vol. 140, no. 10, p. 04014027, Oct. 2014.

[19] M. Bakker, E. A. Trietsch, J. H. G. Vreeburg, and L. C. Rietveld, "Analysis of historic bursts and burst detection in water supply areas of different size," *Water Sci. Technol. Water Supply*, vol. 14, no. 6, pp. 1035–1044, Dec. 2014.

[20] J. M. Makar and Y. Kleiner, "Maintaining water pipeline integrity," in *Proc. AWWA Infrastruct. Conf. Exhibit.*, Baltimore, MD, USA, Mar. 2000, pp. 1–13.

[21] R. Li, H. Huang, K. Xin, and T. Tao, "A review of methods for burst/leakage detection and location in water distribution systems," *Water Sci. Technol., Water Supply*, vol. 15, no. 3, pp. 429–441, Jun. 2015.

[22] Y. Wu and S. Liu, "A review of data-driven approaches for burst detection in water distribution systems," *Urban Water J.*, vol. 14, no. 9, pp. 972–983, Feb. 2017.

[23] S. Datta and S. Sarkar, "A review on different pipeline fault detection methods," *J. Loss Prevention Process Ind.*, vol. 41, pp. 97–106, May 2016.

[24] A. F. Colombo, P. Lee, and B. W. Karney, "A selective literature review of transient-based leak detection methods," *J. Hydro-Environ. Res.*, vol. 2, no. 4, pp. 212–227, 2009.

[25] Z. Liu and Y. Kleiner, "State of the art review of inspection technologies for condition assessment of water pipes," *Measurement*, vol. 46, no. 1, pp. 1–15, Jan. 2013.

[26] A. M. Obeid, F. Karray, M. W. Jmal, M. Abid, S. M. Qasim, and M. S. BenSaleh, "Towards realisation of wireless sensor network-based water pipeline monitoring systems: A comprehensive review of techniques and platforms," *IET Sci. Meas. Technol.*, vol. 10, no. 5, pp. 420–426, Aug. 2016.

[27] J. Kang, Y.-J. Park, J. Lee, S.-H. Wang, and D.-S. Eom, "Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4279–4289, May 2018.

[28] M. R. Halfawy and J. Hengmeechai, "Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine," *Automat. Construction*, vol. 38, pp. 1–13, Mar. 2014.

[29] D. Ozevin and H. Yalcinkaya, "Reliable monitoring of leak in gas pipelines using acoustic emission method," in *Proc. Civil Struct. Health Monit. Workshop (CSHM)*, Berlin, Germany, Nov. 2012, pp. 1–8.

[30] M. Ahadi and M. S. Bakhtiar, "Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to acoustic emission signals," *Appl. Acoust.*, vol. 71, no. 7, pp. 634–639, Jul. 2010.

[31] S. Demirci, E. Yigit, I. H. Eskidemir, and C. Ozdemir, "Ground penetrating radar imaging of water leaks from buried pipes based on back-projection method," *NDT&E Int.*, vol. 47, pp. 35–42, Apr. 2012.

[32] G. Geiger, "State-of-the-art in leak detection and localization," in *Proc. Pipeline Tech. Conf.*, Hannover, Germany, Dec. 2006, pp. 1–25.

[33] H. Kaplan, *Practical Applications of Infrared Thermal Sensing and Imaging Equipment*. 3rd ed. Washington, DC, USA: SPIE, 2007.

[34] B. Shakmak and A. Al-Habaibeh, "Detection of water leakage in buried pipes using infrared technology; A comparative study of using high and low resolution infrared cameras for evaluating distant remote detection," in *Proc. IEEE Jordan Conf. Appl. Elect. Eng. Comput. Technol. (AEECT)*, Amman, Jordan, Nov. 2015, pp. 1–7.

[35] P. M. Bach and J. K. Kodikara, "Reliability of infrared thermography in detecting leaks in buried water reticulation pipes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4210–4224, Sep. 2017.

[36] E. Galdiero, F. De Paola, N. Fontana, M. Giugni, and D. Savić, "Decision support system for the optimal design of district metered areas," *J. Hydroinform.*, vol. 18, no. 1, pp. 49–61, Jan. 2015.

[37] G. Ferrari, D. Savić, and G. Becciu, "Graph-theoretic approach and sound engineering principles for design of district metered areas," *J. Water Resour. Planning Manage.*, vol. 140, no. 12, p. 04014036, Dec. 2014.

[38] R. Perez *et al.*, "Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona," *IEEE Control Syst.*, vol. 34, no. 4, pp. 24–36, Aug. 2014.

[39] D. B. Laucelli, A. Simone, L. Berardi, and O. Giustolisi, "Optimal design of district metering areas for the reduction of leakages," *J. Water Resour. Planning Manage.*, vol. 143, no. 6, p. 04017017, Jun. 2017.

[40] A. Soldevila, J. Blesa, S. Tornil-Sin, E. Duviella, R. M. Fernandez-Canti, and V. Puig, "Leak localization in water distribution networks using a mixed model-based/data-driven approach," *Control Eng. Pract.*, vol. 55, pp. 162–173, Oct. 2016.

[41] M. À. Cugueró-Escofet, V. Puig, and J. Quevedo, "Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: Application to the Barcelona water network," *Control Eng. Pract.*, vol. 63, pp. 1–12, Jun. 2017.

[42] R. Puust, Z. Kapelan, D. Savić, and T. Koppel, "A review of methods for leakage management in pipe networks," *Urban Water J.*, vol. 7, no. 1, pp. 25–45, Feb. 2010.

[43] T. Jitong and C. Jothityangkoon, "Reducing water loss in a water supply system using a district metering area (DMA): A case study of the Provincial Waterworks Authority (PWA), Lop Buri Branch," *Eng. Appl. Sci. Res.*, vol. 44, no. 3, pp. 154–160, Sep. 2017.

[44] J. Gao, B. Shi, W. Zhang, and H. Zhu, "Monitoring the stress of the post-tensioning cable using fiber optic distributed strain sensor," *Measurement*, vol. 39, no. 5, pp. 420–428, Jun. 2006.

[45] L. Ren, T. Jiang, D.-S. Li, P. Zhang, H.-N. Li, and G.-B. Song, "A method of pipeline corrosion detection based on hoop-strain monitoring technology," *Struct. Control Health Monit.*, vol. 24, no. 6, p. e1931, Aug. 2016.

[46] L. Boaz, S. Kaijage, and R. Sinde, "An overview of pipeline leak detection and location systems," in *Proc. Pan Afr. Int. Conf. Inf. Sci., Comput. Telecommun.*, Arusha, Tanzania, Jul. 2014, pp. 133–137.

[47] S. S. C. Madabhushi, M. Z. E. B. Elshafie, and S. K. Haigh, "Accuracy of distributed optical fiber temperature sensing for use in leak detection of subsea pipelines," *J. Pipeline Syst. Eng. Pract.*, vol. 6, no. 2, p. 4014014, May 2015.

[48] D. L. Atherton, "Remote field eddy current inspection," *IEEE Trans. Magn.*, vol. 31, no. 6, pp. 4142–4147, Nov. 1995.

[49] S. Duchesne, N. Bouzida, and J.-P. Villeneuve, "Performance estimation of a remote field eddy current method for the inspection of water distribution pipes," *J. Water Resour. Planning Manage.*, vol. 137, no. 6, pp. 521–530, Nov. 2011.

[50] S. Srirangarajan, M. Allen, A. Preis, M. Iqbal, H. B. Lim, and A. J. Whittle, "Wavelet-based burst event detection and localization in water distribution systems," *J. Signal Process. Syst.*, vol. 72, no. 1, pp. 1–16, Jul. 2013.

[51] S. Grosswig, E. Hurtig, S. Luebbecke, and B. Vogel, "Pipeline leakage detection using distributed fibre optical temperature sensing," in *Proc. 17th Int. Conf. Opt. Fibre Sensors*, Bruges, Belgium, May 2005, pp. 1–4.

[52] D. Chatzigeorgiou, K. Youcef-Toumi, and R. Ben-Mansour, "MIT leak detector: Modeling and analysis toward leak-observability," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 5, pp. 2391–2402, Oct. 2015.

[53] J. A. Liggett and L.-C. Chen, "Inverse transient analysis in pipe networks," *J. Hydraulic Eng.*, vol. 120, no. 8, pp. 934–955, 1994.

[54] A. Soldevila, R. M. Fernandez-Canti, J. Blesa, S. Tornil-Sina, and V. Puig, "Leak localization in water distribution networks using Bayesian classifiers," *J. Proc. Control*, vol. 55, pp. 1–9, Jul. 2017.

[55] S. J. Lee, G. Lee, J. C. Suh, and J. M. Lee, "Online burst detection and location of water distribution systems and its practical applications," *J. Water Resour. Planning Manage.*, vol. 142, no. 1, p. 04015033, Jan. 2016.

[56] H. Zhang, Y. Liang, W. Zhang, N. Xu, Z. Guo, and G. Wu, "Improved PSO-based method for leak detection and localization in liquid pipelines," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3143–3154, Jul. 2018.

[57] M. Romano, Z. Kapelan, and D. Savić, "Automated detection of pipe bursts and other events in water distribution systems," *J. Water Resour. Planning Manage.*, vol. 140, no. 4, pp. 457–467, Apr. 2014.

[58] S. Adachi, S. Takahashi, and H. Tadokoro, "Estimating area leakage in water networks based on hydraulic model and asset information," in *Proc. 16th Conf. Water Distrib. Syst. Anal. (WDSA)*, Bari, Italy, Jul. 2014, pp. 278–285.

[59] J. Meseguer, J. M. Mirats-Tur, G. C. Gabriela, and V. Puig, "Model-based monitoring techniques for leakage localization in distribution water networks," in *Proc. 13th Comput. Control Water Indust. Conf. (CCWI)*, Leicester, U.K., Sep. 2015, pp. 1399–1408.

[60] J. F. Douglas, J. Gasiorek, and J. Swaffield, *Fluid Mechanics*. New York, NY, USA: Longman Scientific and Technical, 1995.

[61] J. Thornton and A. Lambert, "Progress in practical prediction of pressure: Leakage, pressure: Burst frequency and pressure: Consumption relationships," in *Proc. IWA Special Conf. Leakage*, Halifax, NS, Canada, Sep. 2005, pp. 12–14.

[62] D. Chatzigeorgiou, K. Youcef-Toumi, and R. Ben-Mansour, "Design of a novel in-pipe reliable leak detector," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 2, pp. 824–833, Apr. 2015.

[63] M. À. CuguerÃş-Escofet *et al.*, "Model-vs. data-based approaches applied to fault diagnosis in potable water supply networks," *J. Hydroinformatics*, vol. 18, no. 5, pp. 831–850, May 2016.

[64] S. Seifollahi-Aghmiuni, O. B. Haddad, M. H. Omid, and M. A. Mariño, "Effects of pipe roughness uncertainty on water distribution network performance during its operational period," *Water Resour. Manage.*, vol. 27, no. 5, pp. 1581–1599, Mar. 2013.

[65] S. R. Mounce, R. B. Mounce, and J. B. Boxall, "Novelty detection for time series data analysis in water distribution systems using support vector machines," *J. Hydroinformatics*, vol. 13, no. 4, pp. 672–686, Nov. 2010.

[66] G. Mazzolani, L. Berardi, D. Laucelli, A. Simone, R. Martino, and O. Giustolisi, "Estimating leakages in water distribution networks based only on inlet flow data," *J. Water Resour. Planning Manage.*, vol. 143, no. 6, p. 04017014, Jun. 2017.

[67] S. R. Mounce, J. B. Boxall, and J. Machell, "Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows," *J. Water Resour. Planning Manage.*, vol. 136, no. 3, pp. 309–318, May 2010.

[68] D. Jung and K. Lansey, "Water distribution system burst detection using a nonlinear Kalman filter," *J. Water Resour. Planning Manage.*, vol. 141, no. 5, p. 04014070, May 2015.

[69] F. Karray, A. Garcia-Ortiz, M. W. Jmal, A. M. Obeid, and M. Abid, "*EARNPIPE*: A testbed for smart water pipeline monitoring using wireless sensor network," in *Proc. 20th Int. Conf. Knowl. Based Intell. Inf. Eng. Sys.*, York, U.K., vol. 96, Sep. 2016, pp. 285–294.

[70] D. Laucelli, M. Romano, D. Savić, and O. Giustolisi, "Detecting anomalies in water distribution networks using EPR modelling paradigm," *J. Hydroinform.*, vol. 18, no. 3, pp. 409–427, May 2016.

[71] G. Ye and A. Fenner, "Weighted least squares with expectation-maximization algorithm for burst detection in U.K. water distribution systems," *J. Water Resour. Planning Manage.*, vol. 140, no. 4, pp. 417–424, Apr. 2014.

[72] S. A. McKenna, F. Fusco, and B. J. Eck, "Water demand pattern classification from smart meter data," in *Proc. 12th Int. Conf. Comput. Control Water Ind. (CCWI)*, Perugia, Italy, Sep. 2013, pp. 1121–1130.

[73] S. R. Mounce, R. B. Mounce, T. Jackson, J. Austin, and J. B. Boxall, "Pattern matching and associative artificial neural networks for water distribution system time series data analysis," *J. Hydroinformatics*, vol. 16, no. 3, pp. 617–632, May 2014.

[74] K. Aksela, M. Aksela, and R. Vahala, "Leakage detection in a real distribution network using a SOM," *Urban Water J.*, vol. 6, no. 4, pp. 279–289, Aug. 2009.

[75] Y. Wu, S. Liu, X. Wu, Y. Liu, and Y. Guan, "Burst detection in district metering areas using a data driven clustering algorithm," *Water Res.*, vol. 100, pp. 28–37, Sep. 2016.

[76] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[77] Q. Zhang, Z. Y. Wu, M. Zhao, J. Qi, Y. Huang, and H. Zhao, "Leakage zone identification in large-scale water distribution systems using multiclass support vector machines," *J. Water Resour. Planning Manage.*, vol. 142, no. 11, p. 04016042, Nov. 2016.
[78] A. Rajeswaran, S. Narasimhan, and S. Narasimhan, "A graph partitioning algorithm for leak detection in water distribution networks," *Comput. Chem. Eng.*, vol. 108, pp. 11–23, Jan. 2018.
[79] D. Loureiro, C. Amado, A. Martins, D. Vitorino, A. Mamade, and S. T. Coelho, "Water distribution systems flow monitoring and anomalous event detection: A practical approach," *Urban Water J.*, vol. 13, no. 3, pp. 242–252, 2016.
[80] Y. Kim, S. J. Lee, T. Park, G. Lee, J. C. Suh, and J. MinLee, "Robust leak detection and its localization using interval estimation for water distribution network," *Comput. Chem. Eng.*, vol. 92, pp. 1–17, Sep. 2016.
[81] S. E. Christodoulou, E. Kourti, and A. Agathokleous, "Waterloss detection in water distribution networks using wavelet change-point detection," *Water Resour. Manage.*, vol. 31, no. 3, pp. 979–994, Feb. 2017.
[82] S. E. Christodoulou, E. Kourti, and A. Agathokleous, "Waterloss detection in streaming water flow timeseries using change-point anomaly methods," *Eur. Waters*, vol. 58, pp. 429–434, Apr. 2017.
[83] A. Korotcov, V. Tkachenko, D. P. Russo, and S. Ekins, "Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets," *Mol. Pharmaceutics*, vol. 14, no. 12, pp. 4462–4475, Dec. 2017.
[84] I. K. Timotius and S.-G. Miaou, "Arithmetic means of accuracies: A classifier performance measurement for imbalanced data set," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Shanghai, China, Nov. 2010, pp. 1244–1251.
[85] A. A. Mitani, P. E. Freer, and K. P. Nelson, "Summary measures of agreement and association between many raters' ordinal classifications," *Ann. Epidemiol.*, vol. 27, no. 10, pp. 677–685, Oct. 2017.
[86] J. Singh, R. K. Sharma, and A. K. Gupta, "A method of REM-NREM sleep distinction using ECG signal for unobtrusive personal monitoring," *Comput. Biol. Med.*, vol. 78, pp. 138–143, Nov. 2016.
[87] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochem. Med.*, vol. 22, no. 3, pp. 276–282, Oct. 2012.
[88] L. Guan, S. Zhang, and H. Xu, "BAMORF: A novel computational method for predicting the extracellular matrix proteins," *IEEE Access*, vol. 5, pp. 18498–18505, Sep. 2017.
[89] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.
[90] A. Ben-David, "About the relationship between ROC curves and Cohen's kappa," *Eng. Appl. Artif. Intell.*, vol. 21, no. 6, pp. 874–882, Sep. 2008.
[91] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.
[92] M. Quiñones-Grueiro, C. Verde, and O. Llanes-Santiago, "Demand model in water distribution networks for fault detection," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3263–3268, Jul. 2017.
[93] G. Douzas and F. Bacao, "Self-organizing map oversampling (SOMO) for imbalanced data set learning," *Expert Syst. Appl.*, vol. 82, pp. 40–52, Oct. 2017.
[94] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[95] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE—Majority weighted minority oversampling technique for imbalanced data set learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 405–425, Feb. 2014.
[96] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[97] T. K. Chan and S. C. Chin, "Health stages diagnostics of underwater thruster using sound features with imbalanced dataset," *Neural Comput. Appl.*, to be published, doi: 10.1007/s00521-018-3407-3.
[98] V. Chatzigiannakis and S. Papavassiliou, "Diagnosing anomalies and identifying faulty nodes in sensor networks," *IEEE Sensors J.*, vol. 7, no. 5, pp. 637–645, May 2007.
[99] M. El-Koujok, M. Benammar, N. Meskin, M. Al-Naemi, and R. Langari, "Multiple sensor fault diagnosis by evolving data-driven approach," *Inf. Sci.*, vol. 259, pp. 346–358, Feb. 2014.
[100] K. Xie *et al.*, "Recover corrupted data in sensor networks: A matrix completion solution," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1434–1448, May 2017.
[101] L. Gruenwald, H. Chok, and M. Aboukhamis, "Using data mining to estimate missing sensor data," in *Proc. 7th IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Omaha, NE, USA, Oct. 2007, pp. 207–212.
[102] J. Kim and J. H. Ryu, "Quantifying a threshold of missing values for gap filling processes in daily precipitation series," *Water Res. Manage.*, vol. 29, no. 11, pp. 4173–4184, Sep. 2015.
[103] N. Yan, M.-Z. Zhou, and L. Tong, "An estimation algorithm for missing data in wireless sensor networks," *Int. J. Smart Sens. Intell. Syst.*, vol. 6, no. 3, pp. 1032–1053, Jun. 2013.
[104] W. Yuan, D. Guan, L. Shen, and H. Pan, "An empirical study of filter-based feature selection algorithms using noisy training data," in *Proc. 4th IEEE Int. Conf. Inf. Sci. Technol.*, Shenzhen, China, Apr. 2014, pp. 209–212.
[105] H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 3, pp. 304–319, Mar. 2006.
[106] M. M. Seron, J. A. De Dona, and S. Olaru, "Fault tolerant control allowing sensor healthy-to-faulty and faulty-to-healthy transitions," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1657–1669, Jul. 2012.
[107] S. G. Buchberger and G. J. Wells, "Intensity, duration, and frequency of residential water demands," *J. Water Resour. Planning Manage.*, vol. 122, no. 1, pp. 11–19, Jan. 1996.
[108] O. Giustolisi and T. M. Walski, "Demand components in water distribution network analysis," *J. Water Resour. Planning Manage.*, vol. 138, no. 4, pp. 356–367, Jul. 2012.
[109] G. Germanopoulos, "A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models," *Civil Eng. Syst.*, vol. 2, no. 3, pp. 171–179, 1985.
[110] G. Germanopoulos and P. W. Jowitt, "Leakage reduction by excess pressure minimization in a water supply network," *Proc. Inst. Civil Eng.*, vol. 87, no. 2, pp. 195–214, Jun. 1989.
[111] O. Giustolisi, D. Savić, and Z. Kapelan, "Pressure-driven demand and leakage simulation for water distribution networks," *J. Hydraulic Eng.*, vol. 134, no. 5, pp. 626–635, May 2008.
[112] O. Giustolisi, L. Berardi, D. Laucelli, D. Savić, T. Walski, and B. Brunone, "Battle of background leakage assessment for water networks (BBLAWN) at WDSA conference 2014," in *Proc. 16th Conf. Water Distrib. Syst. Anal. (WDSA)*, Bari, Italy, Jul. 2014, pp. 4–12.
[113] D. Laucelli, L. Berardi, and O. Giustolisi, "Assessing climate change and asset deterioration impacts on water distribution networks: Demand-driven or pressure-driven network modeling?" *Environ. Model. Softw.*, vol. 37, pp. 206–216, Nov. 2012.
[114] P. W. Jowitt and C. Xu, "Optimal valve control in water-distribution networks," *J. Water Resour. Planning Manage.*, vol. 116, no. 4, pp. 455–472, Jul. 1990.
[115] A. O. Lambert, "What do we know about pressure: Leakage relationships in distribution systems?" in *Proc. IWA Conf. Syst. Approach Leakage Control Water Distrib. Syst. Manage.*, Brno, Czech Republic, May 2001, pp. 89–96.
[116] J. May, "Pressure dependent leakage," World Water Environ. Eng., Water Environ. Fed., Washington, DC, USA, Tech. Rep., Oct. 1994. [Online]. Available: http://www.leakssuite.com/wp-content/uploads/2016/10/JOHN-MAY-SEMINAL-1994-ARTICLE-4.pdf
[117] K. Vairavamoorthy and J. Lumbers, "Leakage reduction in water distribution systems: Optimal valve control," *J. Hydraulic Eng.*, vol. 124, no. 11, pp. 1146–1154, Nov. 1998.
[118] M. V. Casillas, V. Puig, L. E. Garza-Castañón, and A. Rosich, "Optimal sensor placement for leak location in water distribution networks using genetic algorithms," in *Proc. Conf. Control Fault-Tolerant Syst. (SysTol)*, Nice, France, Oct. 2013, pp. 14984–15005.
[119] M. V. Casillas, L. E. Garza-Castañón, and V. Puig, "Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms," *Water*, vol. 7, no. 11, pp. 6496–6515, Nov. 2015.
[120] F. Nejjari, R. Sarrate, and J. Blesa, "Optimal pressure sensor placement in water distribution networks minimizing leak location uncertainty," *Procedia Eng.*, vol. 119, pp. 953–962, Sep. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705815026491?via%3Dihub
[121] L. M. Brion and L. Mays, "Methodology for optimal operation of pumping stations in water distribution systems," *J. Hydraulic Eng.*, vol. 117, no. 11, pp. 1551–1569, Nov. 1991.
[122] T. Al-Washali, S. Sharma, and M. Kennedy, "Methods of assessment of water losses in water supply systems: A review," *Water Res. Manage.*, vol. 30, no. 14, pp. 4985–5001, Nov. 2016.
[123] D. Laucelli and S. Meniconi, "Water distribution network analysis accounting for different background leakage models," in *Proc. 13th Comput. Control Water Ind. Conf. (CCWI)*, Leicester, U.K., Sep. 2015, pp. 680–689.

T. K. CHAN received the B.Eng. and M.Phil. degrees from Newcastle University, Newcastle upon Tyne, in 2015 and 2018. He is currently pursuing the Ph.D. degree under the Singapore Industrial Postgraduate Programme with Visenti Pte. Ltd. and Newcastle University. His role as a Data Engineer at Visenti Pte. Ltd. includes investigation of advanced information fusion methods to coordinate the deployed large scale heterogeneous sensor network and development of models to achieve a reliable and robust leak detection and localization performance. His current research interests include water leakage detection and localization algorithms, signal processing, machine learning, and handling of imbalanced dataset.

CHENG SIONG CHIN (M'01–SM'09) received the B.Eng. degree (Hons.) in mechanical and aerospace engineering from Nanyang Technological University (NTU), in 2000, the M.Sc. degree (Hons.) in advanced control and systems engineering from The University of Manchester, in 2001, and the Ph.D. degree in applied control engineering from the Research Robotics Centre, NTU, in 2008. He is currently an Adjunct Professor with the School of Automotive Engineering, Chongqing University, and the Director of Innovation at Newcastle University, Singapore. He was a Lecturer in mechatronics engineering with Temasek Polytechnic, in 2008. He worked in the industry for a few years before moving into academia. He has published over 100 journal papers, books, book chapters, and conference papers. He currently holds three U.S. patents, two provisional U.S. patent applications, one Singapore provisional patent, and two trade secrets in electronics and measurement systems. He obtained two research grants from SMI and four EDB-Industrial Postgraduate Programme grants in the areas of intelligent systems design and predictive analytics. His current research interests include intelligent systems design and simulation of complex systems in uncertain environment. He is a fellow of the Higher-Education Academy and IMarEST, a Senior Member of the IET, and a Chartered Engineer. He received the Best Paper Award for the Virtual Reality of Autonomous Vehicle in the 2018 10th International Conference on Modelling, Identification and Control sponsored by IEEE.

XIONGHU ZHONG (M'11) received the B.Eng. and M.Sc. degrees from Northwestern Polytechnical University, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Institute for Digital Communications, The University of Edinburgh, in 2010. He was a Research Fellow with the School of Computer Engineering, and then a Senior Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

In 2017, he joined Xylem Inc., as a Data Scientist. He is currently a Professor with Hunan University, China. His research interests are statistical signal processing, target localization and tracking, nonparametric Bayesian modeling and machine learning methods, and their applications to distant speech enhancement and recognition, V2X communications, and water distribution network monitoring.