Received February 13, 2021, accepted March 1, 2021, date of publication March 8, 2021, date of current version March 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064445

Water Leak Detection Survey: Challenges & Research Opportunities Using Data Fusion & Federated Learning

ABDALLAH MOUBAYED, (Member, IEEE), MOHAMED SHARIF, (Student Member, IEEE), MARCO LUCCINI, (Member, IEEE), SERGUEI PRIMAK, (Member, IEEE), AND ABDALLAH SHAMI, (Senior Member, IEEE)

Electrical and Computer Engineering Department, Western University, London, ON N6A 3K7, Canada

Corresponding author: Abdallah Moubayed (amoubaye@uwo.ca)

ABSTRACT With the increase in pipeline usage for fluid transportation, leak detection has become a major concern. More specifically, detecting water leaks has become a pressing challenge to both governmental and industrial stakeholders due to the financial losses it causes as well as the safety concerns associated with it. This issue is further highlighted in industrial and manufacturing environments such as the steel-making process in which a water leak into a furnace can cause a significant explosion that would threaten both the facility and its operators. Therefore, many different water leak detection methods belonging to different types (hardware-in-the-loop-based, simulation-in-the-loop-based, or hybrid) have been proposed in the literature. However, many of these methods either are computationally complex or only suitable for particular applications. Hence, there is a need to develop innovative and novel frameworks that offer effective and efficient water leak detection mechanisms. To that end, this article discusses two different paradigms, namely sensor data fusion and federated learning, that have the potential to further enhance water leak detection methods. Therefore, this article first surveys the different water leak detection methods proposed in the literature along with their merits and limitations. It then describes the sensor data fusion and federated learning paradigms in more detail. Moreover, it presents different research opportunities in which these paradigms can be implemented to offer a more effective and computationally efficient water leak detection framework.

INDEX TERMS Water leak detection, sensor fusion, federated learning.

I. INTRODUCTION

The number of pipelines and similar structures that are being designed and deployed is rapidly increasing due to the increase in the need to transport gases or fluids (such as oil or water) from production sites to end user areas [1]. These pipelines may carry toxic or hazardous content and often pass through high population areas or environmentally sensitive areas [1]. As such, there is a need to constantly monitor these pipelines to ensure their safety as well as the safety of the surrounding environment and population. Depending on the nature of these pipelines in terms of deployment mode (buried or over the ground), build material (metal, plastic, etc.), and content being transported (gas, oil, water, etc.), different leak detection methods have been proposed in the literature ranging from visual inspection to complex combined hardware-software setups [1]. This work focuses mainly on the problem of water leak detection, especially in industrial and manufacturing environments which may be considered to be harsh environments.

Water leak detection is a major concern for various governmental and industrial stakeholders. This is due to the associated damages and costs. For example, reports in 2013 suggested that 13% of the extracted water by Canadian municipal water suppliers is lost during the distribution process [2]. Similarly, the United States Environmental Protection Agency (EPA) states that an average of 10,000 gallons of water are wasted per year by a household, amounting to close to 1 trillion gallons of wasted water nationwide [3]. This has a negative impact on multiple levels.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

For example, higher levels of water leaks are associated with revenue loss, higher stress on the aquatic ecosystem due to increased levels of water extraction from lakes and streams, a reduction on system reliability, and a contribution to pipe failures [2]. To further illustrate the associated costs, the EPA estimates that the added cost per year for a small broken water distribution line is close to \$ 64,000 [4]. As such, detecting water leaks quickly and effectively is a crucial task to reduce the overall costs and improve the reliability of the systems dependent on the water usage.

In addition to the economic and financial costs associated with water leaks, there is a safety concern especially in industrial and manufacturing environments. This is evident in industries such as steel manufacturing in which furnaces are commonly used. With furnaces being continually being pushed to the limit, ensuring their safe and reliable operation is crucial given the associated risk of water leaks in steel manufacturing. This is highlighted by the many furnace accidents that included water leaks. For example, a steel making company in California had a furnace accident in 2004 that led to a major injury to one of its workers [5]. A furnace exploded while a safety technician and three of his coworkers were trying to stop a water leak in an EAF [5]. The explosion led to hot steam and flying debris being emitted as well as blowing out the front observation glass and the back window of the control room. As a result, the technician was hospitalized with severe burns and injuries. A more recent accident in 2014 that resulted in the death of one steelworker and the injury of five further employees due to a hydrogen explosion [5]. The explosion was due to more than 1,000 gallons of water leaking into the EAF that had a content heat of up to 2900^o F. The leakage resulted in fragments of metal and debris to be tossed out, killing the steelworker and injuring the remaining employees [5].

To illustrate the danger of water leaking into furnaces, a study conducted by Tveit *et al.* showed that the equivalent of 32 kg of TNT is generated if 10 liters of water are entrapped in 1000 kg of aluminum [6]. Given that 1 kg of TNT would destroy all windows within a radius of 30 m, the explosion equivalent to 32 kg is extremely dangerous and potentially fatal with injuries being both directly and indirectly caused by the pressure wave from the explosion [6]. Additionally, moving debris and liquid metal may injure people near the explosion [6]. Hence, efficiently and effectively detecting water leaks in such environments is a crucial yet challenging tasks to ensure that operators and workers are safe by detecting leaks more accurately and in a shorter time duration.

As shown in Fig. 1, there are multiple water leak detection techniques that have been proposed in the literature ranging from hardware-in-the-loop (HIL)-based to simulation-in-the-loop (SIL)-based techniques [7]. Regardless of the technique used, the amount of data generated by water distribution monitoring systems is large. As an example, experiments conducted by Liu *et al.* showed that 100 datasets of size 5000 samples can be generated in 1200s [8]. This is equivalent to 1.5 million samples/hour for a monitoring network consisting of around 600 nodes [8]. To address this issue, multiple potential solutions can be adopted. One such solution is the usage of sensor data fusion techniques to combine and compress the amount of data analyzed. Data fusion refers to the process of integrating data and knowledge from several sources [9]. The goal is to combine information from several sources in order to form a unified picture of the application/task at hand [10]. Applying data fusion techniques, particularly when having a system with multiple sensors, comes with various advantages such as enhanced data authenticity and availability, reduced redundant data exchanged, and reduced energy consumption to transmit this data [10]. As such, they are a promising solution to be adopted for water leak detection systems given the systems' general structure.

Another promising paradigm to adopt is federated learning (FL). FL is a machine learning (ML) paradigm in which a high quality centralized model is trained using data that is distributed over a large number of locations [11]–[13]. The term was first coined by Google in 2016 when they proposed a mechanism in which data at each location is used to independently compute an update of the current ML model [11]. This update is then communicated back to a central service that aggregates these updates to compute a new global model that is distributed back to the different locations [11]. Accordingly, this paradigm adopts the "bringing the code to the data" philosophy rather than "bringing the data to the code" [12]. As such, the FL paradigm addresses concerns regarding the data privacy, ownership, and locality [12]. Given the distributed nature of water leak monitoring systems with sensors collecting data at various geographical locations, FL promises to be a viable solution for extracting meaningful information from the collected data while still maintaining its privacy and locality.

This work focuses on surveying some of the published research tackling the water leak detection problem and discussing their merits and limitations. It also describes the different sensor data fusion techniques and FL paradigms previously proposed in the literature. Moreover, it discusses how they can be adapted and adopted to enhance the leak detection systems.

The remainder of this article is organized as follows: Section II discusses the different water leak detection techniques proposed. Section III describes the different sensor data fusion techniques and proposes potential research opportunities in which they can be deployed for effective water leak detection. Similarly, Section IV provides a mathematical background of the federated learning paradigm and presents the research opportunities to apply it for water leak detection. Finally, Section V concludes the article.

II. WATER LEAK DETECTION

Pipeline leak detection has garnered significant interest from both academia and the industry. As such, several different methods have been proposed in the literature to detect these leaks [7]. These methods can be divided into three main

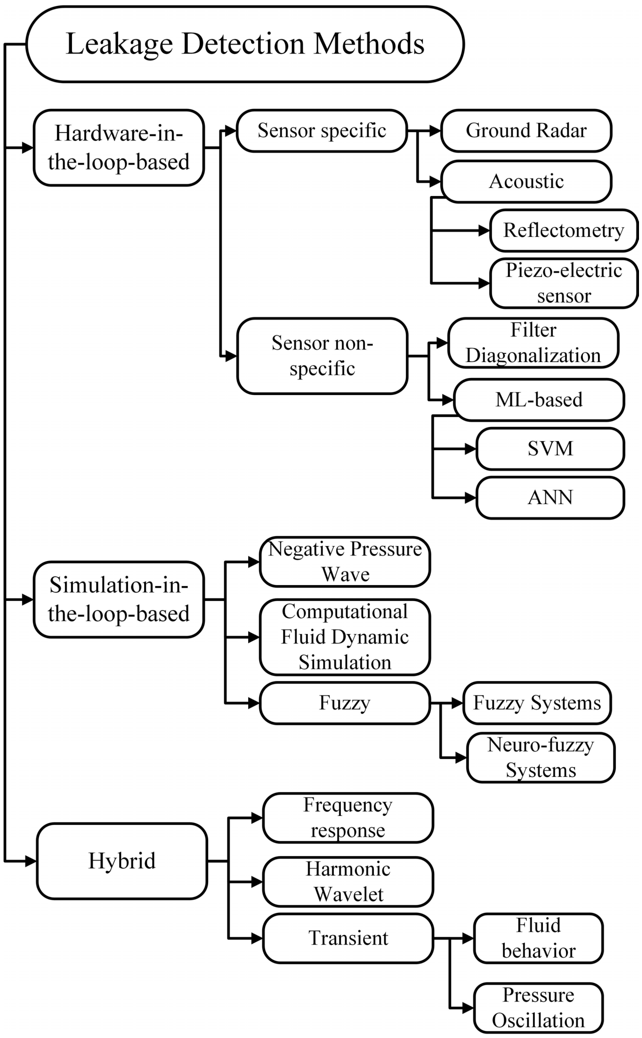


FIGURE 1. **Summary of leak detection methods.**

categories, namely HIL-based methods, SIL-based methods, and hybrid methods [7]. HIL-based methods mainly focus on the analysis of the data collected by special sensing devices to detect leaks in pipes [7]. In contrast, SIL-based methods focus on the use of various software and simulation programs to simulate leakages and develop integrated leakage detection models [14]. A third category combines the HIL-based methods and SIL-based methods by collecting/analyzing data through both the sensors and leakage models to detect leaks within pipes. Fig. 1 summarizes the different methods within each category. Below, a brief description of these detection methods is given along with a discussion of their advantages and limitations.

A. HARDWARE-IN-THE-LOOP-BASED LEAK DETECTION METHODS

The first category of leak detection methods is HIL-based methods. It focuses on the analysis of data collected by special sensing devices such as optical fibers, cable sensors, and hydrophones to detect leaks in pipes [7]. There are a multitude of methods that fall within this category that can be categorized as sensor-specific (*i.e.* dependent on the sensor used) and sensor non-specific (*i.e.* can be applied to data collected by different sensors). Below, some of the prominent HIL-based methods are discussed.

1) SENSOR-SPECIFIC

As the name suggests, this category of methods and algorithms is reliant on the type of sensor used to collect the data. As such, these methods can only be applied when the associated sensor is used for data collection. Below are two of the most prominent algorithms within this category.

a) Ground penetrating radar: One promising sensor specific HIL-based method that has been proposed in the literature is Ground Penetrating Radar (GPR) method. The GPR method belongs to the group of near-surface geophysical (NSG) methods that analyze different types of wave and induction properties in materials [15]. This method sends high frequency electro-magnetic pulses and receives back the reflected echo [15]. Using the reflected signals combined with sophisticated signal processing techniques, a three-dimensional map of the sub-surface can be generated [15]. GPR has been proposed to detect water leaks, particularly for underground pipelines, due to three main reasons. The first being that the GPR radar waves' velocity and reflection strength are significantly affected by water, making them a prime candidate to detect water leaks [15]. The second reason is the high resolution images and non-contact nature of that GPR method entails. More specifically, since the waves are in often in the high MHz to the GHz range, the resulting image resolution is in the millimeter accuracy without having to be in direct physical contact with the pipeline, making it suitable for underground imaging. The third reason is that the wide range of frequency used in GPR method allow it to detect leaks at different physical scales of structure thickness [15]. This allows for the detection of water leaks in roads, slopes, and seawalls possible through in-lab data analysis [15]. Based on the aforementioned advantages, GPR has been proposed in the literature [15], [16].

Lai *et al.* proposed the use of GPR method as a tool to detect water leaks in buried pipelines [15]. Accordingly, the authors studied the perturbation patterns of GPR signals (between 1.6-2 GHz) in different metallic and PVC pipelines [15]. The patterns were then compared to the no-leak condition based on dielectric contrast, reflection coefficient, and the corresponding absorption mechanisms to detect the location and size of water leaks [15]. As a result of this method, water leak signatures and fingerprints were accurately identified at the top of the pipes.

In a similar fashion, Senin *et al.* also used GPR method at 1.6 GHz to detect water leaks in shallow buried PVC pipes [16]. The authors studied the influce of the soil's moisture on the reflection characteristics of the GPR signal to detect the water leakage [16]. Experimental results showed that the reflected signal power was attenuated and that delayed hyperboles appeared when the soil water content increased [16]. Also, it was shown that the proposed method detected leaks up to 24 cm [16].

Despite the great promise shown by GPR method, it still faces some limitations. One limitation is that it cannot detect leaks accurately when the pipes are reinforced by bars in concrete pavements [7]. Moreover, such methods are deemed expensive due to the equipment used [7]. For example, renting a MALA Easy Locator Pro WideRange HDR GPR can cost around $5,625/month [17]. Also, buying a limited version costs around $6,000 [18]. This illustrates the cost of the equipment needed for such a method.

b) Acoustic: Another type of sensor specific HIL-based water leak detection methods is the acoustic-based methods. Such methods rely on the fact that a unique sound/noise is generated when water leaves the pipe with small leaks generating higher-frequency sounds and large leaks generating lower-frequency sounds [19]. Therefore, the goal of these methods is to identify the abnormal sounds/noises resulting from water leaks in pipes [19]. More specifically, these methods rely on devices such as listening rods and hydrophones installed either on the pipe surface or inside it to detect these sound/noise changes and thus detect the water leaks [20]. Among the acoustic-based methods, we discuss below two methods, namely the acoustic reflectometry method and the Acoustic Piezo-electric sensor-based method.

(i) Acoustic Reflectometry: The acoustic reflectometry method relies on injecting a sound pulse into the pipeline and focuses on amplifying the reflected wave [21]. The idea is that any crack in a pipe will represent a change in its cross-sectional area. Thus, this method benefits from the characteristic that the acoustic wave propagation in fluids is extremely sensitive to any discontinuity in the fluid's properties [21]. As a result, this change in cross-sectional area will cause a fraction of the incident acoustic wave to be reflected back and detected by any listening rod or hydrophone installed [21]. Due to its characteristics and cost-effectiveness, several previous works have proposed its use for water leak detection [21], [22].

Papadopoulou *et al.* proposed using acoustic reflectometry to detect leaks in large diameter pipes having long lengths [21]. By injecting different acoustic waves at different frequencies, the experimental results showed that the acoustic reflectometry method successfully detected pipe holes in single pipelines and pipeline networks as it is able to identify extremely small holes representing 1% of the pipeline diameter [21]. Moreover, it was shown that it is not affected by the pipe material [21].

Abdullahi and Oyadiji also proposed the use of acoustic reflectometry along with a modal frequency technique to detect leaks in a pipeline system [22]. Their experimental results showed that the size of leaks has a direct impact on the amplitude of the reflected waves with larger leaks having higher amplitudes. This is applicable to both the first reflection as well as the second reflections [22]. Moreover, it was shown that the relationship between the ratio of the first and second reflected waves in pipes and the percentage of the pipes' leak areas can be characterized by a polynomial function [22].

Acoustic reflectometry does have some limitations. One limitation is that the presence of high noise levels makes the detection process extremely challenging [7]. Moreover, it requires significant signal processing to accurately detect leaks, making it less suitable for large networks [7].

(ii) Acoustic Piezoelectric Sensor: Another acoustic-based method to detect water leaks proposed in the literature is to use piezoelectric sensors coupled with acoustic emissions. In general, piezoelectric sensors are devices that are often used to measure the change in different metrics such as pressure, temperature, or acceleration by converting them to an electric charge [23]. Using this property, such sensors are able to detect changes in the acoustic wave transmitted within the pipe to detect leaks in water pipes [24]. This technique is gaining popularity with multiple works from the literature adopting it to detect leaks [24], [25].

Ozevin proposed the use of acoustic emission and piezoelectric sensors to detect and localize cracks in pipes [24]. The proposed method combines the geometric boundaries and the local coordinate system to determine the shortest direct path between the acoustic source and the sensor [24]. The author's simulation results showed that the proposed method is capable of accurately detecting and localizing the crack even in cases where the acoustic source and corresponding senor are not in a straight path [24].

Ozevin extended the previous work by proposing the use of acoustic emission and piezoelectric sensor to detect leaks in pipeline networks deployed in a two-dimensional configuration [25]. The proposed method determines the difference in arrival times using cross-correlation function and uses geometric connectivity to determine the path that the leak waves followed to reach the sensor [25]. The authors again showed using simulations that the proposed method is effective in detecting and localizing the leak in a multi-dimensional space [25].

Despite its promise in detecting water leaks accurately and facilitating the online monitoring of pipelines, this method suffers from one main limitation. Using piezoelectric sensors with acoustic emission method cannot accurately detect leaks in complex pipeline networks [7]. This limitation makes this method unsuitable for large water distribution networks.

2) SENSOR NON-SPECIFIC

In contrast to the first category, sensor non-specific methods and algorithms do not depend on the type of sensor used to collect the data. As such, these methods can be applied to data collected by different types of sensors. Below, we discuss two well-known algorithms of this category.

a) Filter Diagonalization Method: One sensor non-specific HIL-based water leak detection method that has garnered interest is Filter Diagonalization method (FDM). The FDM method is a nonlinear, parametric method used to fit a time signal to the sum of sinusoids [26]. This is commonly referred to as the solution to the harmonic inversion problem [26]. The beauty of FDM is that it converts the often large and ill-conditioned non-linear fitting problem to a set of pure linear algebra problems of diagonalizing some small data matrices in the frequency domain [26]. Hence, it has been successfully applied to spectral analysis of experimentally measured time signals [26]. Due to its properties, FDM has been proposed in various previous works for water leak detection [27], [28].

Lay-Ekuakille *et al.* proposed the use of FDM for water leak detection to overcome the limitations of the Fast Fourier Transform (FFT)-based spectral analysis [27]. Using the FDM method, the authors were able to solve problems arising from zigzag pipelines and buried water tubes [27]. Experimental results showed that the proposed FDM method outperformed the FFT method by detecting the leaks sooner and localizing it with more accuracy [27].

Another work by Lay-Ekuakille *et al.* also adopted the FDM method to describe the necessary indicators of uncertainty and accuracy in detecting leaking cracks on water pipelines [28]. In this work, the authors processed data collected from pipe-mounted magnetic sensors to detect the leaks [28]. Again, it was shown through experimental results that the FDM method coupled with quadratic regression is better able to detect spectral peaks resulting from leaks [28]. This further highlights the promise of the FDM approach in facilitating the water leak detection problem [28].

Although the FDM method has proven to be a promising approach for water leak detection, it still has some drawbacks. One main drawback is the fact that the localization process is still erroneous [7]. Although it was shown that uncertainty can be reduced to less than 5%, this might not be enough, particularly for long pipelines where a 5% uncertainty may be equivalent to hundreds of meters. As such, such method may be more suitable for short or small sized pipelines.

b) Machine Learning (ML)-based Methods: Another group of sensor non-specific algorithms is ML-based algorithms. These algorithms have been proposed in the literature to detect water leaks due to their high level of dynamicity and adaptiveness as they allow systems to "learn" without being told what to do [29], [30]. Within this group, we discuss two well-known algorithms, namely support vector machines and artificial neural networks.

(i) Support Vector Machine: The first ML-based leak detection method is using support vector machines (SVM). In general, SVM is a supervised machine learning classification algorithm that focuses on finding the optimal hyperplane that separates the labeled training data with the maximum margin from the closest point [31], [32]. It is a more powerful and restrictive classifier than other classification algorithms such as logistic regression [31], [32]. Due to its versatility and classification accurateness, it has been proposed as an effective leak detection method using data collected by different types of sensors ranging from optical signals to pressure sensors and flow sensors [33], [34].

Qu *et al.* proposed an SVM-based leak detection system that classifies data collected from an optical fiber [33]. In the considered system, the optical fibers installed in parallel with the pipeline and acted as vibration sensors [33]. The SVM classifier was then trained to detect three different types of anomalous events, namely pipeline leaks, manual digging, and human walking [33].

In contrast, Mashford *et al.* proposed the use of an SVM classifier to detect water leaks based on data collected through pressure sensors or flow-measuring devices [34]. The authors used a pipe network simulation tool named EPANET to generate a training dataset with various pressure values representing different leak sizes along a pipeline [34]. As part of the simulations, the authors generated 10 data sets, each containing 150 samples. Out of the total 1500 samples generated, 1000 were chosen to act as the training dataset while the remaining 500 samples acted as the testing dataset. Through SVM parameter optimization, the classifier achieved an average classification accuracy of close to 77% with a maximum accuracy of close to 90% in some cases, highlighting the effectiveness of the proposed model [34].

Although SVM-based methods have shown that they can detect leaks with high accuracy, they still suffer from some shortcomings. The first is the fact that they often require a large training dataset to effectively learn leak detection [7]. This is illustrated in [34] and [35] in which the EPANET simulation tool was used to be able to generate the required dataset. Another shortcoming is the high computational complexity associated with such methods as it can be shown that the time complexity of SVM algorithm can reach $O(n^3)$ where n is the number of training samples [36]. This is illustrated by the model training time needed shown in [37] which reached up to 410s for a dataset comprised of around 2000 samples. Moreover, given that such models need to be re-trained periodically, this time complexity becomes a concern.

(ii) Artificial Neural Networks: Another ML-based leak detection technique is artificial neural networks (ANN). ANN is a popular supervised classification algorithm often used whenever abundant labeled training data with many features is available and a nonlinear hypothesis function is desired [38]. It mimics the way our brain works by adopting a similar structure of neurons, dendrites, and axons found within the human brain [38]. More specifically, the features of the dataset act as dendrites with the neurons being the computational units. The output is the value of the hypothesis function [38]. The structure of ANN helps extract more information from the feature set. As such, ANN has been proposed as an effective ML-based leak detection technique [39], [40].

Zadkarami *et al.* proposed a multi-layer perceptron (MLP), a variant of ANN, to detect the presence, location, and severity of pipeline leaks [39]. This MLP acted as a feature extraction and classification technique with multiple statistical and wavelet being collected. The proposed model was applied to a 20-km pipeline in southern Iran (Goldkari-Binak pipeline). The experimental results showed that the proposed model achieved a detection accuracy of 92% with a low false alarm rate, highlighting its effectiveness [39].

In a similar fashion, Li *et al.* proposed the use of ANN as a feature extraction and classification model to detect leaks in water distribution pipelines [40]. More specifically, the authors investigated the acoustic characteristics of leak signals in the socket and spigot pipe segments [40]. Experimental results showed that the proposed model achieved a leak detection accuracy ranging between 96.9%-97.2% for different feature sets such as {Peak, Mean, Peak Frequency, Kurtosis} and {Mean, Peak Frequency}.

Similar to the SVM case, ANN have multiple shortcomings despite their accurate detection of leaks. One shortcoming is the interpretability of ANN. More specifically, given the structure of ANNs and how the different layers interact with each other, it is not easy to interpret how the ANN made the classification decision. This in turn makes them less desirable for public adoption [41]. Another shortcoming is the high computational complexity of ANN. More specifically, the complexity of ANN can reach $O(n^4)$ where n is the number of training samples [42]. This is problematic, especially given that these models need to be re-trained periodically.

B. SIMULATION-IN-THE-LOOP-BASED LEAK DETECTION METHODS

The second category of leak detection methods is SIL-based methods. This category of methods focus on the use of various software and simulation programs to simulate leakages and develop integrated leakage detection models [7]. There are a multitude of methods that fall within this category. Some of the most prominent SIL-based methods are described and discussed below.

1) NEGATIVE PRESSURE WAVE

One of the most prominent simulation-based leak detection methods is the negative pressure wave (NPW) method. The NPW accounts for the pipeline physical model to derive mathematical expressions that describes the amplitude change of the NPW and its attenuation as it travels through the pipeline [43]. Using the these expressions, the method is able to determine the smallest detectable leakage flow rate which in turn helps in detecting the presence and location of the water leak in the pipeline [43]. Due to its characteristics, it has been proposed to detect and localize water leaks [44], [45].

Li *et al.* developed an NPW-based method that relies on the attenuation of the NPW to detect the leaks [44].As such, the authors developed a model that takes into account the pressure difference rather than the time difference to determine the leak location [44]. The authors' experiments showed that the proposed method achieved an average error ranging between 0.355% and 1.161% in detecting and localizing the leaks in pipelines, highlighting the accuracy of the proposed NPW-based method.

Wang *et al.* also designed a new NPW-based leak detection method that studies the pressure change using fiber Bragg grating sensors [45]. The authors built a testing platform to evaluate the performance of the proposed method. Experimental results showed that the proposed method was able to accurately calculate the trends in pressure changes throughout the pipe, facilitating the online calculation of the NPW velocity. Furthermore, it was shown that the proposed method is able to detect smaller leak volumes with higher accuracy.

Although the NPW method has proved to be successful in detecting and localizing leaks in pipelines, it was shown that it is not suitable for short distance transportation pipelines [7]. This is supported by the different experimental setups considered in which the length of the pipe used covered a long distance [45], [46]. As such, this method is more suitable for long-distance water distribution pipelines.

2) COMPUTATIONAL FLUID DYNAMIC SIMULATION

A second simulation-based leak detection method proposed in the literature is the computational fluid dynamic simulation (CFDS) method. This method relies on the principles of computational fluid dynamics (CFD). CFD refers to the use of numerical analysis to analyze and solve problems involving flows of fluids [47]. Accordingly, different simulation tools and computers are used to calculate and simulate the stream flow of fluids and their interaction with the adjacent surfaces [47]. Due to its nature, the CFDS method has been proposed in various previous works in the literature focusing on leak detection [48], [49].

Ben-Mansour proposed the use of the CFDS method to detect leaks in pipes with a small diameter [48]. The goal was to detect small leaks in such systems. Accordingly, a group of simulations were conducted showing the existence of distinct patterns in the pressure and pressure gradient variations [48]. Moreover, the simulations showed that these patterns, particularly for the pressure gradient, also exist even for small leaks (below 1 l/min).

Jujuly *et al.* focused on studying leaks in sub-sea pipelines and their impact on their surroundings. [49]. To do so, the authors adopted a CFDS-based methodology using the ANSYS FLUENT software to better understand the internal flow within the pipeline and the corresponding consequences of leaks [49]. Simulation results showed that an increase in the pipeline's operating pressure resulted in an increase in the escaping fluid's flow rate [49]. Moreover, it was observed that more noise is generated by high-pressure fluid flows than with low-pressure fluid flows [49].

One limitation of the CFDS method is that it is difficult to predict the leaks in pipelines [7]. Although the method succeeds at providing an analytical model to characterize the size and location of the leak, the main challenge is matching this model to the data that is collected. Hence, this method is more suitable for short and straight pipelines in which the visual inspection can be conducted easily based on the data collected [7].

3) FUZZY-BASED METHODS

A third type of simulation-based leak detection methods is fuzzy-based methods. These methods use fuzzy logic fundamentals and principles to characterize and detect potential leaks. Simply speaking, fuzzy logic refers to the notion of following a many-valued logic for which the truth values can take any real number within the range 0 to 1 [50]. Accordingly, it can handle the concept of partial truth. Therefore, fuzzy logic has been adopted as a method to recognize, represent, interpret, and utilize vague data and information [50]. Due to its characteristics, several previous works from the literature proposed the use of fuzzy-based methods to detect leaks including pure fuzzy and neuro-fuzzy systems [51]–[53].

Mamlook and Al-Jayyousi proposed the use of fuzzy set methodology to detect problems in environmental systems [51]. More specifically, they focused on water distribution systems with the goal of detecting water leaks. As such, the authors considered a fuzzy set containing three scenarios, namely no leakage, partial leakage, and leakage [51]. Using data from a water distribution system in Jordan, the authors evaluated the performance of their proposed method and determined that there are four major factors that directly impact leak detection, namely the water demand patterns, pipe age, pipe material, and the associated costs [51].

Similarly, Da Silva *et al.* proposed the use of fuzzy logic methods to detect pipeline faults [52]. To do so, the authors initially trained the fuzzy system offline using a modified dataset that included simulated leaks [52]. The proposed method was then tested on a small-scale pipeline with results showing relatively low false alarm rates and higher leak detection accuracy at low testing computational complexity.

Jalalkamali *et al.* proposed the use of genetic algorithm coupled with adaptive neuro-fuzzy inference systems to detect water leaks [53]. The goal is to learn the often nonlinear and complex relations between pressure changes and leakage rate to facilitate the early and quick detection of leaks [53]. To that end, the authors considered two water distribution networks in Iran to evaluate the effectiveness and efficiency of their proposed model. Their simulation results showed that the proposed model is able to accurately estimate the leakage rate in a compact and efficient manner [53].

Despite the high accuracy and the easy to interpret outputs that fuzzy methods provide, they tend to have high computational complexity. More specifically, different stages of the fuzzy-based systems can have quadratic complexity in terms of the number of inputs, number of outputs, and the number of rules generated [54], [55]. This is also further exacerbated when considering neuro-fuzzy systems that have the additional complexity of the classification method adopted. This makes such methods difficult to process, especially when having large datasets as evident from [52] in which the training had to be conducted offline.

C. HYBRID LEAK DETECTION METHODS

The third category of leak detection methods is hybrid methods. This category of methods combines the HIL-based methods and SIL-based methods by collecting data through both the sensors and leakage models to detect leaks within pipes [7]. Multiple methods belong to this category. In what follows, some of the most prominent hybrid methods are described and discussed.

1) FREQUENCY RESPONSE DIAGRAM

The frequency response diagram method is one of the many hybrid water leak detection methods proposed in the literature. This method relies on the fact that the transient behavior of fluids is impacted by multiple pipeline features such as potential leaks and blockages that are particularly evident in the frequency domain. To that end, the frequency response diagram, a plot that characterizes the dynamics of a system in the frequency domain by providing the output spectrum in response to a specific stimulus [56], is a promising method to facilitate the detection of leaks of water pipelines by analyzing it in the frequency domain. Accordingly, many previous works have proposed its use for leak detection [57], [58].

Lee *et al.* proposed the use of the frequency response diagram to detect leaks in water pipes [57]. The proposed method relies on injecting a fluid into the pipe and analyzing the resultant transient trace in the frequency domain. It was shown that the proposed method is able to efficiently detect and localize the leak within a specific region of the pipe using a combination of the resonance peak-sequencing method and the inverse resonance method [57]. The authors extend their work by studying the performance of the frequency response method in detecting leaks in a single pipe [58]. Again, it is shown that leaks illustrate a varied pattern in the resonance peaks of the diagram, a property that is used to indicate the presence of a leak [58]. The authors also derived analytical expressions to describe the resonance peak patterns [58]. It was shown that the proposed method is able to detect multiple simultaneous leaks along with their size and location on the condition that the loss due to leaks is less than 30% of the system's total flow [58].

One limitation of the frequency response diagram method is that it is difficult to detect leaks that are at the mid-point of the pipe [7]. This is attributed to the fact that when the leak is at the mid-point, its impact in the frequency domain often falls at harmonic frequencies and as such is not always detected. Hence, this limitation has to be considered whenever this method is used.

2) HARMONIC WAVELET

Another hybrid water leak detection method is based on harmonic wavelet analysis. As mentioned earlier, faults in pipelines such as leaks result in a change in the transient behavior of fluids flowing within it which is reflected in both time and frequency domain. Accordingly,the harmonic wavelet analysis method allows users to better understand and characterize the fluid behavior in both time and frequency. This is due to the fact that the harmonic wavelet transform is a linear wavelet-based transformation of a particular function into a time-frequency representation [59], allowing to identify any changes in both time and frequency domain. Due to its nature, it has been proposed in multiple previous works for leak detection [60], [61].

Ferrante and Brunone proposed the use of harmonic wavelet analysis method to detect leaks in pipes [60]. More specifically, the authors use the impulse response to solve the transient fluid flow equations in the frequency domain [60]. Then, the authors derive the analytical expressions describing the the piezometric head spectrum [60]. By comparing the experimental results with the analytical model describing the behavior in an intact pipe, the authors showed that the proposed harmonic wavelet analysis method is capable of detecting leaks within the pipe [60]. The authors extend their work by using the same method to localize the leak location [61]. To that end, the wavelet transform component of the method was applied to the experimental data collected to identify local singularities in the pressure time history [61]. Based on this information, the location of the leak can be determined using the discontinuity in the arrival time of the reflected leak-induced pressure wave.

However, the harmonic wavelet analysis method does suffer from one major shortcoming, namely its computational complexity. It was shown that this method has a high computational complexity as shown in [62] and [63]. Therefore, it is not suitable for real-time monitoring in which the fast detection of leaks is crucial.

3) TRANSIENT-BASED METHODS

A third type of hybrid water leak detection methods is transient-based methods such as transient fluid behavior and transient pressure oscillation. This type of detection method rely on the fact that leaks are considered to be hydraulic phenomenon [64]. As such, the presence of such leaks will inevitably result in a reflected wave for any transient signal which in turn alters the system's flow and pressure response [64]. Based on this concept, several literature works proposed the use of transient-based methods to detect leaks in pipes [65], [66].

Verde *et al.* proposed the use of the transient response to identify two leaks in a single pipe [65]. More specifically, they authors used the transient fluid behavior to determine the leaks' parameters [65]. Using a simulation setup of a 135 m pipe, the authors illustrated the promise and robustness of the proposed method by showing that it was able to accurately estimate the aforementioned parameters even when operation point changes occurred [65].

Duan built on the previous results by investigating the effectiveness of using transient response method to detect leaks in more complex pipe connection systems [66]. To that end, the author initially determined the transient response of the intact system to use it as a baseline and differentiate it from potential leakage scenarios within the system [66]. The author then proceeded to derive the analytical expressions describing the leak-induced patterns within the transient response. The derived models were validated through multiple numerical experiments which highlighted the accuracy of the proposed method for leak detection in complex pipe connection systems [66].

Despite their promise, these methods have several limitations. One limitation is the fact that they are not suitable for online implementation since the data needs to be gathered over a period of time before the transient response is calculated [7]. This contradicts with the goal of quickly detecting the leaks. A second limitation is that they do not perform well in noisy environments, particularly when using the transient pressure oscillation method [7]. Thus, it is not advised that such methods are used in an industrial manufacturing environment given that such environments are typically considered to be significantly noisy.

III. SENSOR DATA FUSION

Studies have shown that the amount of data generated by water distribution monitoring systems is large. For example, Liu *et al.* showed that 100 datasets of size 5000 samples can

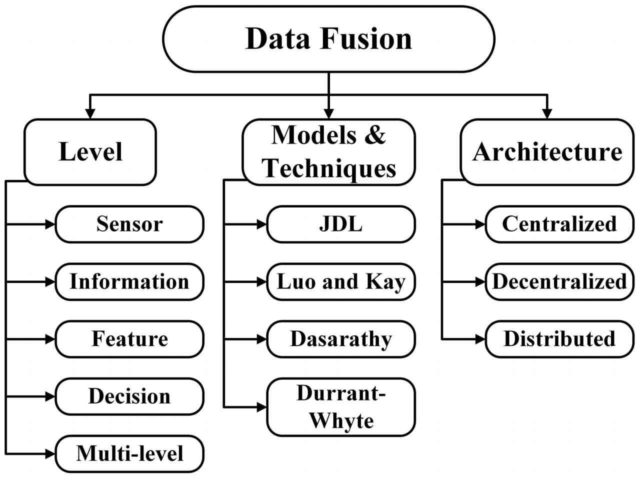


FIGURE 2. **Summary of data fusion methods.**

be generated in 1200s [8]. This is equivalent to 1.5 million samples/hour [8]. Multiple potential solutions can be adopted to tackle this issue. One such solution is using sensor data fusion techniques to combine and compress the amount of data analyzed. Data fusion refers to the process of integrating data and knowledge from several sources [9]. The goal is to combine information from several sources in order to form a unified picture of the application/task at hand [10]. Applying data fusion techniques, particularly when having a system with multiple sensors, comes with various advantages such as enhanced data authenticity and availability, reduced redundant data exchanged, and reduced data transmission energy consumption [10]. As such, these techniques are a promising solution to be adopted for water leak detection systems.

There are multiple ways to categorize data fusion techniques. They can be categorized based on the level at which the fusion occurs, on the model adopted, or the architecture deployed. These different potential categorizations are presented below, as illustrated in Fig. 2, along with a discussion of the different research opportunities of sensor data fusion for efficient water leak detection.

A. BACKGROUND

1) DATA FUSION LEVEL

One way to categorize data fusion techniques is the level at which this fusion occurs. Accordingly, fusion can occur at the sensor level, information level, feature level, decision level, or across multiple levels. Each level allows for different types of knowledge and insights to be extracted.

a) Sensor fusion: refers to the combining of the raw data or signals collected [9]. In this case, data collected directly from the sensors is combine without any processing [9], [10]. This is mostly common when having multiple sensors that are measuring the same characteristic or phenomenon such as temperature [67].
b) Information fusion: refers to the combining of processed data [9], [10]. In this case, raw data is processed to extract some level of information. This extracted information is then fused together to provide a more comprehensive representation of the system state [68]. As such, information fusion represents a higher semantic level of fusion than sensor fusion [68]. As an example, stating that the weather is "hot" when the sensed temperature is above 30^o C would be considered as a piece of "information". When multiple pieces of "information" are combined using information fusion techniques, a more comprehensive representation of the weather in a particular region can be provided.
c) Feature fusion: refers to the selection and combination of features to remove those that are redundant and irrelevant [69]. This represents an intermediate level of fusion as it is considered to be a higher semantic level of fusion when compared information fusion and data fusion [69]. Based on the definition provided above, there are multiple feature fusion techniques such as information gain-based and correlation-based feature selection, principal component analysis and independent component analysis feature extraction, and weighted and non-weighted serial/parallel feature combination [69]. Due to their nature, these techniques have been extensively proposed for ML problems to improve their accuracy as well as reduce their training time [69].
d) Decision fusion: refers to the fusion and combination of decisions [9], [10]. It represents the highest semantic level of fusion by combining the decisions or rules made concerning a particular system [68]. This is often done to reduce the amount of uncertainty made by a decision mechanism [68]. Due to its nature, decision fusion has also been extensively proposed in ML-based systems in which multiple classifiers or learners are combined in what is known as an ensemble model/learner [68]. This again is done to improve the accuracy of the ML model by reducing any bias the learner may have [70]
e) Multi-level fusion: refers to the combination at multiple levels such as at both the information and feature level or at the feature and decision level [9], [10]. Again, such a multi-level fusion is common in ML-based solutions as it allows the system to achieve improved performance (in terms of accuracy for example) with higher computational efficiency (due to the reduced feature size after feature fusion).

2) DATA FUSION MODELS & TECHNIQUES

A second perspective to consider when categorizing data fusion techniques is based on the model adopted. Many of the data fusion techniques are based on one of four main fusion models, namely the Joint Directors of Laboratories (JDL) model, Lou and Kay model, Dasarathy model, and the Durrant-Whyte Model [9]. Each of these models tackles the data fusion process from a different angle.

a) JDL model: The JDL is one of the most popular data fusion models used in the literature [71]. This model consists of three main components: data sources providing the input data, database management system that stores the information and fused data, and the human-computer interface (HCI) the provides the user with the resulting outputs based on the user queries and commands [71]. Within this model, five different fusion levels are defined and categorized as either low-level fusion (consisting of levels 0 and 1) or high-level fusion (consisting of levels 2, 3, and 4). The lowest level (level 0) is the source preprocessing in which data fusion is performed at the signal/sensor level [71]. The second level (level 1) is the object refinement in which further processing of the extracted information is performed such as association, correlation, and/or clustering. The third level (level 2) is the situation assessment in which relationships between objects is detected for high-level inferences [71]. The fourth level (level 3) is the impact assessment in which the risks are evaluated based on the aforementioned fused data and the logical outcome is projected. Finally, the fifth level (level 4) aims at efficiently managing the available resources while accounting for task priorities, scheduling, and the control over them [71].

b) Lou and Kay model: This model classifies the techniques based on the abstraction level at which fusion occurs [72]. Accordingly, it defines four abstraction levels: signal, pixel, characteristic, and symbol. The signal level deals with the data directly acquired from the sensor. The pixel level deals with a slightly higher representation and is often suitable for image processing tasks. The characteristic level fuses the features that are extracted from the signals and images. Finally, the symbol level represents the information as a symbol and is generally thought of as the decision level of fusion [72].

c) Dasarathy model: Another well-known and popular data fusion model is the Dasarathy model [73]. Similar to the Lou model above, it also deals with the abstraction level at which fusion occurs. However, the main contribution of this model is that it defines the abstraction level at both the input and output of the fusion technique, providing a framework to better classify it [73]. Therefore, this model categorizes fusion techniques into five main categories/classes. The first is the data in-data out (DAI-DAO) which is the most basic category as it deals with the raw data both at the input and output. The second category is the data in-feature out (DAI-FEO) which processes raw data from the sensors and extracts features/characteristics to describe the environment [73]. The third category is the feature in-feature out (FEI-FEO) in which both the input and output are features with the goal of improving or extracting new features. This category is also commonly referred to as information fusion or feature fusion [73]. The fourth category is the feature in-decision out (FEI-DEO) in which the features are used as an input to provide a decision. Most of the ML classification algorithms fall within this category [73]. Finally, the fifth category is the decision in-decision out (DEI-DEO) which aims at fusion decisions to provide improved or new decisions. As such, this is commonly referred to as decision fusion [73].

d) Durrant-Whyte model: This model categorizes the data fusion techniques based on the relationship between the data sources [74]. Accordingly, it defines three data fusion categories: complementary, redundant, and cooperative. The "complementary" category represents the case in which sensors are providing information about different parts of the system or scene (for example, two cameras capturing the same object from different angles) to provide a more comprehensive and global view of it. The "redundant" category represents the case where the input sources are providing information about the same target with the goal of improving the confidence upon fusion [74]. Finally, the "cooperative" category represents the case where the data from different types of sensors is combined to produce new data/information (for example, combining the video from a camera and the audio from a microphone to provide a more comprehensive view of the scene) that is often more complex yet more informative than the original data/information [74].

3) DATA FUSION ARCHITECTURE

The architecture adopted is another way to classify data fusion techniques. It refers to the location at which data fusion occurs. Thus, different potential architectures have been proposed such as the centralized, decentralized, and distributed architecture, each having its merits and limitations/challenges.

a) Centralized architecture: As the name suggests, the centralized architecture fuses the data collected at a central unit/location, often being in the cloud [9]. As such, all sensing nodes send the data collected to that central location for fusion and information extraction [9]. The merit of such an architecture is that it allows for a global view of the data that is collected from the different sensors [75]. However, it suffers from two major limitations/challenges, namely the high bandwidth required to send the data to the central location (potentially being the system bottleneck) and the single-point-of-failure that this central location represents. Therefore, multiple resiliency and redundancy mechanisms are often deployed when adopting such an architecture.

b) Decentralized architecture: The decentralized architecture is the complete opposite. In this case, each sensing node acts as a fusion node [9]. Therefore, it is assumed that each node has a minimum level of computing capabilities that allows it to fuse the data it collects along with that shared by its peers [9]. The merit of such an architecture is that it allows for parallel fusion with each node handling its local data and overcomes the single-point-of-failure limitation [76]. In contrast, its main limitation lies in its scalability, mainly due to the high communication cost ($O(n^2)$ at each communication step where n

is the number of nodes) to exchange the data between the peers. Therefore, it is often not recommended for systems consisting of a large number of nodes.

c) Distributed architecture: The distributed architecture represents a middle-ground between the two other architectures. Within such an architecture, a group of sensing nodes send the data collected to a local fusion node (sometimes referred at as a cluster head in wireless sensor networks) [9]. Local fusion nodes then send their fused data or information to a central location for further processing or fusion. Therefore, such an architecture aims at striking a balance between the single-point-of-failure challenge of the centralized architecture and the high communication cost of the decentralized architecture [77].

B. RELATED WORK

The concept of data fusion at different levels has been explored as part of leak detection systems of different fluids (oil, gas, water, etc.). The goal is to combine data or information collected at different levels of the leak detection systems to improve its performance. In what follows, a brief presentation of some of these previous works is given [78]–[82].

Guerriero *et al.* proposed a probabilistic data fusion model to detect and localize leaks in oil and gas pipelines [78]. To that end, the proposed model combined data collected from two distinct systems, namely the fiber optic Distributed Acoustic Sensing (DAS) system and the Internal Leak Detection (ILD) system. The fusion was performed by building a Dynamic Bayesian Network (DBN) using the information collected from the two aforementioned systems. Both the simulation and experimental results showed that the proposed data fusion model achieved lower false alarms, higher response times, and improved sensitivity [78].

Dong *et al.* proposed the use of data fusion to help improve the performance of gas leak detection [79]. More specifically, the authors proposed the use of a weighted fusion model that combines the data of the sensors assumed to be the closes to the leak. This is because it was shown that the measured gas concentration of a sensor decreases as the sensor gets further away from the leak source. Experimental results showed that the proposed gas leak detection model achieved a high detection rate (close to 96.7%) and a low average detection time delay ($\leq$ 30 s). This highlighted the effectiveness of the proposed model in detecting the leaks quickly.

Liu *et al.* proposed a data fusion model at two separate levels for oil leak detection [80]. The proposed scheme first focused on using feature fusion as part of the feature extraction process by combining different statistical features such as the relative fluctuation characteristics of data segment samples and their corresponding variance. This is then followed by a two-stage decision fusion model that accounts for both the short-term and long-term detection models. Experimental results showed that the proposed model achieved high accuracy (above 97%) and low false positive rate (average of 1%).

Doshmanziari *et al.* also investigated the use of sensor data fusion to detect leaks in gas pipelines [81]. To that end, the authors proposed the use of Extended Kalman Filters as state observers to estimate the presence of leaks. Moreover, the authors proposed the use of the Fisher method to combine the data collected from sensor arrays to increase system redundancy and improve leak detection accuracy. The authors used two simulation softwares, namely OLGA and PVTSIM, to mimic the behavior of a high-pressure operational pipeline. Simulation results showed that the proposed model had a low estimation root mean square error and variance. This confirmed the effectiveness of the proposed method.

Nkemeni *et al.* proposed a Distributed Kalman Filtering and Distributed data fusion model for water leak detection in water distribution networks [82]. The goal was to develop a localized detection model that can overcome the issue of long-distance transmissions via several hops to a centralized fusion center. Using the Cupcarbon 4.2 simulation tool (a Smart City and Internet of Things Wireless Sensor Network (SCI-WSN) simulator), the authors created a water distribution network that is equipped with multiple sensors and simulated the occurrence of leaks. Again, simulation results showed a low estimation root mean squared error values, indicating the effectiveness of the proposed model in detecting leaks in water distribution networks.

Despite the promise shown in the use of data fusion for leak detection, most of the previous works focused on its use in large distribution pipelines (e.g. oil pipelines, water distribution systems, etc.). However, very few considered using data fusion within the context of leak detection in industrial and manufacturing settings/environments.

C. RESEARCH OPPORTUNITY

Based on the description provided for the different potential levels, models, and architectures for data fusion, multiple research opportunities come to mind. One such opportunity is a novel hierarchical distributed multi-level data fusion framework. Within such a framework, raw data collected by the sensors installed along the cooling pipe of the furnace can be fused. This represents fusion at the sensor level and can be performed using an industrial level data processing board such as the industrial suited customized Raspberry Pi 3 series "NetPI" [83]. The fused data from multiple furnaces can then be aggregated and fused at a local server within the facility at both the information level and feature level. Again this represents a higher level of fusion as we move up the hierarchy of the framework. After potentially training different leak detection models at the local server within the facility, decision fusion can be performed across multiple facilities by sending the developed detection models to a centralized cloud server (*e.g.* Amazon cloud service) that is capable of fusing the multiple models to generate an improved leak detection model. Such a framework is beneficial as it would be able to leverage the knowledge gained across multiple furnaces within the same facility as well as across multiple facilities to improve the effectiveness of the leak detection model. At the same time, it would reduce the amount of data exchanged and hence improve the computational efficiency of the detection model. Fig. 3 illustrates this hierarchical distributed multi-level framework.

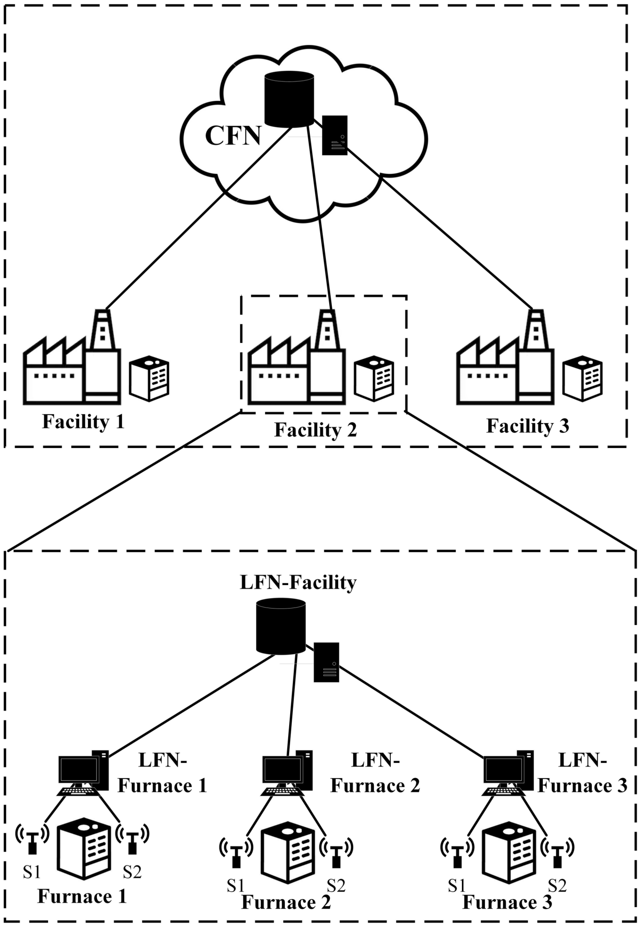


FIGURE 3. Proposed hierarchical distributed multi-level data fusion framework for water leak detection.

IV. FEDERATED LEARNING

As mentioned earlier, FL is a novel distributed ML paradigm that is gaining traction in both academia and industry. When adopting the FL paradigm, a high quality centralized model is trained using data that is distributed over a large number of locations and devices [11]–[13]. The term was first coined by Google in 2016 when they proposed a mechanism in which data at each location is used to independently compute an update of the current ML model [11]. This update is then communicated back to a central service that aggregates these updates to compute a new global model that is distributed back to the different locations [11]. Accordingly, this paradigm brings "the code to the data" rather than "the data to the code" [12]. As such, the FL paradigm addresses concerns regarding the data privacy, ownership, and locality [12]. In what follows, the mathematical concepts and mechanisms at the core of the FL paradigm as well as its potential in tackling the water leak detection problem is presented and discussed.

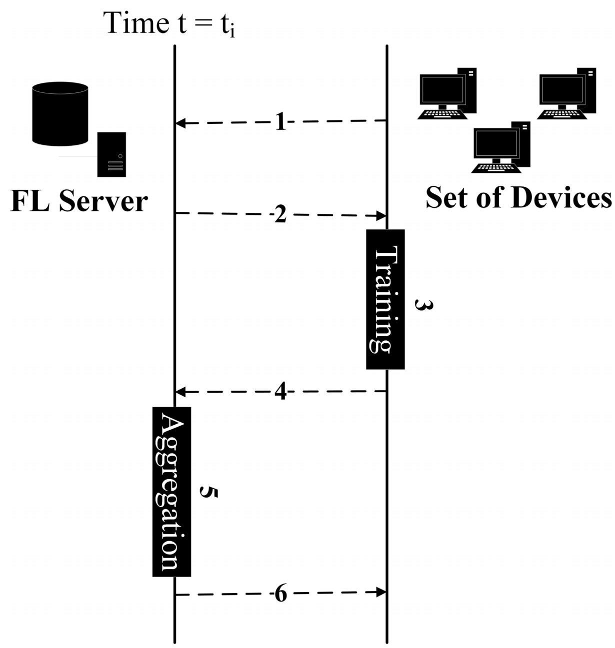


FIGURE 4. FL process workflow.

A. MATHEMATICAL BACKGROUND

To describe the FL paradigm, the general architecture is first described. The FL architecture typically consists of a centralized FL server that can communicate with a group of devices that are ready to perform the required FL task. The workflow can be divided into six main steps [11], [12]:

1) The set of devices send an availability message indicating that they are ready to complete an FL task.
2) The FL server chooses a subset of these available devices and shares with them the ML model at time t_{i-1}.
3) Each device then performs a training process based on the local data to determine a new local ML model.
4) Each device sends the updated parameters of its ML model based on the aforementioned training process.
5) The FL server then aggregates the local models to determine the updated global ML model for time t_i.
6) The FL server sends the updated global ML model to all devices.

This workflow is repeated for every round with the update frequency determined by the FL server. This workflow is described in Fig. 4.

Mathematically speaking, the FL paradigm aims at learning the parameters of the global ML model that can be represented by the matrix W. To do so, the FL server sends the model $W_{t_{i-1}}$ to a subset D_{t_i} devices out of D_{tot} total devices available. Each device $d^j_{t_i} \in D_{t_i}$ conducts a training process through which it determines an updated local model $W^j_{t_i}$. Accordingly, each devices then sends its update $H^j_{t_i} = W^j_{t_i} - W_{t_{i-1}}$ back to the FL server. The FL server then aggregates these local updates to generate the global model as follows [11], [12]:

$$W_{t_i} = W_{t_{i-1}} + \alpha_{t_i} H_{t_i} \tag{1}$$

where α_{t_i} is the learning rate chosen by the FL server and H_{t_i} is the average aggregated device-shared update given by:

$$H_{t_i} = \frac{1}{|D_{tot}|} \sum_{j \in D_{t_i}} H_{t_i}^{j} \tag{2}$$

It is worth nothing that the term H_{t_i} can be calculated as the weighted sum of the device-shared updates rather than the average for specific implementations [12].

B. RELATED WORK

Due to its recency, the FL paradigm has not yet been proposed for leak detection problems to the best of our knowledge. However, it has been proposed as a promising paradigm in many applications and industries that have significant focus on privacy protection and data security [13]. This includes applications such as finance, mobile security, vehicular network management, and Internet of Health Things [84]–[88].

Qin *et al.* proposed the use of an FL-based model for intrusion detection at the network edge [84]. The goal is to detect complex edge network attacks in an accurate, efficient, and scalable manner. To that end, the authors propose to train a binarized neural network (BNN) using the FL approach to reduce the communication overhead while still maintaining high classification accuracy. Experimental results using the CICIDS 2017 dataset showed that the proposed model achieved high detection accuracy (close to 98%) while having a lower complexity due to the binarization process. Moreover, it is shown that the packet processing latency is low (less than 2 ms for 95% of the packets) due to the adoption of the FL paradigm.

Lu *et al.* on the other hand proposed the use of FL paradigm to preserve the data privacy in vehicular cyber-physical systems [85]. The goal is to mitigate the leakage of data from different entities involved in vehicular networks such as autonomous vehicles and road side units (RSUs). In their proposed model, the aggregation process occurs asynchronously in each vehicle. This is done to improve both the security and efficiency of the trained ML models. To evaluate the performance of their proposed model, the authors used the 20 Newsgroups dataset. Experimental results showed that the proposed model achieved high accuracy and utility (around 0.9) while having a low running time (in the range of 0.8-1.4 s). This highlighted the near real-time defending capabilities of the proposed model.

Within the context of Internet of Things (IoT), Li *et al.* proposed the use of an FL-based model to detect advanced persistent threats (APTs) titled FLAPT [86]. APTs are menacing and stealthy multiple-steps attacks in IoT-related applications. The goal of the FLAPT model is to learn the different APT attack patterns by maintaining a global model across multiple clients. The performance of the proposed FLAPT model was evaluated using the UNSW-NB15 dataset with simulation results showing that it achieved a detection accuracy of 96.7%, highlighting its effectiveness in achieving its desired goal.

Similarly, Rahman *et al.* proposed a hybrid FL model within the context of Internet of Health Things (IoHT) [87]. The proposed hybrid FL model combines blockchain smart contracts and differential privacy to ensure the security, privacy, and provenance of IoHT data. Using their own testbed, the authors showed that the proposed model achieved a training accuracy above 90% and testing accuracy above 85%. Additionally, the proposed model achieved an average accuracy of around 89% across the different applications considered (e.g. patient recognition, pill detection, fever detection, human fall, etc.).

Hsu *et al.* a privacy-preserving FL (PPFL) system to detect malware in android devices [88]. Within the proposed system, mobile devices collaborate to train the detection classifier without revealing any sensitive information. The authors created their own dataset based on APK files collected between January and September 2014 from the Opera Mobile Store. Experimental results showed that the proposed PPFL model outperformed both the local models as well as the centralized model, particularly in terms of precision, recall, and F1-score. Additionally, it was shown that the PPFL system had a lower execution time compared to the centralized model due to the fact that the training (which is the time consuming part of the model) is distributed among multiple nodes with the aggregation step having negligible execution time. Thus, the proposed PPFL model not only improved the detection performance, but also reduced the required computation time, highlighting its effectiveness and efficiency.

As can be seen, the FL paradigm has illustrated its promise as a viable solution in multiple applications due to its distributed computing nature and privacy-preserving characteristics, particularly when applied for security purposes as illustrated above. Hence, given the nature of the architecture involved, the FL paradigm can be adapted to act as a potential solution for the water leak detection problem, particularly in industrial and manufacturing settings/environments.

C. RESEARCH OPPORTUNITY

Based on the above description, multiple research opportunities for water leak detection using the FL paradigm can be proposed due to its distributed computing nature and privacy-preserving capabilities, two prevalent and desirable characteristics in industrial and manufacturing environments. One such opportunity is applying this paradigm to multiple furnaces within one location. In this case, sensors placed within the cooling pipes of the furnaces are used to collect the data through an industrial level data acquisition (DAQ) board such as the industrial suited customized Raspberry Pi 3 series "NetPI" [83]. These industrial DAQ boards, acting as the FL devices, would be connected to a centralized server acting as the FL server. In this case, different ML algorithms for leak detection such as SVM or artificial neural networks can be trained on these DAQ boards (since they have enough computational power) based on the data collected at each individual furnace. These local models can then be aggregated at the centralized FL server within the facility to get the global leak

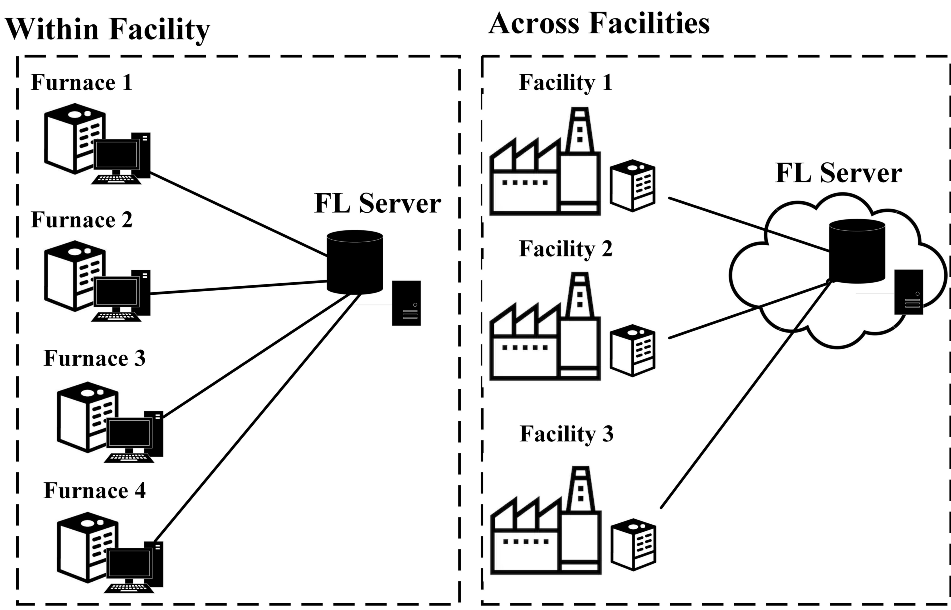


FIGURE 5. Proposed FL-based architectures for leak detection within a facility and across facilities.

detection model. The beauty of such an architecture is that it leverages information about leak behavior from multiple furnaces and extends this knowledge to other furnaces. This is particularly important given that it is highly unlikely that multiple furnace cooling pipes have leaks simultaneously. Hence, such an architecture allows the facility to learn from data collected at one furnace and apply this to react in a faster manner to any leaks at any other furnace within the facility.

In a similar fashion, this paradigm can be extended to facilities across multiple geographical locations. This is suitable for companies with multiple manufacturing facilities located at different geographical locations. Again, using the fact that it is unlikely to have multiple simultaneous leaks at different facilities, the FL paradigm allows companies to benefit from knowledge gained due to a water leak in one facility and apply this knowledge to other manufacturing facilities. In this case, each facility would act as an FL device by deploying a set of servers that are used to perform the local training. Different facilities would all be connected to a centralized cloud server (*e.g.* Amazon cloud service) that acts as the FL server aggregating the local models and sending back the global ML models used. As mentioned earlier, ML detection models such as SVM and artificial neural networks can be trained (as they have proven to be effective leak detection methods) at the facility level with the centralized FL server providing the global ML model after aggregation. Note that with such an architecture, it is expected that the data is either heterogeneous or is not independent identically distributed (non-iid) due to the facilities having different capacities or using different hardware equipment. To address this, multiple approaches can be adopted. One approach is to cluster similar facilities together and have one of them send the updates on behalf of cluster [89]. Such an approach would tackle the data heterogeneity issue as well as the heterogeneity of the computing capabilities at each of the facilities. The second approach is to globally share a portion of the data from each facility [90]. This would allow the local models being trained at each facility to view and observe data from other facilities. For example, Zhao *et al.* showed that sharing only 5% of the local data globally can significantly improve the quality of the global model [90]. Hence, a similar approach can be adopted for the multi-facility architecture to ensure that the global model at the centralized FL server is of higher quality. Fig. 5 illustrates these architectures with the left hand side describing the FL architecture within a facility and the right hand side describing it across multiple facilities.

V. CONCLUSION

The number of pipelines that are being designed and deployed is rapidly increasing due to the increase in the need to transport gases or fluids (such as oil or water) from production sites to end user areas [1]. Water leak detection in these pipelines is a major concern for various governmental and industrial stakeholders. This is due to the associated damages and costs [2]. In addition to the economic and financial costs associated with water leaks, there is a safety concern especially in industrial and manufacturing environments. This is evident in industries such as steel manufacturing where furnaces have been used at longer and faster rates to ramp up production. Therefore, ensuring their safe and reliable operation is crucial given the associated risk of water leaks in steel manufacturing. This is highlighted by the many furnace accidents that included water leaks which resulted in property damage as well as injuries and deaths of technicians [5]. Hence, efficiently and effectively detecting water leaks in such environments is essential to ensure that operators and workers are safe by detecting leaks more accurately and quickly.

There are multiple water leak detection techniques that have been proposed in the literature ranging from hardware-in-the-loop-based to simulation-in-the-loop-based techniques [7]. Regardless of the technique used, the amount of data generated by water distribution monitoring systems is large. To address this issue, multiple potential solutions can be adopted. One such solution is the usage of sensor data fusion techniques to combine and compress the amount of data analyzed. Applying data fusion techniques, particularly when having a system with multiple sensors, comes with various advantages such as enhanced data authenticity and availability, reduced redundant data exchanged, and reduced energy consumption to transmit this data [10]. Another promising paradigm is federated learning (FL). FL is a machine learning (ML) paradigm in which a high quality centralized model is trained using data that is distributed over a large number of locations [11]–[13]. Given the distributed nature of water leak monitoring systems with sensors collecting data at various geographical locations, FL promises to be a viable solution for extracting meaningful information from the collected data while still maintaining its privacy and locality.

This work focused on surveying some of the previous work that tackled the problem of water leak detection. It also described the different sensor data fusion models and FL paradigm previously proposed in the literature. Moreover, it presented a hierarchical distributed multi-level data fusion framework that leverages the knowledge gained across multiple furnaces within the same facility as well as across multiple facilities to improve the effectiveness of the leak detection model. Similarly, this work also proposed two FL-based architectures that can be implemented either within one facility or across multiple facilities. Therefore, combining the sensor data fusion and the FL paradigms can enhance the water leak detection systems, making them more effective and efficient.

REFERENCES

[1] T. Chis, "Pipeline leak detection techniques," 2009, *arXiv:0903.4283*. [Online]. Available: http://arxiv.org/abs/0903.4283

[2] S. Renzetti, D. Dupont, and D. P. Dupont, "Buried treasure: The economics of leak detection and water loss prevention in Ontario," Dept. Econ., Environ. Sustainability Res. Centre, Brock Univ., St. Catharines, ON, Canada, Tech. Rep. ESRC-2013, 2013, vol. 1.

[3] United States Environmental Protection Agency. *Fix a Leak Week*. Accessed: Sep. 1, 2020. [Online]. Available: https://www.epa.gov/watersense/fix-leak-week

[4] K. Slowey. (Jul. 2019). *Leak Detection Tech Can Save Contractors the Cost of Water Waste and Damage*. [Online]. Available: https://www.constructiondive.com/news/leak-detection-tech-can-save-contractors-the-cost-of-water-waste-and-damage/558967/

[5] S. Furgeson and N. Zsamboky, "Electric arc furnace explosions: A deadly but preventable problem," *Iron Steel Technol.*, vol. 14, pp. 30–37, Jan. 2017.

[6] H. Tveit, M. Garcia, H. Delbeck, A. T. Haug, B. Saugestad, and I. J. Eikeland, "Water leakages in ferroalloy and silicon reduction furnaces-experience gained from a severe accident in 2006," in *Silicon for the Chemical and Solar Industry IX*. Oslo, Norway: Department of Materials Technology, Norwegian Univ. of Science and Technology, 2008.

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

[8] Y. Liu, X. Ma, Y. Li, Y. Tie, Y. Zhang, and J. Gao, "Water pipeline leakage detection based on machine learning and wireless sensor networks," *Sensors*, vol. 19, no. 23, p. 5086, Nov. 2019.

[9] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, pp. 1–19, Sep. 2013.

[10] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.

[11] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*. [Online]. Available: http://arxiv.org/abs/1610.05492

[12] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, H. B. McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," 2019, *arXiv:1902.01046*. [Online]. Available: http://arxiv.org/abs/1902.01046

[13] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[14] H. Jin, L. Zhang, W. Liang, and Q. Ding, "Integrated leakage detection and localization model for gas pipelines based on the acoustic wave method," *J. Loss Prevention Process Industries*, vol. 27, pp. 74–88, Jan. 2014.

[15] W. W. L. Lai, R. K. W. Chang, J. F. C. Sham, and K. Pang, "Perturbation mapping of water leak in buried water pipes via laboratory validation experiments with high-frequency ground penetrating radar (GPR)," *Tunnelling Underground Space Technol.*, vol. 52, pp. 157–167, Feb. 2016.

[16] S. Senin, M. Jaafar, and R. Hamid, "Locating underground water pipe leakages via interpretation of ground penetrating radar signals," *Int. J. Eng. Technol.*, vol. 8, no. 1.2, pp. 72–77, 2019.

[17] Kwipped. *Mala Easy Locator Pro Widerange HDR Ground Penetrating Radar*. Accessed: Sep. 1, 2020. [Online]. Available: https://www.kwipped.com/rentals/product/mal-easy-locator-pro-widerange-hdr-ground-penetrating-radar/15611

[18] GP Technical Equipment. *Mala Easy Locator HDR Pro Ground Penetrating Radar*. Accessed: Sep. 1, 2020. [Online]. Available: https://gp-technical.com/product/mala-easy-locator-hdr-pro-ground-penetrating-radar/

[19] Echologics. *What is Non-Invasive Leak Detection & How Does it Work?* Accessed: Sep. 1, 2020. [Online]. Available: https://www.echologics.com/resources/articles/non-invasive-detection/

[20] Y. A. Khulief, A. Khalifa, R. B. Mansour, and M. A. Habib, "Acoustic detection of leaks in water pipelines using measurements inside pipe," *J. Pipeline Syst. Eng. Pract.*, vol. 3, no. 2, pp. 47–54, May 2012.

[21] K. A. Papadopoulou, M. N. Shamout, B. Lennox, D. Mackay, A. R. Taylor, J. T. Turner, and X. Wang, "An evaluation of acoustic reflectometry for leakage and blockage detection," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 222, no. 6, pp. 959–966, Jun. 2008.

[22] M. Abdullahi and S. O. Oyadiji, "Simulation and experimental measurement of acoustic wave reflectometry for leak detection in pipes," *Proc. SPIE*, vol. 10972, Apr. 2019, Art. no. 109721S.

[23] P. M. Moubarak, P. Ben-Tzvi, and M. E. Zaghloul, "A self-calibrating mathematical model for the direct piezoelectric effect of a new MEMS tilt sensor," *IEEE Sensors J.*, vol. 12, no. 5, pp. 1033–1042, May 2012.

[24] D. Ozevin, "Geometry-based spatial acoustic source location for spaced structures," *Struct. Health Monitor.*, vol. 10, no. 5, pp. 503–510, Sep. 2011.

[25] D. Ozevin and J. Harding, "Novel leak localization in pressurized pipeline networks using acoustic emission and geometric connectivity," *Int. J. Pressure Vessels Piping*, vol. 92, pp. 63–69, Apr. 2012.

[26] V. A. Mandelshtam, "FDM: The filter diagonalization method for data processing in NMR experiments," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 38, no. 2, pp. 159–196, Mar. 2001.

[27] A. Lay-Ekuakille, G. Vendramin, and A. Trotta, "Spectral analysis of leak detection in a zigzag pipeline: A filter diagonalization method-based algorithm application," *Measurement*, vol. 42, no. 3, pp. 358–367, Apr. 2009.

[28] A. Lay-Ekuakille, G. Griffo, J. Kidiamboko, R. Velazquez, and J. C. Garcia, "Assessment indicators for expressing accuracy in sensing infrastructure health: Case study of leakage in water pipelines," in *Proc. 11th Int. Conf. Sens. Technol. (ICST)*, Dec. 2017, pp. 1–5.

[29] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Machine learning towards intelligent systems: Applications, challenges, and opportunities," *Artif. Intell. Rev.*, pp. 1–50, Jan. 2021. [Online]. Available: https://rdcu.be/cgn4K

[30] M. Injadat, "Optimized machine learning models towards intelligent systems," Ph.D. dissertation, Univ. Western Ontario, London, ON, Canada, Jul. 2020.

[31] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "E-learning: Challenges and research opportunities using machine learning & data analytics," *IEEE Access*, vol. 6, pp. 39117–39138, 2018.
[32] A. Moubayed, "Optimization modeling and machine learning techniques towards smarter systems and processes," Ph.D. dissertation, Univ. Western Ontario, London, ON, Canada Aug. 2018.
[33] Z. Qu, H. Feng, Z. Zeng, J. Zhuge, and S. Jin, "A SVM-based pipeline leakage detection and pre-warning system," *Measurement*, vol. 43, no. 4, pp. 513–519, May 2010.
[34] 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.
[35] S. Porwal, S. A. Akbar, and S. C. Jain, "Leakage detection and prediction of location in a smart water grid using SVM classification," in *Proc. Int. Conf. Energy, Commun., Data Anal. Soft Comput. (ICECDS)*, Aug. 2017, pp. 3288–3292.
[36] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (SVM) in LibSVM," *Int. J. Comput. Appl.*, vol. 128, no. 3, pp. 28–34, Oct. 2015.
[37] 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.
[38] R. J. Schalkoff, *Artificial Neural Networks*. New York, NY, USA: McGraw-Hill, 1997.
[39] M. Zadkarami, M. Shahbazian, and K. Salahshoor, "Pipeline leakage detection and isolation: An integrated approach of statistical and wavelet feature extraction with multi-layer perceptron neural network (MLPNN)," *J. Loss Prevention Process Industries*, vol. 43, pp. 479–487, Sep. 2016.
[40] S. Li, Y. Song, and G. Zhou, "Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition," *Measurement*, vol. 115, pp. 39–44, Feb. 2018.
[41] F. Fan, J. Xiong, M. Li, and G. Wang, "On interpretability of artificial neural networks: A survey," 2020, *arXiv:2001.02522*. [Online]. Available: http://arxiv.org/abs/2001.02522
[42] K. Fredenslund, "Computational complexity of neural networks," Internet Kasperfred, Copenhagen, Denmark, Tech. Rep., 2019. Accessed: Aug. 28, 2020. [Online]. Available: https://kasperfred.com/series/introduction-to-neural-networks/computational-complexity-of-neural-networks
[43] C. Ge, G. Wang, and H. Ye, "Analysis of the smallest detectable leakage flow rate of negative pressure wave-based leak detection systems for liquid pipelines," *Comput. Chem. Eng.*, vol. 32, no. 8, pp. 1669–1680, Aug. 2008.
[44] J. Li, Q. Zheng, Z. Qian, and X. Yang, "A novel location algorithm for pipeline leakage based on the attenuation of negative pressure wave," *Process Saf. Environ. Protection*, vol. 123, pp. 309–316, Mar. 2019.
[45] J. Wang, L. Zhao, T. Liu, Z. Li, T. Sun, and K. T. V. Grattan, "Novel negative pressure wave-based pipeline leak detection system using fiber Bragg grating-based pressure sensors," *J. Lightw. Technol.*, vol. 35, no. 16, pp. 3366–3373, Aug. 15, 2017.
[46] P. Ostapkowicz, "Leakage detection from liquid transmission pipelines using improved pressure wave technique," *Oper. Rel.*, vol. 16, no. 1, pp. 9–16, 2014.
[47] T. Chung, *Computational Fluid Dynamics*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[48] R. Ben-Mansour, M. A. Habib, A. Khalifa, K. Youcef-Toumi, and D. Chatzigeorgiou, "Computational fluid dynamic simulation of small leaks in water pipelines for direct leak pressure transduction," *Comput. Fluids*, vol. 57, pp. 110–123, Mar. 2012.
[49] M. Jujuly, P. Thodi, A. Rahman, and F. Khan, "Computational fluid dynamics modeling of subsea pipeline leaks in arctic conditions," in *Proc. Arctic Technol. Conf.*, 2016, pp. 1–6.
[50] R. Kruse, J. E. Gebhardt, and F. Klowon, *Foundations of Fuzzy Systems*. Hoboken, NJ, USA: Wiley, 1994.
[51] R. Mamlook and O. Al-Jayyousi, "Fuzzy sets analysis for leak detection in infrastructure systems: A proposed methodology," *Clean Technol. Environ. Policy*, vol. 6, no. 1, pp. 26–31, Dec. 2003.
[52] H. V. da Silva, C. K. Morooka, I. R. Guilherme, T. C. da Fonseca, and J. R. P. Mendes, "Leak detection in petroleum pipelines using a fuzzy system," *J. Petroleum Sci. Eng.*, vol. 49, nos. 3–4, pp. 223–238, Dec. 2005.
[53] A. Jalalkamali and M. Eftekhari, "Estimating water losses in water distribution networks using a hybrid of Ga and neuro-fuzzy models," *World Appl. Sci. J.*, vol. 18, no. 4, pp. 528–537, 2012.
[54] S. C. Ahn, Y. H. Kim, and W. H. Kwon, "Computational complexity of fuzzy controllers and its application to a real-time loop controller," *IFAC Proc. Volumes*, vol. 28, no. 5, pp. 441–447, May 1995.
[55] Y. H. Kim, S. C. Ahn, and W. H. Kwon, "Computational complexity of general fuzzy logic control and its simplification for a loop controller," *Fuzzy Sets Syst.*, vol. 111, no. 2, pp. 215–224, Apr. 2000.
[56] L. Satish and S. C. Vora, "Amplitude frequency response measurement: A simple technique," *IEEE Trans. Educ.*, vol. 53, no. 3, pp. 365–371, Aug. 2010.
[57] P. J. Lee, J. P. Vítkovský, M. F. Lambert, A. R. Simpson, and J. A. Liggett, "Frequency domain analysis for detecting pipeline leaks," *J. Hydraulic Eng.*, vol. 131, no. 7, pp. 596–604, Jul. 2005.
[58] P. J. Lee, J. P. Vítkovský, M. F. Lambert, A. R. Simpson, and J. A. Liggett, "Leak location using the pattern of the frequency response diagram in pipelines: A numerical study," *J. Sound Vib.*, vol. 284, nos. 3–5, pp. 1051–1073, Jun. 2005.
[59] D. E. Newland, "Harmonic wavelet analysis," *Proc. Roy. Soc. London A, Math. Phys. Sci.*, vol. 443, no. 1917, pp. 203–225, 1993.
[60] M. Ferrante and B. Brunone, "Pipe system diagnosis and leak detection by unsteady-state tests. 1. Harmonic analysis," *Adv. Water Resour.*, vol. 26, no. 1, pp. 95–105, Jan. 2003.
[61] M. Ferrante and B. Brunone, "Pipe system diagnosis and leak detection by unsteady-state tests. 2. Wavelet analysis," *Adv. Water Resour.*, vol. 26, no. 1, pp. 107–116, Jan. 2003.
[62] L. Eren, M. Unal, and M. J. Devaney, "Harmonic analysis via wavelet packet decomposition using special elliptic half-band filters," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2289–2293, Dec. 2007.
[63] B. K. S. Kumar, "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform," *Signal, Image Video Process.*, vol. 7, no. 6, pp. 1125–1143, Nov. 2013.
[64] 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, Apr. 2009.
[65] C. Verde, N. Visairo, and S. Gentil, "Two leaks isolation in a pipeline by transient response," *Adv. Water Resour.*, vol. 30, no. 8, pp. 1711–1721, Aug. 2007.
[66] H.-F. Duan, "Transient frequency response based leak detection in water supply pipeline systems with branched and looped junctions," *J. Hydroinformatics*, vol. 19, no. 1, pp. 17–30, Jan. 2017.
[67] B. P. L. Lau, S. H. Marakkalage, Y. Zhou, N. U. Hassan, C. Yuen, M. Zhang, and U.-X. Tan, "A survey of data fusion in smart city applications," 2019, *arXiv:1905.11933*. [Online]. Available: http://arxiv.org/abs/1905.11933
[68] W. Xu and J. Yu, "A novel approach to information fusion in multi-source datasets: A granular computing viewpoint," *Inf. Sci.*, vol. 378, pp. 410–423, Feb. 2017.
[69] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Tech. Rev.*, vol. 27, no. 4, pp. 293–307, 2010.
[70] R. Gupta, K. U. R. Laghari, and T. H. Falk, "Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization," *Neurocomputing*, vol. 174, pp. 875–884, Jan. 2016.
[71] P. Rashinkar and V. S. Krushnasamy, "An overview of data fusion techniques," in *Proc. Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA)*, Feb. 2017, pp. 694–697.
[72] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," *IEEE Sensors J.*, vol. 2, no. 2, pp. 107–119, Apr. 2002.
[73] B. V. Dasarathy, "Decision fusion strategies in multisensor environments," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1140–1154, Sep. 1991.
[74] H. Durrant-Whyte, M. Stevens, and E. Nettleton, "Data fusion in decentralised sensing networks," in *Proc. 4th Int. Conf. Inf. Fusion*, 2001, pp. 302–307.
[75] T. Zhao, J. Fu, Y. Wang, and B. Zhang, "An underwater measurement and control network centralized data fusion localization algorithm based on Chan-algorithm method," *J. Acoust. Soc. Amer.*, vol. 142, no. 4, p. 2732, 2017.
[76] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, "Decentralized data fusion with inverse covariance intersection," *Automatica*, vol. 79, pp. 35–41, May 2017.
[77] M. E. Campbell and N. R. Ahmed, "Distributed data fusion: Neighbors, rumors, and the art of collective knowledge," *IEEE Control Syst.*, vol. 36, no. 4, pp. 83–109, Aug. 2016.
[78] M. Guerriero, F. Wheeler, G. Koste, S. Dekate, and N. Choudhury, "Bayesian data fusion for pipeline leak detection," in *Proc. 19th Int. Conf. Inf. Fusion (FUSION)*, 2016, pp. 278–285.

[79] L. Dong, Z. Qiao, H. Wang, W. Yang, W. Zhao, K. Xu, G. Wang, L. Zhao, and H. Yan, ''The gas leak detection based on a wireless monitoring system,'' *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6240–6251, Dec. 2019.

[80] J. Liu, D. Zang, C. Liu, Y. Ma, and M. Fu, ''A leak detection method for oil pipeline based on Markov feature and two-stage decision scheme,'' *Measurement*, vol. 138, pp. 433–445, May 2019.

[81] R. Doshmanziari, H. Khaloozadeh, and A. Nikoofard, ''Gas pipeline leakage detection based on sensor fusion under model-based fault detection framework,'' *J. Petroleum Sci. Eng.*, vol. 184, Jan. 2020, Art. no. 106581.

[82] V. Nkemeni, F. Mieyeville, and P. Tsafack, ''A distributed computing solution based on distributed Kalman filter for leak detection in WSN-based water pipeline monitoring,'' *Sensors*, vol. 20, no. 18, p. 5204, Sep. 2020.

[83] Hilscher. *Industrial Raspberry Pi IoT Gateway netPI RTE 3*. Accessed: Sep. 1, 2020. [Online]. Available: https://www.hilscher.com/products/product-groups/industrial-internet-industry-40/netiot-edge/niot-e-npi3-51-en-reecs/

[84] Q. Qin, K. Poularakis, K. K. Leung, and L. Tassiulas, ''Line-speed and scalable intrusion detection at the network edge via federated learning,'' in *Proc. IFIP Netw. Conf. (Networking)*, 2020, pp. 352–360.

[85] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, ''Federated learning for data privacy preservation in vehicular cyber-physical systems,'' *IEEE Netw.*, vol. 34, no. 3, pp. 50–56, May 2020.

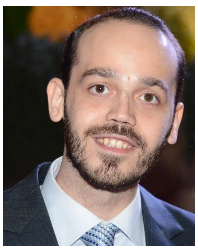
[86] Z. Li, J. Chen, J. Zhang, X. Cheng, and B. Chen, ''Detecting advanced persistent threat in edge computing via federated learning,'' in *Proc. Int. Conf. Secur. Privacy Digit. Economy*. Singapore: Springer, 2020, pp. 518–532.

[87] M. A. Rahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, ''Secure and provenance enhanced Internet of health things framework: A blockchain managed federated learning approach,'' *IEEE Access*, vol. 8, pp. 205071–205087, 2020.

[88] R.-H. Hsu, Y.-C. Wang, C.-I. Fan, B. Sun, T. Ban, T. Takahashi, T.-W. Wu, and S.-W. Kao, ''A privacy-preserving federated learning system for Android malware detection based on edge computing,'' in *Proc. 15th Asia Joint Conf. Inf. Secur. (AsiaJCIS)*, Aug. 2020, pp. 128–136.

[89] A. Ghosh, J. Hong, D. Yin, and K. Ramchandran, ''Robust federated learning in a heterogeneous environment,'' 2019, *arXiv:1906.06629*. [Online]. Available: http://arxiv.org/abs/1906.06629

[90] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, ''Federated learning with non-IID data,'' 2018, *arXiv:1806.00582*. [Online]. Available: http://arxiv.org/abs/1806.00582

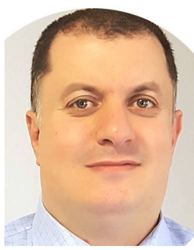


ABDALLAH MOUBAYED (Member, IEEE) received the B.E. degree in electrical engineering from Lebanese American University, Beirut, Lebanon, in 2012, the M.Sc. degree in electrical engineering from the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, in 2014, and the Ph.D. degree in electrical & computer engineering from the Western University, in August 2018. He is currently a Postdoctoral Associate with the Optimized Computing and Communications (OC2) Lab, Western University. His research interests include wireless communication, resource allocation, wireless network virtualization, performance & optimization modeling, machine learning & data analytics, computer network security, cloud computing, and e-learning.

MOHAMED SHARIF (Student Member, IEEE) is currently pursuing the bachelor's degree in computer engineering with Western University, where he is also pursuing the accelerated master's degree in electrical & computer engineering.

MARCO LUCCINI (Member, IEEE) received the Ph.D. degree in information technology from the Politecnico di Milan and also from Western University, in 2013. He is currently a R&D Process Engineering Researcher. He has worked extensively in the area of process estimation and prediction, digitalization, and other areas including anomaly detection and water leak detection.

SERGUEI PRIMAK (Member, IEEE) received the M.S.E.E. degree from the St. Petersburg University of Telecommunications, St. Petersburg, Russia, in 1991, and the Ph.D. degree in electrical engineering from the Ben-Gurion University of the Negev, Be'er Sheva, Israel, in 1996. He is currently an Associate Professor with the ECE Department, Western University, London, ON, Canada. His current research interests include field of ultrawideband radar applications, random signal generations, the modeling of wave propagation in random media, information distribution over complex structures, MIMO communications, and related disciplines.



ABDALLAH SHAMI (Senior Member, IEEE) is currently a Professor with the ECE Department, Western University, London, ON, Canada, where he is also the Director of the Optimized Computing and Communications Laboratory.

Dr. Shami has chaired key symposia for IEEE GLOBECOM, IEEE ICC, IEEE ICNC, and ICCIT. He is currently an Associate Editor of the IEEE Transactions on Mobile Computing, the IEEE Network, and the IEEE Communications Tutorials and Survey.