Water Leak Detection and Classification using Multiple Sensors

Daniele Ugo Leonzio*, Paolo Bestagini*, Marco Marcon*, Gian Paolo Quarta† and Stefano Tubaro*
*Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano - Milan, Italy
†Onyax S.r.l. - Vigevano, Italy

***Abstract*—Water leakage represents an important concern in managing Water Distribution Networks (WDNs) as more than sixty countries are still facing high water stress risk. These leaks constitute major financial loss, soil contamination and other environmental hazards, all contributing to further water scarcity. Fast-expanding water supply networks need the development of better leak detection technologies. Current technologies such as real-time monitoring of the WDNs and Machine Learning (ML) allow us to limit these losses by developing data-driven methods for leak detection. This work aims at contributing to this issue through a study of the capacity of ML methods to detect and classify types of leakage in WDNs by using acoustic measurements. For this purpose, we used a dataset generated by a laboratory-scale WDN including three types of sensors, namely accelerometers, hydrophones, dynamic pressure sensors, no-leak conditions, four leak types, various background conditions, and two network topologies. The results achieved both for the detection and classification problem are successful, suggesting that the proposed solution can be adopted to solve both the problems. Moreover in this work we have faced a problem not yet investigated in the literature, which is the different leak types classification.**

***Index Terms*—Anomaly Detection, Water Leak Classification, Water Leak Detection**

I. Introduction

The issue of water scarcity has become a growing concern in recent years, as demand has increased while infrastructure has aged and environmental degradation has taken its toll. Among the many causes of urban water loss, leaks or pipe breaks account for a significant portion, amounting to 30% of the total water supply [1]. This is a major contributor to the global water crisis, as it represents the largest portion of the estimated 126 billion cubic meters of Non Revenue Water (NRW) lost annually [2]. In addition to causing water supply disruptions during leak repairs, false alarms and inaccurate leak location assessments can result in substantial repair costs. Therefore, identifying leaks in a timely and accurate manner is crucial to improving water distribution and supply efficiency. The economic and environmental benefits of quickly and reliably identifying leaks in a pipe network are significant, as leaks increase operating costs, reduce service levels, and cause water pollution [3], [4]. As such, leak detection is not just an economic problem, but also an environmental and safety concern [5].

With the evolution of electronic and information technologies, various sensing systems for water leak detection have emerged. Indeed, with the development of Internet of Things (IoT) solutions, sensing technologies have become critical [6].

Leak detection in water distribution systems is an ongoing research area, and several solutions have been proposed in recent years. Two strategies are commonly used: developing methods to solve the leak detection and localization problem for a single pipe [7]–[9]; or creating algorithms that solve the problem for an entire water distribution network [10]–[12].

In case of leakage, the water escaping from the pipe generates a particular noise, therefore, acoustic sensors can be used to capture the signal generated.

Vibroacoustic methods have shown satisfactory results, however usually more effective for metal pipes than plastic ones. This is due to the fact that leak noise propagating in plastic pipes is attenuated way more than in metal pipelines [13]. There is also more uncertainty in the noise propagation speed in Polyvinyl Chloride (PVC) pipes [6]. This is why the problem of detecting leaks in plastic pipelines using vibroacoustic methods deserves special attention. Hydrophones, which are acoustic devices, have the potential for both short and long-term leak monitoring and control [14], and are of particular interest for their ability to capture in-pipe acoustic signatures of leak signals, which is significantly different from out-of-pipe technologies like accelerometers [15].

Hydrophones have shown to be particularly effective in producing high-resolution correlation in the case of plastic pipes with a large diameter and a high attenuation of sound waves, making them more successful than other technologies [16], [17]. Experimental studies have demonstrated that hydrophones produce similar or higher accuracy results compared to accelerometers, pressure sensors, or ground microphones [18]. Acoustic methods provide high-resolution data for the detection

This work was supported by the Italian Ministry of University and Research (MUR) and the European Union (EU) under the PON/REACT project.

ISBN 978-3-903176-63-8 ©2024 IFIP

of small leaks, as noted by Cody et al. [8]. However, using hydrophones to cover the entire network can be time-consuming, costly, and ineffective compared to other hardware technologies [19].

In this paper we investigate the problem of leak detection and leak classification in water pipes through different type of sensors, namely accelerometers, dynamic pressure sensor and hydrophone. The first task will be a binary classification task that consists in training ML models to distinguish between leak and no-leak signals. The second experiment will be a multi-class classification task where the models will be trained to recognize the four different leak types as well as the no-leak condition. For each experiment we implement also a feature selection scheme in order to select the best feature for each type of sensor.

The results achieved demonstrate that classical feature can be used to distinguish both leak from no-leak signals and differentiate the leak types. In addition this is one of the first work that investigate the problem of different leak types classification.

The paper is organized as follows. Section II presents the background about water leak detection solutions. In Section III we formalize the problem and in Section IV we describe the proposed method to solve it. Section V describes the dataset used in this work and how we setup the parameters of the proposed model. In Section VI we present the results obtained in this study. In Section VII we investigate the effect of downsampling of the balanced accuracy for accelerometer and dynamic pressure sensors. paoSection VIII concludes the paper.

II. Background

The advancement of electronic and information technologies has led to the emergence of numerous sensing systems for water leak detection. The rise of IoT solutions has made sensing technologies increasingly important [6]. Sensors are devices that detect or measure a physical property and convert it into a form that can be easily interpreted by humans or machines.

One way to detect leaks is to investigate the inherent properties of measured signals. For instance, Martini et al. [20] successfully employed the auto-correlation method for detecting and localizing water leaks in small-diameter plastic pipes using vibroacoustic signals. They tested a technique based on the normalized auto-correlation function's shape, during measurements with a High Density Polyethylene (HDPE) pipe. The researchers artificially generated leaks and studied both active leak and non-leaking conditions. They measured the signals with a hydrophone and two accelerometers in a full-scale test facility and generated two datasets at different times of the year (winter and spring) to account for seasonal fluctuations. The experiment yielded good results with the accelerometers when the sensing direction was parallel to the pipe axis (i.e., analysis of longitudinal vibration signals). However, monitoring of radial vibrations did not allow for accurate leak detection. Overall, the hydrophone performed better as it showed higher sensitivity to leaks.

The study by Scussel et al. [13] aimed to estimate the spectrum of leak noise based on data obtained from acoustic correlators. Both accelerometers and hydrophones were considered in their investigation. The proposed method enabled the prediction of the spectral shape of the leak noise for both sensors and was tested on datasets obtained from three test sites with varying conditions. The authors estimated the Frequency Response Function (FRF) between the acoustic pressure of the leak noise and the acoustic pressure at the sensors. The Power Spectral Density (PSD) of the leak noise was subsequently evaluated from the FRFs, which were then used to estimate the spectrum of the leak noise [13]. Results showed that when hydrophones were used, both the magnitude and shape of the spectrum could be estimated, whereas accelerometers only allowed the estimation of the shape of the spectrum.

Shen et al. [7] undertook the development and comparison of three tree models, specifically a decision tree, a random forest, and an AdaBoost model, to facilitate leak detection. To identify relevant features for this task, parameters such as the main frequency, Spectral Flatness (SF), Spectral Roll-Off (SRO), and one-dimensional Mel-Frequency Cepstral Coefficients (MFCC) were considered. The authors collected tens of thousands of sets of on-site leak detection signals for their study [7].

In their research, Nam et al. [21] utilized a Convolutional Neural Network (CNN) to detect water leaks using recorded leak sounds and the hold-out method. They collected acoustic datasets of leak sounds from sensors placed at ten different sites and found that 19 out of 20 cases achieved an accuracy above 70%, with 15 of them above 80% [21]. However, the hold-out method used in their study validated the method using data that was different from the training data, indicating the need for further research to utilize this technique for real-world applications.

According to observations, the use of accelerometers for leak detection is not effective when the leaks are located far away from the sensors. Additionally, similar to geophones, accelerometers face challenges in detecting low-frequency leak noise, particularly when there is a significant attenuation factor [22]. Another limitation in studying the spectral properties of leak noise is the difficulty of placing sensors at the actual leak location inside the pipeline [13]. While hydrophones are efficient for detecting leaks in plastic pipes, their deployment cost is high as they are invasive sensors.

III. Problem Formulation

In our work we propose a method for water leak detection and classification by analyzing the signals acquired from different sensors distributed in a WDN.

Let us consider the signal $\mathbf{x}(n)$ which represent the signal acquired with a sensor, where $n = 0, 1, ..., N-1$. The objective of this study is develop a model capable of detecting the presence of a leak in the signal $\mathbf{x}$ for the detection problem, and capable of classify the leak type in case of the classification problem.

Formally, given a generic input acoustic signal $\mathbf{x}(n)$, solving the detection problem is equivalent to associating a label $c \in [0, 1]$, where 0 represents a signal in no-leak condition, while 1 means a leak signal.

Instead, solving the multi-class classification problem is equivalent to associating a label $c \in [0, 1, 2, 3, 4]$ where 0 represents a signal in no-leak condition, while 1, 2, 3, and 4, represent a signal characterized by the presence of a leak of type 1, 2, 3, and 4, respectively.

IV. Proposed Method

In this section, we discuss the proposed approach by describing the processing pipeline which consists of three steps, namely preprocessing, feature extraction, and detection or classification depending on the task we want to tackle. Figure 1 describes our proposed method pipeline.

A. Preprocessing

The first block of the proposed pipeline is the preprocessing stage. In the preprocessing stage we need to prepare the input signal $\mathbf{x}$ for the feature extraction. To do so we divide the signal $\mathbf{x}$ into N overlapping window, with an overlap value of O. Formally, a window is defined as:

$$\mathbf{w} = \mathbf{x}[m, m+1, ..., m+L-1], \tag{1}$$

where $\mathbf{x}$ is the input signal, m is the starting sample of the window w, and L is the length of the window.

The windows obtained as output of the preprocessing stage will be the input for the features extraction process.

B. Feature Extraction

The second block of our method pipeline is the feature extraction. It takes the windows $\mathbf{w}$ obtained from the preprocessing stage as input and gives feature vectors $\mathbf{f}$ as output. The feature vector $\mathbf{f}$ has the dimensionality $(1, K-1)$, where K is the number of feature selected. Formally, this process can be defined as follows:

$$\mathbf{f} = F(\mathbf{w}), \tag{2}$$

where F represents the feature extractor that takes signal windows $\mathbf{w}$ as input.

C. Detection or Classification

The last step of the method is the detection or classification step based on the problem we want to resolve, so binary or multi-class respectively. In this phase we used a ML algorithm to predict a label $\hat{c}$ for the feature vector in input.

$$\hat{c} = \mathcal{ML}(\mathbf{f}), \tag{3}$$

where f represents the selected features used by the model and $\mathcal{ML}$ is the ML classifier.

1) Detection: For the detection part, which means solve the binary problem, the predicted label $\hat{c}$ can have only two values, to distinguish leak from no-leak signals. Formally,

$$\hat{c} = \begin{cases} 1, & \text{if Leak} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

2) Classification: In case of the classification, instead, the predicted label $\hat{c}$ can have five values, four different leak types plus the no-leak case. Formally,

$$\hat{c} = \begin{cases} 0, & \text{No Leak} \\ 1, & \text{Leak type 1} \\ 2, & \text{Leak type 2} \\ 3, & \text{Leak type 3} \\ 4, & \text{Leak type 4} \end{cases} \tag{5}$$

V. Experimental setup

In this section we report details about the dataset and the parameters we chose in our pipeline, from the preprocessing parameters to the ML method selection.

A. Dataset

To develop our method we have used the dataset from the work [23]. The dataset proposed by Mohsen et al. [23] presents two 218 sensory measurements generated via a laboratory-scale WDN.

Measurements were made using three different types of sensors, namely accelerometers, hydrophone and dynamic pressure sensors. Different leak types were studied: Orifice Leak (OL), Longitudinal Crack (LC), Circumferential Crack (CC), Gasket Leak (GL), and No-Leak (NL) condition, under two different network topologies, Looped or Branched topology.

Finally, six background conditions with various noise and demand conditions were considered for this data acquisition. The sampling frequency for the accelerometers and the dynamic pressure sensors was 25.6 kHz, and 8 kHz for the hydrophones. The WDN testbed consisted of 152.4 mm diameter PVC pipes and 47 m total pipe length.

Fig. 1. Proposed method pipeline for the leak detection problem.

B. Parameters selection

The parameters selection is referred to the preprocessing stage.

Signals from different sensors have different length in [23], but to have a common value in all the experiment we decided to crop the signals to a total length T of $30s$. We set the window length to 0.5s, so L is equal to 4000 samples. The overlap 0 is equal to 75%.

C. Feature selection

Various spectral features were considered for training our models, namely the Spectral Centroid (SC), the Spectral Bandwidth (SB), the Spectral Roll-Off (SRO), the MFCC and the Spectral Flatness (SF). Time-domain features such as the Root Mean Square (RMS) and the Zero-Crossing Rate (ZCR) were considered as well. For each of these features we considered the first two statistical moments, namely mean, standard deviation. The final feature vector has therefore the dimensionality $(1, 14)$.

To select the best feature for the detection problem we used a Sequential Feature Selector (SFS) method, in particular we used a forward SFS pipeline.

We ran the SFS algorithm for each sensor type and for each topology, both for leak detection and classification.

At the end the most relevant feature for both the problems and all sensors turned to be the RMS, ZCR, SC and the MFCC.

D. Model selection

To select the best model for each sensor and topology we ran a variety of classical machine learning algorithms, including Linear Regression (LR), Logistic Regression (LogR), Random Forest (RF), Gradient Boosting (GB), and Support Vector Machines (SVM). To choose the best model we considered the balanced accuracy to compare the various models.

In Table I and Table II we show the ML model selected for each sensor and topology, considering the detection and classification tasks respectively.

TABLE I
MODEL SELECTION RESULTS FOR THE DETECTION PROBLEM

Sensor type	ML Model	
	Branched	Looped
Hydrophone	LightGBM	LightGBM
Accelerometer	XGBoost	LightGBM
Dynamic Pressure	ExtraTrees	ExtraTrees

TABLE II
MODEL SELECTION RESULTS FOR THE CLASSIFICATION PROBLEM

Sensor type	ML Model	
	Branched	Looped
Hydrophone	ExtraTrees	LightGBM
Accelerometer	ExtraTrees	Random Forest
Dynamic Pressure	XGBoost	LightGBM

VI. RESULTS

This section presents the results achieved by our water leak detection and classification method. The results are divided in two subsections: Detection and Classification. The detection part is referred to the results obtained solving the binary problem, which is distinguish leak and no leak signals. The classification section, instead, presents the results obtained solving the multi-class problem, which is classify the leak type based on the four classes that are provided in the dataset.

A. Detection

A said above in this section we show the results of the detection problem. We have computed the results for each sensor and for each topology, following the feature and model selected as explained in Section V-C and Section V-D. In Table III we reported the results achieved for each case in terms of balanced accuracy. In Figure 2 there is an example of confusion matrix results obtained solving the detection problem for the accelerometer sensor in the looped scenario.

TABLE III
DETECTION RESULTS

Sensor type	Balanced Accuracy	
	Branched	Looped
Hydrophone	0.900	0.970
Accelerometer	0.987	0.997
Dynamic Pressure	0.931	0.956

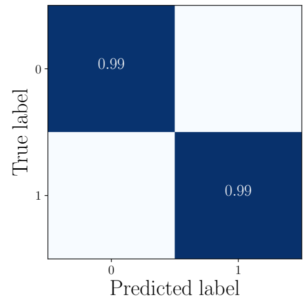

Fig. 2. Example of confusion matrix detection results obtained for accelerometer signals for looped configuration.

B. Classification

As done for the detection part, we report in this section the results of the classification problem, so differentiate the multiple type of leak that can occur in a water pipe. Table IV presents the balanced accuracy reached for each sensor and network topology. In Figure 2 there is an example of confusion matrix results obtained solving the classification problem for the hydrophone sensor in the looped scenario.

TABLE IV
CLASSIFICATION RESULTS

Sensor type	**Balanced Accuracy**	
	Branched	**Looped**
Hydrophone	0.894	0.910
Accelerometer	0.984	0.934
Dynamic Pressure	0.895	0.865

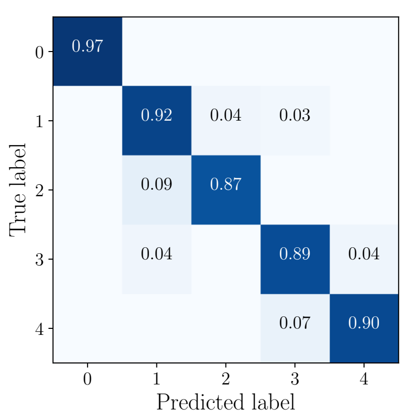

Fig. 3. Example of confusion matrix classification results obtained for hydrophone signals for looped configuration.

C. Comparison against Baseline

In order to compare our proposed method with the latest state-of-the-art approach, we conducted tests using the methodology presented in Shen et al.'s paper [7] along with the dataset proposed by Mohsen et al. [23]. Shen et al. employed an ensemble of features, namely dominant frequency, spectral roll-off, spectral flatness, and 1-D MFCC, in combination with the AdaBoost algorithm to address the leak detection problem. Since, their work focused on signals from geophones, we utilized hydrophone data for this comparison.

Although the method performed reasonably well, in particular for the looped topology, in this scenario; it exhibited slightly inferior performance compared to the approach proposed in our paper. Additionally, we introduced a feature selection step to aid in identifying the most effective features for each problem. Moreover we added in our method the feature selection step, in order to choose the best feature for each scenario, whereas this approach is missing in [7].

TABLE V
BASELINE COMPARISON

Method	**Balanced Accuracy**	
	Branched	**Looped**
Baseline [7]	0.793	0.927
Proposed	0.900	0.970

VII. DOWNSAMPLING TEST

As last analysis in this section we present some experiment on the accuracy results after different level of downsampling of the accelerometers and dynamic pressure sensor data. The purpose was to assess the system's performance when downsampling aggressively to achieve sampling rates comparable to commercially available sensors commonly used on a large scale. The sensors utilized to acquire the data in the dataset [23] had a sampling rate of 25.6 KHz, which is considered too high for commercial low-power IoT devices. As such, we sought to investigate the system's performance under lower sampling rates, more suitable for these devices that typically operate around 1 KHz sampling rate.

Furthermore, it is worth noting that consistent with the physics of the phenomenon being measured, the fundamental frequencies for detection are below 500 Hz.

Figure 4 illustrates the balanced accuracy achieved at different downsampling levels, for each sensor and each topology. The x-axis represents the sampling rate, while the y-axis represents the balanced accuracy value.

The results clearly indicate that even at highly reduced sampling rates, the system maintains a high level of accuracy. Notably, the accuracy remains consistently above 85%, confirming that this proposed method can be adopted also in case of lower sampling rate, that is typical of IoT applications for water monitoring.

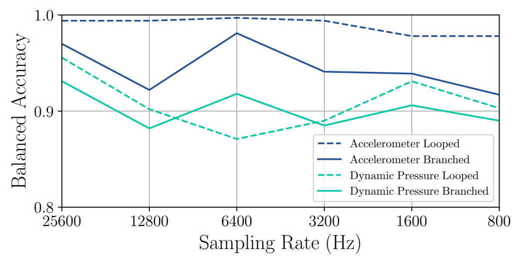

Fig. 4. Effect of downsampling over balanced accuracy metric for accelerometer and dynamic pressure sensor

VIII. Conclusion

We have introduced a novel approach in this study for the detection and classification of leaks in water pipelines. Our methodology focuses on two primary objectives:

- Identifying the presence of a leak within a pipe.
- Categorizing the type of leak.

To accomplish these goals, we employ a feature ensemble technique and a ML algorithm that has been specifically trained to handle either the detection or classification task.

To evaluate our proposed method, we utilize a dataset suggested in the publication by Aghashani et al. in [23]. This dataset comprises multiple sensor readings from pipes with and without leaks, considering two distinct network scenarios: branched and looped networks.

Our proposed method has demonstrated promising outcomes in both leak detection and classification. Particularly, the pipeline we describe in this study exhibits exceptional speed and attains a high level of accuracy. Notably, this work stands out as one of the initial investigations into the classification of different types of leaks.

For the detection problem, we achieve a balanced accuracy exceeding 90% for each sensor and network topology. Regarding classification, our balanced accuracy surpasses 85% in all tested cases. These results emphasize that the proposed method offers a valid solution for both detection and classification, outperforming the current state-of-the-art baseline.

As a future direction, our focus will be on exploring sensor fusion techniques and domain adaptation approaches to assess the generalizability of the proposed method.

References

[1] S. Mounce, J. Boxall, and J. Machell, "Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows," *ASCE Journal of Water Resources Planning and Management*, pp. 309–318, 2010.

[2] R. Liemberger and A. Wyatt, "Quantifying the global non-revenue water problem," *Water Supply*, vol. 19, p. 831–837, 2018.

[3] S. Sarkamaryan, A. Haghighi, and A. Adib, "Leakage detection and calibration of pipe networks by the inverse transient analysis modified by Gaussian functions for leakage simulation," *Journal of Water Supply: Research and Technology-Aqua*, pp. 404–413, 2018.

[4] G. Guo, X. Yu *et al.*, "Leakage Detection in Water Distribution Systems Based on Time–Frequency Convolutional Neural Network," *ASCE Journal of Water Resources Planning and Management*, pp. 1–11, 2021.

[5] A. Colombo and B. Karney, "Energy and costs of leaky pipes: Toward comprehensive picture," *ASCE Journal of Water Resources Planning and Management*, pp. 441–450, 2002.

[6] N. A. Mohd Yussof and H. W. Ho, "Review of water leak detection methods in smart building applications," *Buildings*, vol. 12, 2022.

[7] Y. Shen and W. Cheng, "A tree-based machine learning method for pipeline leakage detection," *Water*, vol. 14, no. 18, 2022.

[8] R. A. Cody, P. Dey, and S. Narasimhan, "Linear prediction for leak detection in water distribution networks," *Journal of Pipeline Systems Engineering and Practice*, vol. 11, p. 04019043, 2020.

[9] D. U. Leonzio, P. Bestagini *et al.*, "Water Leak Detection via Domain Adaptation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

[10] A. Soldevila, G. Boracchi *et al.*, "Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models," *Neural Computing and Applications*, pp. 4759–4779, 2022.

[11] D. U. Leonzio, P. Bestagini *et al.*, "Water Leak Detection and Localization using Convolutional Autoencoders," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[12] ——, "Robust Water Leak Detection and Localization with Graph Signal Processing," in *IEEE Industrial Electronics, Control, and Instrumentation Conference (IECON)*, 2023.

[13] O. Scussel, M. Brennan *et al.*, "Estimating the spectrum of leak noise in buried plastic water distribution pipes using acoustic or vibration measurements remote from the leak," *Mechanical Systems and Signal Processing*, vol. 147, 2021.

[14] J. Xu, K. T.-C. Chai *et al.*, "Low-cost, tiny-sized mems hydrophone sensor for water pipeline leak detection," *IEEE Transactions on Industrial Electronics*, vol. 66, pp. 6374–6382, 2019.

[15] K. Marmarokopos, D. Doukakis *et al.*, "Leak detection in plastic water supply pipes with a high signal-to-noise ratio accelerometer," *Measurement and Control*, vol. 51, pp. 27–37, 2018.

[16] Y. Gao, M. J. Brennan *et al.*, "Improving the shape of the cross-correlation function for leak detection in a plastic water distribution pipe using acoustic signals," *Applied Acoustics*, vol. 127, pp. 24–33, 2017.

[17] O. Hunaidi and W. T. Chu, "Acoustical characteristics of leak signals in plastic water distribution pipes," *Applied Acoustics*, vol. 58, pp. 235–254, 1999.

[18] Y. Gao, M. Brennan *et al.*, "On the selection of acoustic/vibration sensors for leak detection in plastic water pipes," *Journal of Sound and Vibration*, vol. 283, pp. 927–941, 2005.

[19] R. Li, H. Huang *et al.*, "A review of methods for burst/leakage detection and location in water distribution systems," *Water Supply*, vol. 15, p. 429–441, 2014.

[20] A. Martini, A. Rivola, and M. Troncossi, "Autocorrelation analysis of vibro-acoustic signals measured in a test field for water leak detection," *Applied Sciences*, vol. 8, 2018.

[21] Y. W. Nam, Y. Arai *et al.*, "Water leak detection based on convolutional neural network using actual leak sounds and the hold-out method," *Water Supply*, vol. 21, no. 7, pp. 3477–3485, 2021.

[22] F. Almeida, M. Brennan *et al.*, "On the acoustic filtering of the pipe and sensor in a buried plastic water pipe and its effect on leak detection: An experimental investigation," *Sensors*, vol. 14, pp. 5595–610, 2014.

[23] M. Aghashahi, L. Sela, and M. K. Banks, "Benchmarking dataset for leak detection and localization in water distribution systems," *Data in Brief*, vol. 48, 2023.