Water Research X 29 (2025) 100333

Contents lists available at ScienceDirect

Water Research X

journal homepage: www.sciencedirect.com/journal/water-research-x

Leak detection in water supply networks using two-stage temporal segmentation and incremental learning for non-stationary acoustic signals

Xingke Ma [a], Yipeng Wu [a,*], Guancheng Guo [a], Shuming Liu [a,*], Yuexia Xu [c], Jingjing Fan [b], Hongbin Wang [c], Liren Xu [b]

[a] School of Environment, Tsinghua University, 100084, Beijing, PR China
[b] Shanghai Lingang Water & Wastewater Development Co., Ltd., 201306, Shanghai, PR China
[c] Zhengzhou Water Investment Holding Co., Ltd., 450007, Zhengzhou, PR China

ARTICLE INFO

Keywords:
Water supply networks
Acoustical leak detection
Non-stationary signal
Convolutional neural network
Incremental learning

ABSTRACT

Acoustic detection is a primary method for identifying leaks in urban water supply networks. However, acoustic signals within pipelines are highly susceptible to dynamic interference noise. This complicates the differentiation between leak and non-leak signals. To address this challenge, this paper presents a temporal segmentation-based approach for processing acoustic signals. Specifically, the two-stage temporal segmentation approach, which applies long-term segments to isolate non-stationary characteristics and short-term segments for capturing quasi-stationary features in acoustic signals, is introduced. We then applied the CNN model to recognize the Mel spectrogram features of the two-stage segmented signals and compared its performance with other models. Results indicate that this approach enhances both the accuracy and stability of leak detection, with the model achieving an average detection accuracy of 95 %. Moreover, the model is designed as an adaptive and continuous learning model, integrating its detection outcomes and newly labeled data segments into its training dataset. In practical applications, this continuous learning capability enables the model to improve its detection efficacy over time as data volume expands.

1. Introduction

The urban water supply networks (WSNs) form a vital component of modern urban infrastructure, yet they remain vulnerable to leaks, resulting in substantial water wastage and economic losses (Ahopelto and Vahala, 2020). Consequently, efficient and accurate leak detection is essential for maintaining the integrity and sustainability of these networks. Acoustic detection has emerged as a key technology for identifying leaks because it can capture the distinctive acoustic signals generated when water escapes through pipeline defects (Adegboye et al., 2019; Datta and Sarkar, 2016). Machine learning-based methods have garnered substantial attention and become the focus of extensive research to streamline the traditionally time-consuming process of manual signal analysis. These approaches offer the ability of automatically distinguishing leak signals from large volumes of monitoring data, enhancing both the efficiency and accuracy of leak detection processes (Bae et al., 2018).

Machine learning techniques for leak detection in WSNs using acoustic signals can be broadly categorized into two approaches, namely traditional machine learning methods with manual feature extraction and deep learning methods with automatic feature extraction (Wu et al., 2024a). Traditional machine learning methods, such as Support Vector Machines (SVM) (Wang et al., 2022), K-Nearest Neighbors (KNN) (Quy et al., 2019; Ravichandran et al., 2021; Yussif et al., 2023), and Random Forest (Guo et al., 2020), often rely on manually designed features (Bykerk and Miro, 2022a). In the context of acoustic signals, feature extraction requires identifying signal characteristics such as frequency and amplitude (Martini et al., 2018). Deep learning models, such as Convolutional Neural Networks (CNN) (Ahmad et al., 2023; Guo et al., 2021) and Recurrent Neural Networks (RNN) (Ramezani et al., 2020), can learn features directly from raw or transformed acoustic signals. These models minimize the need for manual feature engineering, allowing them to automatically detect complex signal patterns that may indicate leaks (Gunatilake and Miro, 2024; Zhang et al., 2023a). Moreover, a feature fusion method has been developed to further enhance performance by integrating expert knowledge with deep

* Corresponding authors.
E-mail addresses: wu_yp2021@mail.tsinghua.edu.cn (Y. Wu), shumingliu@tsinghua.edu.cn (S. Liu).

https://doi.org/10.1016/j.wroa.2025.100333
Received 12 November 2024; Received in revised form 14 February 2025; Accepted 12 March 2025
Available online 12 March 2025

learning (Zhang et al., 2023b). By leveraging both human expertise and deep learning, this hybrid method enhances the model's sensitivity to complex leak patterns, demonstrating strong practical effectiveness and robustness (Wu et al., 2023).

In previous studies, feature extraction driven by manual or deep learning methods has typically been applied to the entire signal or through fixed-duration segmentation (Fares et al., 2023; Islam et al., 2022). While these approaches are feasible, they often fail to account for the non-stationary nature of acoustic signals collected from real-world WSNs (Peng et al., 2024a). Non-stationary signals are those whose statistical properties, such as mean and variance, vary over time (Jiang et al., 2021). Considering that the duration of one collected acoustic signal is limited (e.g., 5 s), the non-stationarity within this limited time frame is critical (Yang et al., 2021). Ignoring these temporal variations can result in inaccuracies when distinguishing leak signals, as fixed-duration segmentation or whole-signal analysis may fail to capture essential time-varying characteristics (Peng et al., 2024a). Moreover, data has been a primary concern in research on recognition using deep learning methods (Cody et al., 2020; Liu et al., 2024). While techniques such as multiple slicing or data augmentation can be employed to generate additional training samples (Wu et al., 2024b), the new data generated from original datasets often lacks diversity. This limitation challenges recognition performance when addressing the diverse data sources encountered in real-world application environments.

The study employs the CNN model to address the challenge of recognizing non-stationary signals in leak detection. By utilizing a two-stage temporal segmentation approach, the model applies long temporal segmentation to isolate the non-stationary characteristics of signals, and short temporal segmentation captures the quasi-stationary features within acoustic signals. Ultimately, a multi-segment joint recognition technique assigns classification labels to each acoustic signal, allowing the model to adapt more effectively to the complexities encountered in real-world applications. Furthermore, by utilizing incremental learning methods, additional labeled data are introduced to increase diversity. The entire model structure is dynamic, continuously enhancing its ability to discern non-stationary signals throughout the application process.

This study proposes a novel leak detection framework based on temporal segmentation and Mel spectrogram transformation, combined with the CNN deep learning model. This article is organized into four main sections to comprehensively elaborate on the proposed research: methodology, experiment, results and discussion, conclusions. The main contributions of this research are as follows. (1) This study bridges advanced deep learning techniques with the challenges of analyzing non-stationary acoustic signals. (2) A novel two-stage segmentation approach effectively reduces fluctuating noise and enhances critical signal features, improving detection accuracy. (3) Mel-spectrogram transformation emphasizes low-frequency components, where the most relevant information for leak detection resides. (4) The proposed model employs an adaptive incremental learning strategy, enabling continuous integration of filtered, newly collected data to optimize detection performance in real-world environments.

2. Methodology

2.1. Overview of the method

The dataset employed comprises two components: an ideal dataset, collected in a controlled quiet environment, and a dataset with interference, gathered from actual operating conditions. The ideal dataset was curated by professional technicians, distinguished by its accurate labeling and minimal interference. In contrast, the dataset with interference was obtained from a monitoring platform, characterized by a significant level of random interference. The study was based on these two types of datasets, and the research framework was outlined as follows: it centered on data processing, model construction, and result output to complete the task of recognizing and classifying acoustic signals. In the data processing module, activities include data cleaning, temporal segmentation, and data transformation. In the model construction section, convolution the operation was applied to the processed data, with the recognition target being the data segments. In the result output section, the recognition results of long-term segments were integrated to define the recognition category of that data.

2.2. Temporal segmentation

The collected acoustic signals underwent cleaning, which included trend removal and filtering (Luo and Johnston, 2010). After filtering, the data were segmented into time series, performing both long temporal segmentation and short temporal segmentation. Each long-term segment served as a basis for determining whether the original signal indicated a leak. By voting on these judgments, the final detection result is obtained, effectively reducing the impact of a signal's non-stationary characteristics (e.g., fluctuations caused by intermittent environmental noise). Each short-term segment served as a feature extraction segment.

2.2.1. Long temporal segmentation

"Long temporal segmentation" involved the subdivision of acoustic signals into extended segments. The acoustic signals, when collected in real sampling environment, were susceptible to various interferences and signal fluctuations. For instance, during sample collection, external factors such as the honking of passing vehicles or the calls of nearby animals may introduce interference, resulting in changes to the acoustic signals (Fares et al., 2023). A key process of this paper was to segment the acquired signals into some long-term segments, each potentially containing specific longer-scale perturbative factors.

2.2.2. Short temporal segmentation

"Short temporal segmentation" involved the partitioning of the long-term segments into short-term segments. The acoustic signals obtained from an actual pipeline network constitute a category of non-steady-state random signals. This non-steady-state nature arose because, during a leak event, the turbulence intensity within the pipe increased, resulting in greater friction with the pipe walls. The instability of the turbulence results in a more nuanced form of non-stationarity within the acoustic signals. Employing short temporal segmentation on acoustic signals enabled us to posit that within exceedingly brief time intervals, the signal segment attained a quasi-stationary state, maintaining consistent information characteristics (Maheswari and Umamaheswari, 2017).

2.2.3. Temporal segmentation steps

This subsection illustrated how the data were segmented and the selection of the segmentation scale. The detailed steps were as follows :

First, a single acoustic signal was divided into multiple sub-samples (i.e., long temporal segmentation). For example, a segment of data denoted as X was segmented into $[X_1 \ldots X_m]$, where each X_i is a partial extraction of the original data, representing one characterization of the original signal. Next, each X_i underwent short temporal segmentation into $[X_{i1} \ldots X_{in}]$. Within each extremely short segment X_{ij}, the data could be regarded as a quasi-stationary signal.

During temporal segmentation, the specific selection of long-scale and short-scale parameters was determined by a series of experimental results. In this study, the long temporal segmentation scales were set at 0.125 s, 0.25 s, 0.5 s, 0.75 s, and 1.00 s; the short temporal segmentation points were set at 64, 128, 256, 512, and 1024. This study analyzed the recognition performance results of temporal segmentation at different scales to determine the optimal segmentation parameters for data preprocessing during the model development process.

2.3. Data transformation

After the data underwent temporal segmentation, transformation was performed. This study employed Mel spectrograms for data transformation. Its application aimed to portray the spectral distribution of a signal based on the Mel scale (Meng et al., 2019). The Mel scale was derived through a collection of Mel filters. This decision stemmed from the long-established practice of leak detection through auditory perception. The human ear was more sensitive to low-frequency sounds and less sensitive to high-frequency ones when processing auditory information. Moreover, the characteristic features of leakage acoustic signals were typically concentrated in the low-frequency range. Drawing on the work of Guo et al., who had employed time-frequency spectrograms for leak detection in water supply networks, it was observed that there was minimal information in the high-frequency portion of the leakage acoustic signal (Guo et al., 2021). Therefore, compressing the high-frequency data while amplifying the low-frequency data could assist the machine learning model in better feature learning. The generation process of the Mel spectrogram primarily encompassed frame segmentation, windowing, Fourier transformation, and Mel filter bank. Beneath the Mel scale, the correlation between Mel frequency and actual frequency followed a non-linear pattern, as delineated by Formula $fmel = 2595 \times lg(1 + f/700)$. Notably, at lower frequencies, Mel exhibited a more rapid variation concerning Hz, whereas, at higher frequencies, the rate of change was more gradual (Utebayeva et al., 2023). Experimental observations revealed a phenomenon within leakage sound waves, where low-frequency components harbored abundant information, contrasting with sparse information in high-frequency components. Consequently, features extracted utilizing the Mel scale showcased heightened compression and more robust representativeness, particularly in capturing leakage information.

Mel-frequency features were generated from each X_{ij} using a Mel filter bank, resulting in a one-dimensional matrix. $[X_{i1} \ldots X_{in}]$ formed a two-dimensional matrix, with the horizontal axis representing the time scale and the vertical axis representing the frequency scale. $[X_1 \ldots X_m]$ formed a three-dimensional matrix, where each X_i represents partial information about the original signal X.

2.4. Leak detection model

2.4.1. Convolutional neural network

The leak detection model in this study was based on CNNs. CNNs represented a category of deep learning models extensively utilized for the analysis of visual images (Li et al., 2016), making them well-suited for leak detection tasks when signals were transformed into Mel spectrograms. The model operated by leveraging temporal segmentation to preprocess data, followed by convolutional operations aimed at compressing and extracting features. Ultimately, this process enabled the classification and recognition of collected signals.

The fundamental principle of this method was to minimize the influence of non-stationary factors when identifying classification results for non-stationary signals obtained from real-world environments. The algorithm structure was presented in Fig. 1. In the temporal segmentation part, data preprocessing was performed to extract the two-stage acoustic signal features. The data then underwent training with a neural network model, which included data input, convolution, pooling, fully connected layers, and the output convolutional neural network result matrix. Finally, a decision function and voting mechanism were applied to recognize the data. During the incremental learning loop, first, an initial model was trained using ideal data with accurate labels. The model is then used to identify acoustic signals collected from the real-world environment (dataset with interference). The determination of the model's classification results relied on the recognition outcomes of long-term segmented sub-segments. Each sub-segment outputted a value between 0 and 1, with values closer to 1 indicating a greater likelihood of being a leakage segment, while values closer to 0 suggested that the sub-segment was normal or influenced by interference. By employing a voting mechanism, if 50 % or more sub-segments were classified as leakage segments, the entire signal X is considered to be a leakage signal; otherwise, it was classified as normal or interfered. Subsequently, based on the classification results outputted by the model and the sample labels, a filtering procedure was applied to retain 50 % of the longer segments, creating an incremental dataset for model optimization. The specific filtering criteria and rationale were detailed in Section 2.4.2. The incremental dataset was utilized for model optimization, and in the next round of recognition tasks, the optimized model was employed to repeatedly carry out the above processes.

2.4.2. Incremental learning

During the introduction to incremental learning, the concepts of various types of data may become easily confused. To clarify the different types of data used in this study, Fig. 2 was created to illustrate the source and intended purpose of each type. To construct a leak detection model that performed well, an ideal dataset was first used to serve as the foundation for training the initial model. These data were collected and labeled by professional technicians. In the ongoing application process, large amounts of data were required for further

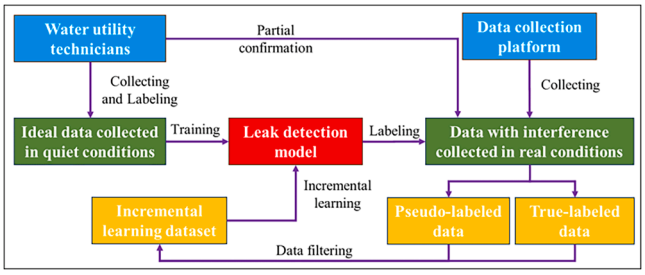

Fig. 2. Data classification and application.

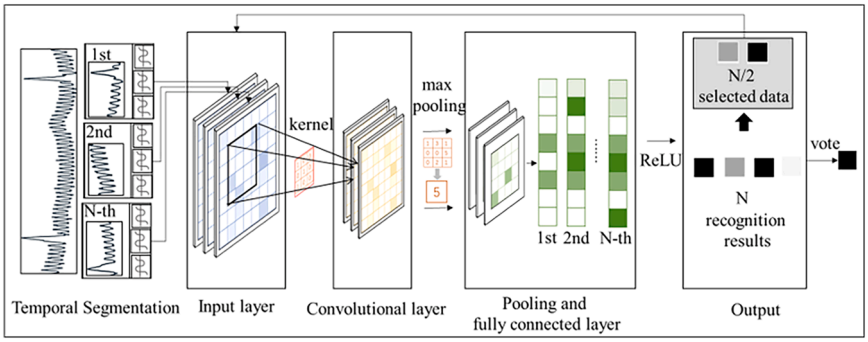

Fig. 1. Structure of model.

model optimization (i.e., incremental learning). However, among the vast number of newly collected signals, it was important to note that the data obtained from the monitoring platform was likely to contain substantial amounts of interference noise and lacked accurate labels, making them unsuitable for direct use. When utilizing this data, it was essential to perform data filtering to extract high-quality datasets suitable for model optimization.

The data with interference included true-labeled data that had been promptly verified on-site, as well as data that was pseudo-labeled by the already constructed leak detection model. Once acoustic monitoring devices were widely deployed, the amount of pseudo-labeled data was likely to far exceed that of true-labeled data, as water utilities lacked sufficient personnel to quickly process all the collected data. The quality of the data in this section varied, and only selected segments were extracted to serve as the incremental learning dataset for the model. The pseudo-labels for these data were partially validated in practice. This validation primarily came from feedback on-site leak detection results provided by pipeline managers during the leak detection process. These detection results were used to correct the data labels.

During the data filtering process, the definition of filtering criteria was specifically tailored for both true-labeled data and pseudo-labeled data.

(1) Rules for pseudo-labeled data

For samples identified by the model as leakage, the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the highest probability values were selected as leakage samples for the incremental dataset. Here, X_m referred to the m-th segment after the long temporal segmentation of the original acoustic signal X.

For samples identified by the model as normal, the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the lowest probability values were selected as normal samples for the incremental dataset.

(2) Rules for true-labeled data

In this study, leak signals were considered positive samples, while non-leak signals are negative samples. Based on the comparison of the model's output with the actual data labels, the following rules were applied for filtering incremental datasets.

For samples identified by the model as leakage and confirmed as leakage by manual inspection (i.e., true positive, TP), the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the highest probability values were selected as leakage samples for the incremental dataset.

For samples identified by the model as leakage but confirmed as normal by manual inspection (i.e., false positive, FP), the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the lowest probability values were selected as normal samples for the incremental dataset.

For samples identified by the model as normal and confirmed as normal by manual inspection (i.e., true negative, TN), the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the lowest probability values were selected as normal samples for the incremental dataset.

For samples identified by the model as normal but confirmed as leakage by manual inspection (i.e., false negative, FN), the original acoustic signal X was extracted. 50 % of the data segments from $[X_1 \ldots X_m]$ with the highest probability values were selected as leakage samples for the incremental dataset.

3. Experiments

3.1. Data collection

The collection of experimental data was conducted in three cities in China: SX, ZZ, and SH. In order to ensure data diversity, the data collection experiment was conducted across multiple locations and encompassed various types of pipe materials, such as steel pipes, ductile iron pipes, and PVC pipes. Data collection utilized a sensor designed to capture the vibration signals of pipeline leaks. According to the research by Bykerk et al., the impact of sensors from different manufacturers on the recognition results was relatively small (Bykerk and Miro, 2022b). The sensors used throughout the experiment were capacitive vibration sensors produced by a Chinese company. The data from each sensor was sampled for a duration of 5 s, with a sampling frequency of 8192 Hz and a signal amplification gain of 1000. This decision was informed by the observation that, under normal circumstances, the primary spectral peak of leak acoustic signals typically falls below 2000 Hz. According to the sampling theorem, a sampling frequency of 8192 Hz adequately met the experimental requirements (Jerri, 1977). During the experiment, 10 devices were used for on-site sampling of leakage acoustic signals, while 800 devices were utilized by water supply network operators on their management platforms.

The collected leakage sample data were obtained after confirming the presence of leakage points during road ground excavation, with data collection conducted in the early morning hours. This decision was made to ensure that the leakage sample data collected during the early morning hours would have minimal interference, thus providing a clearer representation of the leakage scenario. After repairing the pipeline, data collection was conducted on background noise to serve as non-leakage samples.

The ideal data collected on-site included 341 leak samples and 467 non-leak samples. Among the 341 leak acoustic samples, 31 were from PVC pipes, 71 were from steel pipes, 115 were from cast iron pipes, and 124 samples lacked pipe material information. Analyzing the data with material information revealed subtle frequency differences between leak acoustic signals from metallic and non-metallic pipes. Specifically, the frequency center of leak acoustic signals from metallic pipes was generally slightly higher than that from non-metallic pipes. To increase the data volume, each raw signal was sliced, with the first 3 s and the last 3 s treated as two independent samples for further analysis. Therefore, we obtained 1616 ideal samples with accurate labels. In addition, the water utilities provided us with a total of 1072 samples (data with interference) from the monitoring platform that require identification.

3.2. Data processing

According to the introduction in the methodology section, preprocessing was performed on the acquired data. Throughout the research process, we gradually established an acoustic database consisting of ideal data and an incremental data refined from data with interference. The ideal data and data with interference were both preprocessed, and the latter was randomly divided into four groups for incremental learning research. The first three groups of data underwent incremental learning, while the last group was used for model test. After the data with interference were pseudo-labeled, we communicated with pipeline management personnel to obtain on-site inspection results for the data with interference. This served as an important means to validate model performance.

3.3. Model implementation

The initial model training utilized the ideal data. The model input consisted of Mel spectrogram feature information extracted after two stages of data processing. The data were divided into training, validation, and test sets in a 7:2:1 ratio. The training set was used to train the model, the validation set evaluated the model's performance during training, and the test set assessed the final performance of the model (Li et al., 2016). The parameters of the neural network, including those of the convolutional kernels and pooling layers, were trained. Furthermore, hyper-parameters such as the number of neural network layers

and the convolutional kernel size were optimized and adjusted throughout the model training process using grid search. The Adam optimizer was set with a learning rate of 0.001 and a batch size of 64.

The model was also compared with other machine learning models, including Support Vector Machine (SVM) (Chauhan et al., 2019), Random Forest (RF) (Speiser et al., 2019), and XGBoost (Niazkar et al., 2024). The features processed by the RF, SVM, and XGBoost models were the 16 MFCC features (Mel-Frequency Cepstral Coefficients) (Peng et al., 2024b). MFCC features are spectral characteristics that represent the signal. They transform the frequency spectrum of the signal into a set of coefficients by simulating the human ear's perception of different frequencies. Meanwhile, I used a standard CNN model (Li et al., 2016) to directly process the raw signals. The original signal was divided into time segments, and then transformed from a one-dimensional matrix into a two-dimensional matrix to serve as the model input. To distinguish between these two different types of data inputs, I referred to them as CNN (Mel) and CNN (Raw Data), respectively. The hyper-parameters of the machine learning model were optimized using grid search to identify the optimal model parameters. Table 1 lists the hyper-parameters used in these models.

In CNN (Mel), the temporal segmentation process provided the model with two types of data, each with different meanings. The short-term segments offered information to characterize the state of the quasi-stationary segments, while the long-term segments represented specific potential non-stationary conditions caused by interference, serving as a single sampling for the classification of recognition samples. Temporal segmentation was included as a supportive condition for the classification recognition of original samples built upon the basic CNN model. Additionally, the four groups of data with interference used in the incremental learning process facilitated the study of changes in the model's cognitive level through the recognition-feedback-optimization-recognition cycle.

The computer used in this study is a Lenovo desktop with a 3.60 GHz Intel Core i7–9700k processor and 32GB of RAM. The model-building compilation environment is Python 3.7, and the primary Python library utilized was Tensorflow and Scikit-learn.

3.4. Model performance metrics

The performance metrics used to evaluate the model reflect its effectiveness in classification. These metrics are calculated using TP, TN, FP, and FN values, as illustrated. The model's performance is expressed using accuracy, precision, sensitivity (recall), and specificity, with the following calculations: Accuracy = (TP + TN) / (TP + FP + FN + TN), Precision = TP / (TP + FP), Sensitivity (Recall) = TP / (TP + FN), Specificity = TN / (FP + TN). Additionally, the AUC (Area Under the Curve) value is used to further evaluate the model's classification performance (Wang et al., 2013).

Table 1
Hyper-parameter Settings of Models.

	Hyper-parameters
RF	n_estimators = 500, max_features = 'auto', min_samples_leaf = 16, min_samples_split = 2
SVM	kernel = "rbf", C = 100, probability = True
XGBoost	objective = binary: logistic, n_estimators = 1000, learning_rate = 0.1, max_depth = 10
CNN (raw)	batch_size = 64, learning_rate = 0.002, dropout rate = 0.25, epochs = 200, filters = 32, kernel_size = (5,5), pooling size = (2, 2), strides = (2, 2)
CNN (Mel)	batch_size = 64, learning_rate = 0.001, dropout rate = 0.25, epochs = 200, filters = 32, kernel_size = (3,3), pooling size = (2, 2), strides = (2, 2)

4. Results and discussion

4.1. Temporal segmentation data processing results

Data are processed to compare performance across different temporal segmentation scales. For long segments: 0.125 s (Group A), 0.25 s (Group B), 0.50 s (Group C), 0.75 s (Group D), and 1.00 s (Group E). For short segments: 64 points (Group 1), 128 points (Group 2), 256 points (Group 3), 512 points (Group 4), and 1024 points (Group 5). This yields 25 groups (5 long × 5 short). Group A5, where long and short segments coincide, is invalid; thus, 24 valid groups remain. Model performance is assessed using metrics: accuracy, precision, sensitivity, specificity, and AUC.

The results (Fig. 3) indicate that Group B4 (0.25 s long temporal segmentation and 512 points for short temporal segmentation) provides the best model performance. First, by comparing the results of the long temporal segments, it is found that Group B2 outperforms the other groups in terms of overall performance. Additionally, as the window length changes, the model performance reaches its optimal level when using 512 points. In the leak detection, a long temporal segmentation of 0.25 s and a short temporal segmentation of 512 points are used as data segmentation methods, yielding superior recognition performance. In the subsequent sections of this research, all temporal segmentation processes are conducted following the optimal segmentation method.

The combination of 0.25 s for long temporal segmentation and 512 points for short temporal segmentation exhibits superior performance due to several key factors. The 0.25-second segmentation effectively captures rapid changes and transient features, adequately representing the non-stationary characteristics of the signals. For instance, during data collection, a frog's sound in the well lasts about 0.5 s, affecting the data within the 3-second window. Longer segments dilute short-term variations, while shorter segments lack contextual information. The 512-point segmentation provides sufficient data samples for generating Mel frequencies, allowing the model to learn relevant patterns while minimizing noise from excessive data. This combination enhances the model's sensitivity to subtle differences in acoustic features.

4.2. Comparison of CNN (Mel) with other models

When comparing the performance of the CNN (Mel) model with other machine learning models with manually extracted features, it is found that CNN (Mel) model outperformed them. In terms of accuracy, CNN (Mel) outperforms RF, SVM, and XGBoost by 23 %, 15 %, and 12 %, respectively. Regarding AUC values, CNN (Mel) surpasses RF, SVM, and XGBoost by 16 %, 11 %, and 7 %, respectively (Table 2). The CNN (Mel) model's core advantage is its two-stage segmentation process, which effectively preprocesses raw data. Compared to solely extracting MFCC coefficients for classification and recognition, CNN (Mel) demonstrates a significant improvement in recognition performance. This structured preprocessing in temporal segmentation ensures a better representation of acoustic signals, contributing to the superior performance of the model.

To elucidate the efficacy of temporal segmentation, this study also conducts a comparative analysis between the feature extraction capabilities of CNN (Raw Data) and CNN (Mel) for leak detection. The CNN (Mel) model demonstrates significant enhancements in performance across multiple metrics. The CNN (Mel) model shows improvements over the CNN (Raw Data) model in terms of accuracy, precision, sensitivity, and specificity, with enhancements of 3.1 %, 3.3 %, 4.5 %, and 2.1 %, respectively. This suggests that the architectural framework of CNN (Mel) effectively integrates the non-stationarity of the signals, facilitating the model's ability to capture relevant features for improved differentiation. Specific examples can be found in Section 4.3, where a voting mechanism utilizing long signal segments is employed to mitigate the effects of unstable signal disturbances.

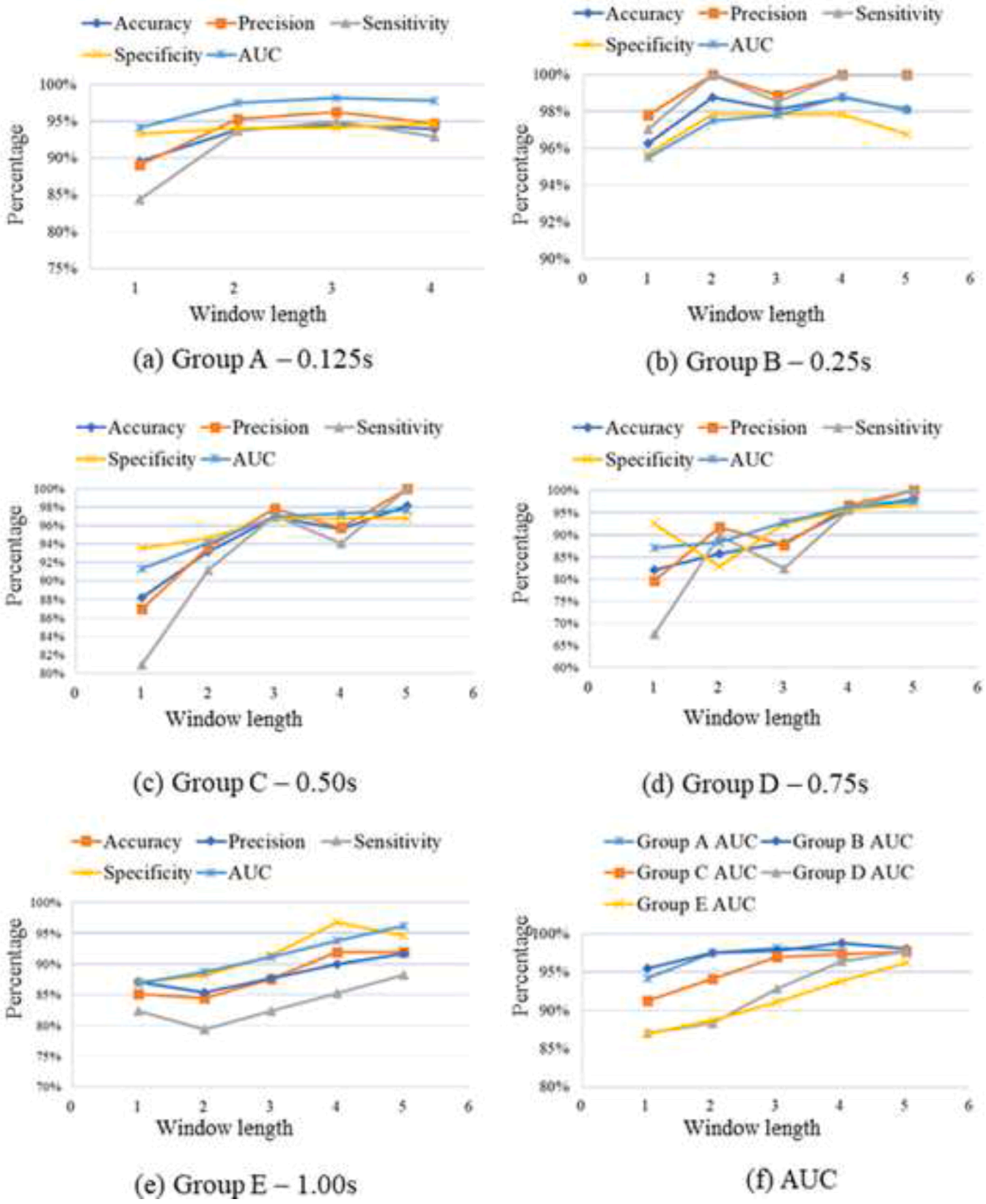

Fig. 3. Comparison of temporal segmentation performance.

Table 2
Comparison of performance metrics across different models.

	Accuracy	Precision	Sensitivity	Specificity	AUC
RF	0.7578	0.8261	0.7419	0.7677	0.8230
SVM	0.8385	0.9011	0.8548	0.8283	0.8770
XGBoost	0.8634	0.8667	0.7742	0.9192	0.9130
CNN (raw)	0.9568	0.9670	0.9552	0.9579	0.9566
CNN (Mel)	0.9876	1.0000	1.0000	0.9785	0.9880

4.3. Data selection in incremental learning

The objectives of dataset selection in incremental learning are twofold: to enhance the model's data comprehension and to ensure label accuracy during training. This section analyzes two representative signals: one labeled as leakage data and the other as normal data. These signals are divided into 23 long-term segments, and the detailed feature extraction process is outlined in the case study section. After model identification, each segment receives a recognition result. Overall, 14 of the 23 leakage segments are classified as leaks, suggesting it is likely leakage data, while 15 of the 23 normal segments are confirmed as normal. Misclassified segments probably contain significant interference, which may involve acoustic signals such as human speech, object collisions, and vehicle honking.

Therefore, when filtering this type of data, we select 50 % of the long-term segments that match the true labels or pseudo-labels as the incremental learning dataset. It should be noted here that the filtering of pseudo-label data is related to the leak detection information provided to us by the water supply managers. For data that has been inspected and verified, a filtering approach tailored to real data is applied; for data that has not been inspected or verified, a filtering approach for pseudo-label data is used. The detailed rules are provided in Section 2.4.2. This method preserves the authenticity of the data, enriches the diversity of the database, and prevents the model's recognition performance from degrading due to learning from an excessive amount of interference data.

4.4. Results of incremental learning

The dataset for incremental learning is sourced from the continuous collection of pipeline acoustic signals from WSNs (monitoring platform). However, these data often contain noise or suffer from missing labels. By applying the data filtering rules outlined in the paper, part of the noisy data can be removed, allowing us to extract higher-value data for incremental learning. In this process, the data with interference are utilized to optimize the model. According to the experimental setup, the data with interference are randomly divided into four groups. The first three groups are used sequentially for incremental data selection, model updating, and recognition result comparison, while the fourth group is reserved for evaluating the model's recognition performance.

For the first three sets of data, we identify the model's recognition errors, recording both false positive and false negative instances. We apply label correction to these misidentified data points, and the corrected labels are used in the subsequent incremental learning phase. Over the course of three rounds of incremental learning, the false positive rates are 5.2 %, 1.1 %, and 0.0 %, while the false negative rates are 7.1 %, 5.9 %, and 3.7 %. To demonstrate the effectiveness of pseudo-label data selection, we train the model using a dataset that excludes pseudo-labels and compare its recognition performance. On a separate test set, the models that incorporate pseudo-labeled data achieve accuracies, precisions, sensitivities, and specificities of 0.9739, 0.9554, 0.9407, and 1.0000, respectively, while the models without pseudo-labeled data perform with values of 0.9515, 0.9477, 0.9322, and 0.9667.

Overall, during the incremental learning process, the model's recognition performance metrics shows a consistent improvement, enhancing its ability to classify and recognize data. Initial recognition detection results on the data with interference using the original model show a decline in various evaluation metrics compared to the results (Fig.3b) on the original dataset, as shown in Fig. 4. This indicates that the original model lacks sufficient recognition capability when faced with a new dataset with interference, highlighting the necessity of enhancing the model's understanding of the data. Incremental learning is an essential step in practice to ensure continuous model updates and improve recognition capabilities. After incremental learning, the model's recognition and understanding of the new data improve, which also validates the model's feasibility in real-world operation. Newly discovered leak acoustic signals in future operations can be effectively utilized, continuously enriching the model's understanding of the data.

5. Conclusions

This study investigates leak detection through the use of acoustic signals collected from real-world pipeline networks. Key conclusions drawn are as follows:

(1) Segmentation across different scales effectively handles non-stationary signals. Long-term segments reduce noise interference and aid in leakage detection via comprehensive multi-segment analysis, while short-term segments decompose signals into quasi-stationary components for easier feature extraction.
(2) Mel-spectrograms are well-suited for analyzing leakage signals as they provide high resolution at low frequencies where key leakage components are concentrated. This makes them particularly effective for extracting and representing relevant features.
(3) Incremental learning enhances model performance by incorporating newly collected data and applying filtering rules to exclude irrelevant noise. This process accounts for the complexities of real-world pipeline data, including non-stationarity, noise, and the lack of accurate labels, thereby ensuring efficient learning and enhancing the model's ability to generalize to real-world scenarios.

Future research should focus on continuously collecting actual leakage and interference data to build a standardized database and enhance the model's ability to capture leakage patterns. Additionally, the acoustic interference caused by pipeline components instead of environmental factors, such as valves and elbows, must be considered, as these fittings complicate the distinction between leakage and background noise. Since data collection often occurs near valve wells and meter wells, future studies should address leak detection challenges in these interferences.

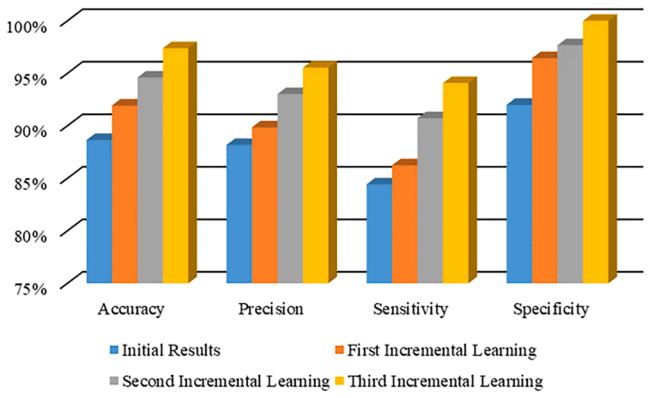

Fig. 4. Incremental learning results.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT 4o in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Xingke Ma: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Yipeng Wu:** Writing – review & editing, Visualization, Validation, Software, Methodology, Conceptualization. **Guancheng Guo:** Software, Methodology, Formal analysis, Data curation. **Shuming Liu:** Supervision, Project administration, Funding acquisition, Conceptualization. **Yuexia Xu:** Validation, Resources, Data curation. **Jingjing Fan:** Validation, Resources, Data curation. **Hongbin Wang:** Resources, Data curation. **Liren Xu:** Resources, Data curation.

Declaration of competing interest

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

Acknowledgments

This work was financially supported by the National Key R&D Program of China (Grant No 2023YFC3208100).

Data availability

The authors do not have permission to share data

References

Adegboye, M.A., Fung, W.K., Karnik, A., 2019. Recent advances in pipeline monitoring and oil leakage detection technologies: principles and approaches. Sensors 19 (11).
Ahmad, Z., Nguyen, T.K., Kim, J.M., 2023. Leak detection and size identification in fluid pipelines using a novel vulnerability index and 1-D convolutional neural network. Eng. Applic. Comput. Fluid Mech. 17 (1).
Ahopelto, S., Vahala, R., 2020. Cost-benefit analysis of leakage reduction methods in water supply networks. Water. (Basel) 12 (1).
Bae, J.H., Yeo, D., Yoon, D.B., Oh, S.W., Kim, G.J., Kim, N.S., Pyo, C.S., 2018. DEEP-LEARNING-BASED PIPE LEAK DETECTION USING IMAGE-BASED LEAK FEATURES, pp. 2361–2365.
Bykerk, L., Miro, J.V., 2022a. Detection of water leaks in suburban distribution mains with lift and shift Vibro-acoustic sensors. Vibration. 5 (2), 370–382.
Bykerk, L., Miro, J.V., 2022b. Vibro-acoustic distributed sensing for large-scale data-driven leak detection on urban distribution mains. Sensors 22 (18).
Chauhan, V.K., Dahiya, K., Sharma, A., 2019. Problem formulations and solvers in linear SVM: a review. Artif. Intell. Rev. 52 (2), 803–855.
Cody, R.A., Tolson, B.A., Orchard, J., 2020. Detecting leaks in water distribution pipes using a deep autoencoder and hydroacoustic spectrograms. J. Comput. Civil Eng. 34 (2).
Datta, S., Sarkar, S., 2016. A review on different pipeline fault detection methods. J. Loss. Prev. Process. Ind. 41, 97–106.
Fares, A., Tijani, I.A., Rui, Z., Zayed, T., 2023. Leak detection in real water distribution networks based on acoustic emission and machine learning. Environ. Technol. 44 (25), 3850–3866.
Gunatilake, A., Miro, J.V., 2024. Multimodel neural network for live classification of water pipe leaks from Vibro-acoustic signals. IEEe Sens. J. 24 (9), 14825–14832.
Guo, G.C., Yu, X.P., Liu, S.M., Ma, Z.Q., Wu, Y.P., Xu, X.Y., Wang, X.T., Smith, K., Wu, X., 2021. Leakage detection in water distribution systems based on time-frequency convolutional neural network. J. Water. Resour. Plan. Manage 147 (2).
Guo, G.C., Yu, X.P., Liu, S.M., Xu, X.Y., Ma, Z.Q., Wang, X.T., Huang, Y.J., Smith, K., 2020. Novel leakage detection and localization method based on line spectrum pair and cubic interpolation search. Water Res. Manag. 34 (12), 3895–3911.
Islam, M.R., Azam, S., Shanmugam, B., Mathur, D., 2022. A review on current technologies and future direction of water leakage detection in water distribution network. IEEe Access. 10, 107177–107201.
Jerri, A.J., 1977. Shannon Sampling Theorem - its various extensions and applications - tutorial review. Proceed. Ieee 65 (11), 1565–1596.
Jiang, W.X., Wang, H.H., Liu, G.J., Liu, Y.H., Cai, B.P., Li, Z.X., 2021. A novel method for mechanical fault diagnosis of underwater pump motors based on power flow theory. IEEe Trans. Instrum. Meas. 70.
Li, Y., Hao, Z., Lei, H., 2016. Survey of convolutional neural network. J. Comput. Appl. (China) 36 (9), 2508-2515, 2565.
Liu, R.S., Zayed, T., Xiao, R., 2024. Advanced acoustic leak detection in water distribution networks using integrated generative model. Water. Res. 254.
Luo, S., Johnston, P., 2010. A review of electrocardiogram filtering. J. Electrocardiol. 43 (6), 486–496.
Maheswari, R.U., Umamaheswari, R., 2017. Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train - A contemporary survey. Mech. Syst. Signal. Process. 85, 296–311.
Martini, A., Rivola, A., Troncossi, M., 2018. Autocorrelation analysis of vibro-acoustic signals measured in a test field for water leak detection. Appl. Sci.-Basel 8 (12).
Meng, H., Yan, T.H., Yuan, F., Wei, H.W., 2019. Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEe Access. 7, 125868–125881.
Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P., Righetti, M., 2024. Applications of XGBoost in water resources engineering: a systematic literature review (Dec 2018-May 2023). Environ. Modell. Softw. 174.
Peng, H., Xu, Z., Huang, Q.L., Qi, L.Q., Wang, H.T., 2024a. Leakage detection in water distribution systems based on logarithmic spectrogram CNN for continuous monitoring. J. Water. Resour. Plan. Manage 150 (6).
Peng, L.G., Zhang, J.C., Li, Y.Q., Du, G.F., 2024b. A novel percussion-based approach for pipeline leakage detection with improved MobileNetV2. Eng. Appl. Artif. Intell. 133.
Quy, T.B., Muhammad, S., Kim, J.M., 2019. A reliable acoustic EMISSION based technique for the detection of a small leak in a pipeline system. Energies. (Basel) 12 (8).
Ramezani, M.G., Hasanian, M., Golchinfar, B., Saboonchi, H., 2020. Automatic Boiler Tube Leak Detection With Deep Bidirectional LSTM Neural Networks of Acoustic Emission Signals. Electr Network.
Ravichandran, T., Gavahi, K., Ponnambalam, K., Burtea, V., Mousavi, S.J., 2021. Ensemble-based machine learning approach for improved leak detection in water mains. J. Hydroinform. 23 (2), 307–323.
Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert. Syst. Appl. 134, 93–101.
Utebayeva, D., Ilipbayeva, L., Matson, E.T., 2023. Practical study of recurrent neural networks for efficient real-time drone sound detection: a review. Drones 7 (1).
Wang, J., Huang, P.L., Sun, K.W., Cao, B.L., Zhao, R., 2013. Ensemble of Cost-Sensitive Hypernetworks For Class-Imbalance Learning, pp. 1883–1888.
Wang, Z.F., He, X.Q., Shen, H.L., Fan, S.J., Zeng, Y.L., 2022. Multi-source information fusion to identify water supply pipe leakage based on SVM and VMD. Inf. Process. Manage 59 (2).
Wu, Y.P., Liu, S.M., Kapelan, Z., 2024a. Addressing data limitations in leakage detection of water distribution systems: data creation, data requirement reduction, and knowledge transfer. Water. Res. 267.
Wu, Y.P., Ma, X.K., Guo, G.C., Huang, Y.J., Liu, M.Y., Liu, S.M., Zhang, J., Fan, J.J., 2023. Hybrid method for enhancing acoustic leak detection in water distribution systems: integration of handcrafted features and deep learning approaches. Process Saf. Environ. Protec. 177, 1366–1376.
Wu, Y.P., Ma, X.K., Guo, G.C., Jia, T.L., Huang, Y.J., Liu, S.M., Fan, J.J., Wu, X., 2024b. Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective. Water. Res. 261.
Yang, D., Ren, W.X., Hu, Y.D., 2021. Non-stationary assessment of structural operational measurements using recurrence quantification analysis. Measurement 171.
Yussif, A.M., Sadeghi, H., Zayed, T., 2023. Application of machine learning for leak localization in water supply networks. Buildings 13 (4).
Zhang, C., Alexander, B.J., Stephens, M.L., Lambert, M.F., Gong, J.Z., 2023a. A convolutional neural network for pipe crack and leak detection in smart water network. Struct. Health Monitor. Int. J. 22 (1), 232–244.
Zhang, P., He, J.G., Huang, W.Y., Zhang, J., Yuan, Y.Q., Chen, B., Yang, Z., Xiao, Y.F., Yuan, Y.X., Wu, C.G., Cui, H., Zhang, L.D., 2023b. Water pipeline leak detection based on a pseudo-siamese convolutional neural network: integrating handcrafted features and deep representations. Water. (Basel) 15 (6).