AQUA
Water Infrastructure, Ecosystems and Society

 AQUA — Water Infrastructure, Ecosystems and Society Vol 73 No 8, 1741 doi: 10.2166/aqua.2024.184

Leak detection in water distribution networks based on deep learning and kriging interpolation method

Huimin Yu[a,b], Sen Lin[c], Hua Zhou[d], Xiaodan Weng[d], Shipeng Chu[a,b] and Tingchao Yu[a,b,*]

[a] Zhejiang Key Laboratory of Drinking Water Safety and Distribution Technology, Zhejiang University, Hangzhou 310058, China
[b] Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing 314100, China
[c] Ningbo Donghai Group Co., Ltd, Ningbo 315537, China
[d] Huadong Engineering Corporation Limited, Hangzhou 311122, China
*Corresponding author. E-mail: yutingchao@zju.edu.cn

ABSTRACT

The burgeoning growth of urban areas has escalated the necessity for efficient and precise leak detection in water distribution networks. Automatic detection methods based on deep learning are a state-of-the-art research topic. In this paper, a methodology that combines deep learning and data imaging is proposed. The framework employs pressure monitoring data and is anchored on the following three pillars: (1) the generation of a comprehensive dataset, encompassing one year of leak-free demand data derived from Fourier Series analysis and monitoring pressure under normal and leak conditions, (2) the transformation of pressure time series into images using kriging interpolation, (3) establishing convolution neural network (CNN) and evaluating its performance of abnormal identification. The effectiveness of the proposed methodology is assessed in different image sets under various leak conditions. The findings reveal that this method meets dependable and effective outputs for leak detection, with the deep learning model achieving a high true positive rate (TPR) of 98% and an area under the curve (AUC) of 94%. This study provides invaluable information for strategic action planning and the enhancement of water loss management protocols, especially in situations where water utilities and regulatory authorities grappling with limited budgets and diminishing revenues.

Key words: abnormal detection, data imaging, deep learning, pressure variation, water distribution networks

HIGHLIGHTS

- Generating a dataset with monitoring pressure data under normal and leak conditions.
- Utilizing data imaging technology to convert discrete data into continuous pseudocolor spatial pressure distribution images.
- Developing a leakage detection model for water distribution networks using VGG16, a convolutional neural network.
- Comparing the performance of the proposed model with other deep learning models.

1. INTRODUCTION

Water loss in water distribution networks (WDNs) is a significant issue in water resource management. Serious leakage incidents not only precipitate substantial water and energy wastage but also catalyze secondary disasters, which lead to the invasion of external pollutant contamination and bacterial through leak holes (Karim *et al.* 2003; LeChevallier *et al.* 2003; Fontanazza *et al.* 2015), and even ground collapse (Guo *et al.* 2013). According to statistics, the annual global volume of water loss amounts to approximately 126 billion cubic meters, translating to a financial loss of around US $39 billion (Liemberger & Wyatt 2018). Therefore, accurate and advanced leakage identification technologies are crucial for saving valuable water resources and fostering the sustainable development of society.

Numerous factors contribute to the occurrence of leakages, including the poor quality of pipes, exceeding service time limits, and alterations in mechanical conditions due to subway construction, etc. However, WDNs are typically buried underground and span vast geographical areas, which makes them vulnerable to structural integrity failure (Van Leuven 2011). Consequently, many minor leaks remain undetected over short periods, complicating the assessment of both the timing and location of a leak. To enhance the performance of leakage identification technologies, two primary categories of anomaly detection methods are extensively employed in the field (Li *et al.* 2015). The first category is hardware-based and focuses on the use of physical devices to detect discrepancies. The second category relies on software, which is designed to analyze and interpret data to identify anomalies (Valizadeh *et al.* 2009).

Many technologies rely on specialized hardware equipment, such as leak noise loggers (Muggleton *et al.* 2006), leak noise correlators (Guo *et al.* 2021), and infrared photography (Fahmy & Moselhi 2010). These technologies utilize distinct physical phenomena to accurately detect and locate leaks in WDNs. However, their application to large-scale WDNs is limited by significant challenges, including the high costs of the equipment, the labor-intensive nature of the operations, and the generally low efficiency of these methods when deployed on a large scale (Romano *et al.* 2011).

With the rapid advancement of supervisory control and data acquisition (SCADA) systems, the integration of advanced computer technologies such as machine learning for analyzing and processing monitoring data is becoming increasingly prevalent (Zhou *et al.* 2019). SCADA systems can control and monitor key hydraulic data of WDNs in real-time, such as node pressure and demand, providing critical data support for abnormal event identification technologies. Compared to flow meters, pressure sensors are more cost-effective and capable of providing instantaneous data (Sun *et al.* 2020), and enabling the collection of more comprehensive information on changes in operating conditions. Therefore, this study concentrates on analyzing variations in node pressure and their spatial distribution to enhance the accuracy of leakage identification.

However, the ability of SCADA systems to provide rich spatiotemporal information is limited by the vast scale, complex topology of WDNs and the limited number of monitoring points. Zhou *et al.* (2022) leveraged graph signal processing to reconstruct slow-varying components, which improves estimation accuracy and provides unknown nodal pressures. To further visualize the spatially continuous pressure variations in WDNs, this study employs spatial interpolation methods to generate continuous spatial data from point samples. Li & Heap (2008) categorized spatial interpolation methods into non-geostatistical interpolation methods, geostatistical interpolation methods, and combined methods, briefly describing a total of 38 methods. Subsequently, Li & Heap (2011) summarized over 70 spatial interpolation methods/sub-methods in environmental science, identifying inverse distance weighting (IDW), ordinary kriging (OK), and ordinary co-kriging (OCK) as the most frequently compared methods. According to the results of 53 comparative studies, OK demonstrated a favorable application effect.

Data-driven leak detection methodologies utilizing machine learning techniques have undergone extensive investigation, primarily concentrating on the analysis of pressure (Yu *et al.* 2023), flow (Moors *et al.* 2018), and transient wave data (Zhou *et al.* 2020). However, these approaches predominantly excel in detecting severe leaks, such as pipe bursts, or in identifying abnormal events within relatively simplistic WDNs. To effectively surmount the constraints inherent in machine learning, it is imperative to develop a novel approach characterized by: (1) the capacity to analyze complex features, (2) robustness sufficient to withstand uncertainties such as variability in water consumption and monitoring noise, (3) scalability to accommodate extensive detection datasets, (4) adaptability to WDNs of diverse scales and configurations, and (5) the ability for autonomous learning with reduced dependence on manual design.

Deep learning, an artificial intelligence (AI) approach, has undergone a significant resurgence since 2006. This revival has been driven by the development of new algorithms that project input spaces into progressively lower-dimensional latent representations in a hierarchical manner. These algorithms do not require domain expertise or human supervision, and can automatically extract complex patterns and relationships. And their performance significantly surpasses that of conventional anomaly detection methods (Pang *et al.* 2022). Recent research has leveraged deep learning to enhance leak detection capabilities in WDNs, demonstrating the effectiveness of this methodology. Romano *et al.* (2014) utilized AI techniques to forecast pressure and flow signal values and applied statistical methods to analyze signals from multiple and distinct District-Metered Areas (DMAs). These DMAs can be dynamically recalibrated as conditions in the WDN change. However, the effectiveness of this approach is influenced by the size of the DMA. As the scale of the network increases, there is a corresponding decrease in sensitivity to anomalous events. Rajabi *et al.* (2023) introduced a Conditional Deep Convolutional Generative Adversarial Network (CDCGAN) to convert images of node demand into corresponding pressure images. To differentiate between normal and leakage images, a threshold was set based on one year of normal operational monitoring data and the 3σ (three-sigma) principle. A leak is identified in the WDN when the similarity between the real-time monitoring images and the pressure images predicted by the CDCGAN falls below the threshold for five consecutive time steps. However, this threshold-based method does not accommodate real-time updates based on variations in weekly and seasonal changes, which can lead to diminished accuracy.

Inspired by the aforementioned studies on leak detection, this paper aims to delve into the complex relationships between hydraulic parameters and attributes of operational conditions based on a deep learning model. The core of this study is to extract deep features from pressure distribution images under varying conditions to enhance the accuracy of leak detection methods. This framework integrates geographic information of urban WDNs with monitored pressure data, and encodes these data into images using the Kriging interpolation method. These spatial pressure distribution images, along with derived labels, are fed into a convolutional neural network (CNN)-based backbone model. The model then extracts image features and identifies anomalous images indicative of leaks, thereby facilitating more accurate and efficient leak detection.

2. MATERIALS AND METHODS

The proposed method is a novel leak detection method based on data imaging techniques and a deep learning model. Different operation condition means different spatial distributions of monitoring data. According to this otherness, abnormal events are identified through the classification process. Section 2.1 introduces the framework structure of the proposed leakage detection method. Section 2.2 introduces the generation of training data. Section 2.3 introduces the data imaging technology based on Kriging interpolation. Section 2.4 introduces the backbone network of the leak detection based on visual geometry group 16 (VGG16).

2.1. Leakage detection framework based on VGG16

Figure 1 shows the flowchart of a novel leak detection framework. The framework can be divided into four parts, which are dataset generation; data imaging; VGG16 training; and leak detection process. The first step of the framework is to add different sizes of leak holes to the leak pipes and traverse potential pipes to simulate various scenarios. First, normal demand with a seasonal trend was generated by section 2.2.1, and normal pressure at each node was calculated by the hydraulic model. The acquisition of leak pressure at every sensor is based on the method of section 2.2.2. Then, depending on the pressure data of sensors, pressure distribution images were generated by the kriging interpolation method, which is demonstrated in section 2.3. Part of these images is randomly selected and input into the VGG16 model. The output of the classification is the probability value P_n of the leak pipe, and the maximum probability value is used to determine whether each image contains a leak.

2.2. Generation of hydraulic parameters

The framework proposed in the paper is data-driven, so the amount of information stored in the dataset has a significant influence on the performance of the deep learning models. The dataset consists of pressure monitoring data for two operating conditions, including normal conditions and various leakages.

2.2.1. Generation of node demand under normal conditions

Flow and pressure are two important hydraulic indexes reflecting the operating state of the network. Under normal conditions, the variation of node water consumption or pressure data consists of daily, weekly, seasonal patterns and other uncertain factors, which is calculated by Equation (1).

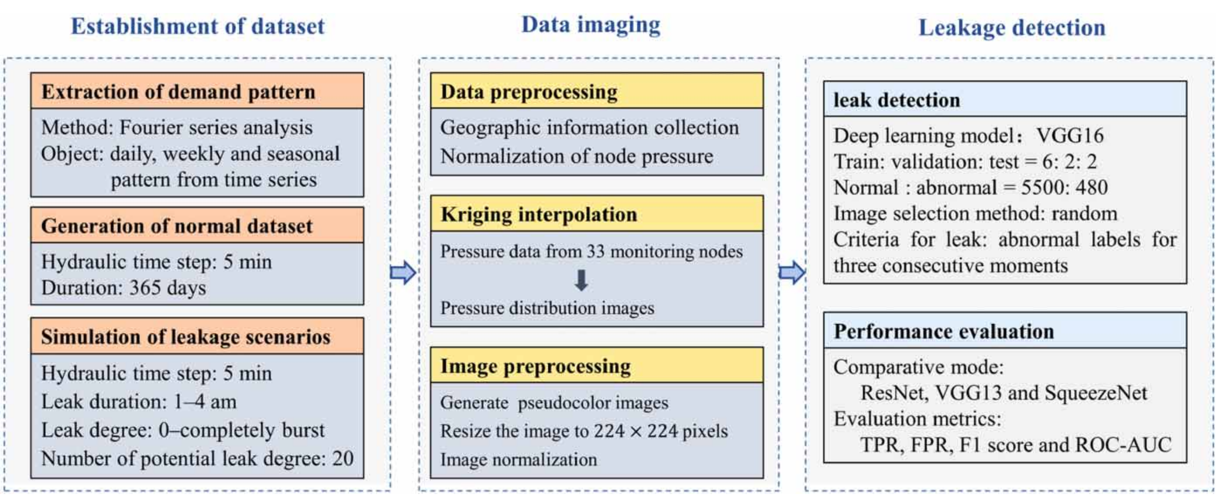

Figure 1 | Flowchart of the leak detection framework.

Node demand collection under normal conditions is the first step for leakage detection. This section adopted empirical mode decomposition (EMD) (Karthikeyan & Nagesh Kumar 2013) to extract different patterns of two DMAs in the same town, Area A is used to extract the seasonal trend in a year because it has no leak, and Area B is used to extract the weekly pattern in a non-leak condition. Combining these two patterns, the normal node demand of Area B in one year can be estimated (Wang *et al.* 2022).

$$Y(t) = S(t) \times T(t) \times R(t) \tag{1}$$

where Y(t) is the normal water consumption of a WDN, S(t) is the seasonal trend, T(t) is the weekly pattern, and R(t) is a random term in the data.

2.2.2. Generation of node pressure under normal and leak conditions

Although the flow rate is more sensitive to leakages, pressure data are more likely to reflect real leak zones due to the amount of pressure sensors being far more than inlet flowmeters. The variation of different pressure sensors differs from the spatial position of sensors when a leak occurs. Therefore, some leak data could make the detection result more accurate. The core step of simulating leakages is adding a hole to the middle of the pipe, and setting different sizes of the hole. As the probability of a leak occurring on every pipe is the same, each pipe is set as a potential leak pipe, and Table 3 lists the parameters for simulating various potential leak scenarios.

For the degree of different leaks, the general equation to calculate leak flow is proposed by Crowl & Louvar (2001), which is expressed as:

$$Q_\mathrm{l} = C_\mathrm{d} A \sqrt{2gh}, \ \text{when } \alpha = 0.5 \tag{2}$$

where Q_l is the leak flow, C_d is the discharge coefficient and the default value is 0.75 for turbulent flow, A is the area of the leak hole, α is an exponent and always set to 0.5 (Greyvenstein & Van Zyl 2007), g is the acceleration of gravity (m/s^2), and h is the pressure head (m).

2.3. Transforming pressure data to images

In order to deeply mine the information contained in the various pressure data, the pressure data was transformed into images and a deep learning model was used to extract image features. The data imaging method used in this paper is kriging interpolation, one of the popular spatial interpolation methods, that could interpolate the values of the primary variable at unsample locations to form a continuous spatial field (Li & Heap 2008).

The basic estimation formula of this method can be represented as weighted averages of sampled data, as Equation (3). Due to spatial variability did not increase linearly with distance, the semivariance, $\gamma(h)$, could be estimated by Equation (4), and Exponential, Gaussian, and Spherical models were developed to quantify the similarity of spatial variability.

The kriging estimator used in this paper is OK, the equations can be expressed as Equations (5) and (6). By calculating the mean average error (MAE), the Gaussian variogram model was used in this paper.

$$\hat{z}(x_0) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_\mathrm{i} \mathrm{z}(x_\mathrm{i}) \tag{3}$$

$$\gamma(h) = \frac{1}{2n} \sum_{\mathrm{i}=1}^{\mathrm{n}} [z(x_\mathrm{i}) - z(x_\mathrm{i} + h)]^2 \tag{4}$$

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_\mathrm{i} \gamma(x_\mathrm{i}, x_\mathrm{j}) - \mu = \gamma(x_\mathrm{i}, x_0) \tag{5}$$

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_\mathrm{i} = 1 \tag{6}$$

where $\hat{z}(x_0)$ is the estimated value at the point of x_0, $z(x_i)$ is indicated as an observed value at the point of x_i, λ_i is the weight allied to data, n indicates the number of point pairs that are separated by a distance of h, μ means Lagrange multiplier to ensure the predicted value is unbiased, and $\gamma(x_i,x_j)$ means the value of variogram reciprocal.

Pressure data at all monitoring points in a single time step are inputted into the kriging interpolation model to create images. The results are then scaled into the interval 0–255 to generate rectangular images with 224×224 pixels. The mean and standard deviation of the normal image are calculated from the normal image set. Then the image set is normalized according to the Z-score (Roshan *et al.* 2019), given as Equation (7). All the images are normalized and input into the deep learning model.

$$X_{\text{norm}} = \frac{X - \mu}{\sigma} \tag{7}$$

where X = pixel value per image, and μ and σ = the average and the standard deviation of all images.

2.4. Leak detection based on VGG16

Convolutional networks (ConvNets) are a prominent method in the research of deep visual representation. To enhance the accuracy of convolutional neural networks in large-scale image recognition, VGGNET (Simonyan & Zisserman 2015) utilizes an architecture that incorporates very small (3×3) convolution filters and increases the model's depth to between 13 and 19 layers. VGG16, a particularly high-performing network within VGGNET, consists of 16 layers that contain parameters. The structure of the model, depicted in Figure 2, is divided into two main parts: a feature extraction layer and a classification layer. The feature extraction component comprises five blocks, where blocks 1–2 contain two convolutional layers each, and blocks 3–5 contain three convolutional layers each. Following each convolution, a ReLU activation function is applied, and a maximum pooling layer with a 2×2 kernel size is added at the end of each block. The final segment of the VGG16 model consists of three fully connected layers, culminating in an output obtained via the sigmoid function.

3. CASE STUDY AND RESULTS

3.1. Description of the case study

This case study network serves a total population of 10,000, including residential and commercial customers (Vrachimis *et al.* 2022). The network has 782 junctions, 2 reservoirs, 1 pump, 1 tank, and 905 pipes. In addition, 33 pressure sensors and 2 flow sensors have been deployed at the most sensitive nodes in the network to monitor changes in operating conditions rapidly and comprehensively. It is divided into three key DMAs: DMA A includes residential and commercial facilities, DMA B includes a pressure relief valve to reduce incipient leaks, and DMA C includes a pump and a tank to provide adequate pressure to users in that zone, and is equipped with 82 water demand reading sensors, as shown in Figure 3.

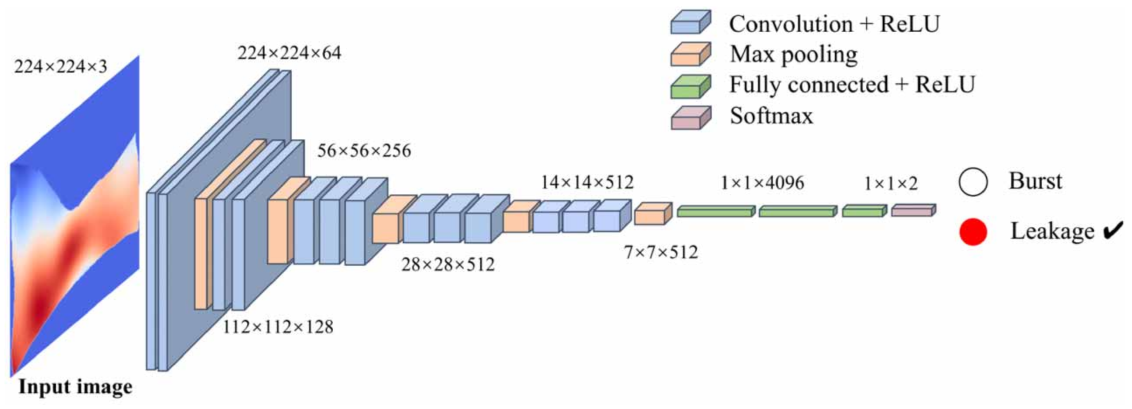

Figure 2 | The architecture of VGG16 and how it extracts the image features.

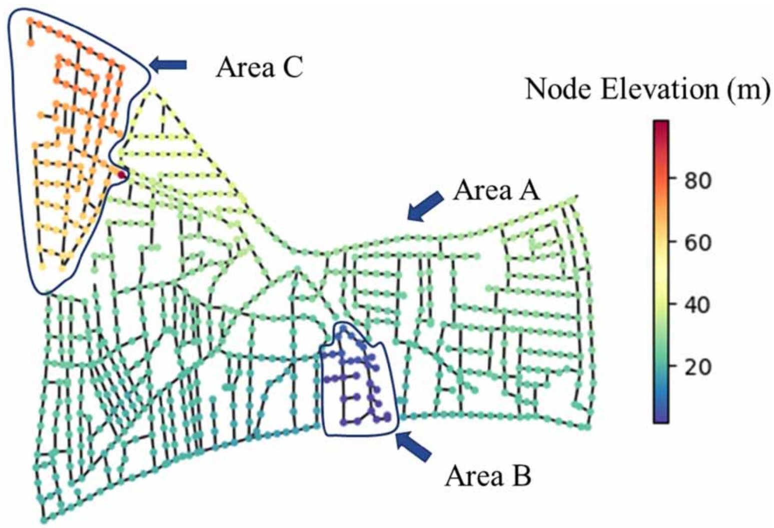

Figure 3 | Diagram of the L-town network.

3.2. Establishment of dataset

There are many methods to simulate leakages such as the artificial reservoir method (Ang & Jowitt 2006), the emitter method (Giustolisi *et al.* 2008), and the additional node demand method (Shao *et al.* 2019). The simulating method adopted in this paper is to add a leak hole in the middle of the pipe and analyze the hydraulic operating parameters by the Water Network Tool for Resilience (WNTR) (Klisel *et al.* 2018). WNTR is an open-source Python package designed to help water utilities simulate and analyze the resilience of WDN, which is compatible with EPANET 2.00.12 and EPANET 2.2. The information on simulation parameters is listed in Table 2.

The collection of normal node demand and pressure is the basis of leak detection. According to the method in section 2.2, the Fourier series is used to extract the weekly and seasonal pattern of the L-town. Combining the patterns and the base demand, the normal node demand for one year was calculated by the inverse Fourier transformation, the result is shown in Figure 4. These demand data are then input into the WNTR model to calculate the pressure time series of monitoring sensors.

Pressure data under different leakage scenarios is also important for VGG16. It is assumed that the possibility of a leak in 905 pipes is the same, and the size of the leak hole represents the severity of the leak. Table 1 shows the leak start time, the duration of the leak, the degree of leak and other simulation parameters.

3.3. VGG16 training

The experiments were conducted on a computer that is configured with NVIDIA GeForece GTX 1660 SUPER; Windows 11 64-bit operating system. CUDA version 11.8, Pytorch 2.1.2, Python 3.9. The software mainly used is the OpenCV image-processing package. For this experiment, pressure data under two operating conditions were combined to form a dataset, and then the kriging interpolation method was used to generate the pressure images, with a uniform size of 224×224. All these images compose an image set. A real image set should be an unbalanced image set, with a much larger proportion of normal images than abnormal ones. Therefore, 5,500 normal images and 480 abnormal images were randomly and repeatedly selected from the set for model optimization and then divided into three parts with the proportions for training, validation, and test set as 60, 20, and 20%, respectively.

When leaks of varying magnitudes occur in different pipes, the resulting pressure drops at monitoring points vary, leading to changes in the spatial pressure distribution in the network. Figure 5 illustrates the spatial pressure distribution across three consecutive time steps under various operational scenarios, where areas of high pressure are indicated in red and low pressure in blue. Figure 5(a) depicts the network in a leak-free state, characterized by relatively minor changes in the image's pixel values across different times, indicating slight pressure fluctuations during this period. Conversely, in Figure 5(b), when a leak occurs, the pressure sensor closest to the leak initially reacts, with the extent of the leakage directly influencing the magnitude of change observed in the image's pixel values.

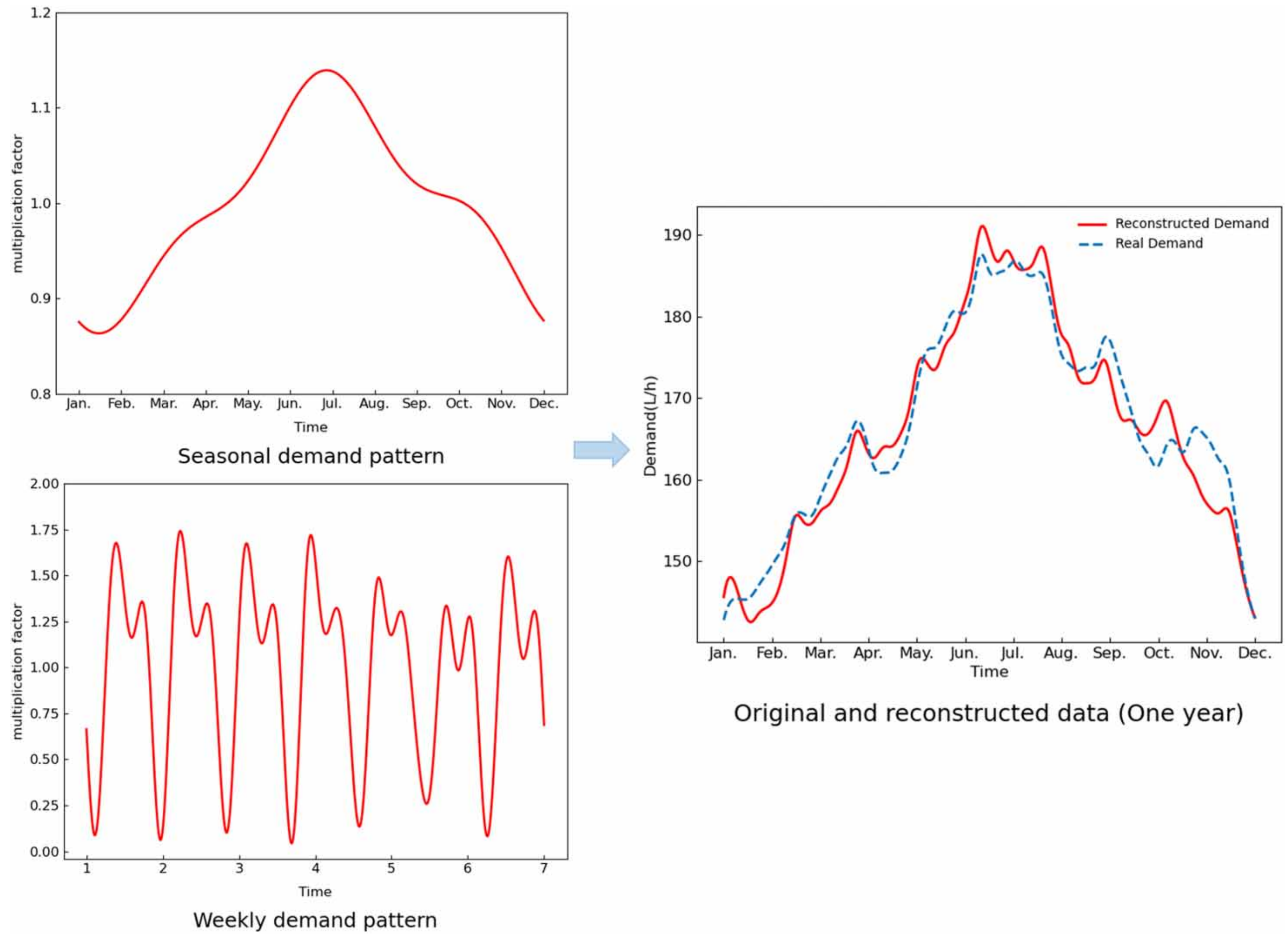

Figure 4 | The result of demand data decomposition and reconstruction based on Fourier transform.

Table 1 | Parameters of leak scenario simulation

Operating state Parameter	Normal	Leak
Time step	5 min	5 min
Duration	365 days	7 days
Number of leak scenarios per pipe	–	20 (1/20 DN–DN)
Number of potential leak pipes	–	905
Time of leak occurrence	–	1–4 am
Total number of samples	105,120	868,800

Note: DN represents the nominal diameter of the leak pipe, measured in millimeters (mm).

The hyperparameter optimization provides the identification model with the best learning rate and weight updating, which ensures that the model has the least loss and highest performance. By comparing the validation loss of VGG16 model under different settings, the final hyperparameter information adopted is shown in Table 2.

Normal images and abnormal images are imputed into the deep learning model, and then the features of the images are deeply extracted by continuous convolution and pooling operations. Finally, the full connection layer is used to classify the images into normal pictures and abnormal pictures. The model classifies the image labels of three consecutive moments

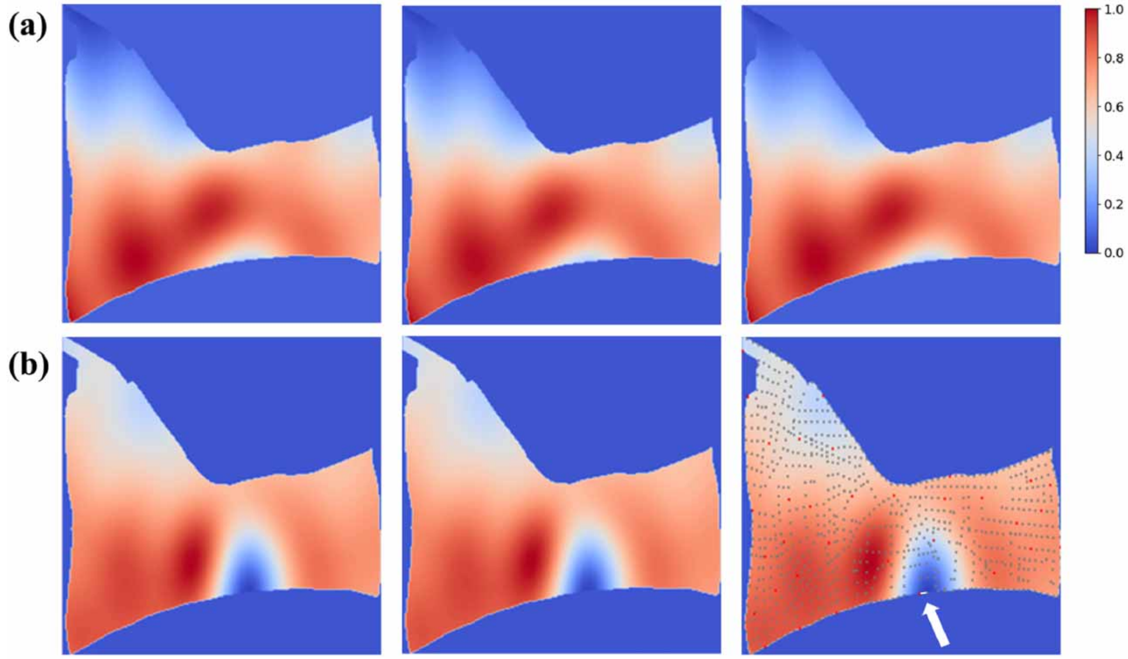

Figure 5 | Pressure distribution diagrams of the pipe network, captured consecutively at the same time under different operational conditions: (a) leak-free state and (b) leaky conditions. The white line pointed by the arrow represents the leak pipe in the network.

Table 2 | Hyperparameter values for the VGG16

Name	Optimal choice
Optimizer	SGD
Learning rate	0.001
Batch size	64
Epochs	200
Learning rate decay	0.2
Momentum	0.6
Weight decay (L2 regularization)	0.01
Decay milestones	[60,120,160]
Gamma	0.2
Loss function	Cross entropy loss

as abnormal labels, which is regarded as the leakage event of the pipe network at that moment. The loss function of the training and validation set is shown in Figure 6.

3.4. Performance evaluation

The weight and bias of the model are preserved by learning and verifying the training and verification set until the model converges. In order to comprehensively evaluate the robustness of the model, the parameter information with the best performance on the verification set is loaded on the test set for evaluation. For qualitative analysis of detection results, four metrics are used to evaluate the performance of the model: true positive rate (TPR), false positive rate (FPR), F1 score, and receiver operating characteristic (ROC) curve and its area under the curve (AUC) (Cha *et al.* 2022; Ji *et al.* 2022). The

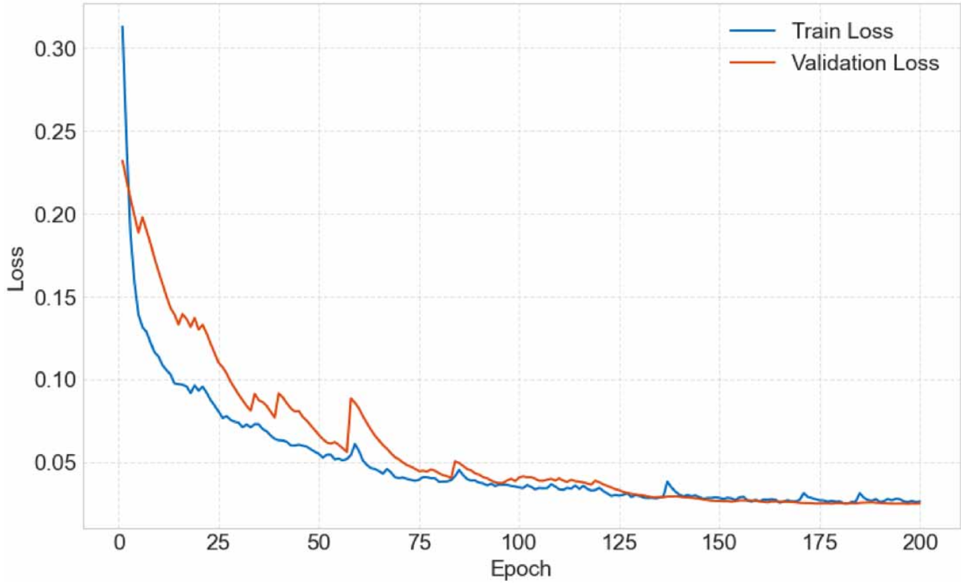

Figure 6 | Plot of the training and validation loss about VGG16.

formulas are as follows:

$$\mathrm{TPR} = \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \tag{9}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{10}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{11}$$

where TP = true positive, indicating that a leakage in the network is correctly detected; FN = false negative, indicating that a leakage in the network is not detected; TN = true negative, indicating that a leakage does not occur and is correctly not detected as an occurrence; and FP = false positive, indicating that a leakage is falsely detected as an occurrence when it did not actually occur.

To demonstrate the ability of different deep learning models for the conduct of leak detection, we compared CNN-based models like the Residual Network (ResNet18) (Li *et al.* 2022), VGG13, SqueezeNet to VGG16. The comparison models and VGG16 have the same hyperparameter values, and the experimental results are listed in Table 3. In descending order, the TPR of VGG16, VGG13, ResNet18 and SqueezeNet are 0.98, 0.83, 0.80 and 0.67. The model with the greatest TPR in leakage detection was VGG16, with a TPR of 0.98. However, we observed that the SqueezeNet model did not show significant performance in this task, which achieved a TPR of less than 0.8 and the lowest TPR. Among these CNN-based models, VGG-

Table 3 | Identification performance of different deep learning models

Models	TPR	FPR	F1	AUC
VGG16	0.98	0.01	0.94	0.94
VGG13	0.83	0.00	0.89	0.91
ResNet18	0.80	0.00	0.89	0.91
SqueezeNet	0.67	0.00	0.79	0.82

based achieved relatively high TPR. In addition, AUC is a key and comprehensive metric to evaluate model performance, it represents the area under the roc curve, ranging from 0.1 to 1, which is regarded as a numerical intuitive evaluation of the classification results. As shown in Figure 7, VGG16 achieves the maximum AUC, which means the performance of it is the best. Compared with the performance of other deep learning models, even if VGG16 has a certain FPR, it is within the acceptable range.

4. DISCUSSION

Compared with other related research methods, the leakage identification method proposed in this paper, based on a deep learning model, exhibits superior accuracy. This enhancement is attributed to the integration of spatiotemporal information from pressure monitoring data. By employing spatial interpolation techniques, limited real-time monitoring data are transformed into a continuous spatial distribution image of pressure. A well-established CNN model then extracts spatial distribution and change characteristics from these pressure images, identifying anomalies through the deployment of a specific sliding time window.

This method offers significant advantages over traditional hydraulic models due to its minimal need for preliminary tasks, such as model establishment and validation. By utilizing basic network geographic information, extensive time series data of normal operational pressure, and limited sequences of abnormal pressure, it enables the analysis of continuous spatial pressure distributions. This not only enhances real-time monitoring but also improves adaptability. In contrast, methods based on machine learning algorithms – such as Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and clustering algorithms – typically require manual configuration of feature extraction parameters. This requirement can complicate the process of learning complex features and impede the attainment of high performance.

Deep neural networks, as demonstrated by empirical evidence, tend to achieve better recognition outcomes. Rajabi *et al.* (2023) utilized a conditional convolutional generative adversarial network (CDCGAN) to produce grayscale maps of pressure distributions under normal conditions, comparing the structural similarity index (SSIM) of real-time images and the threshold images. However, the threshold value of this method is determined according to the normal pictures in the database, and the latest pressure fluctuation pattern cannot be updated according to seasonal changes, and the recognition accuracy of this method is 70%. Li *et al.* (2022) developed a leak detection model leveraging the ResNet18, designed to mitigate issues related to gradient vanishing and network depth. In our study, we conducted a comparative analysis of the VGG16 model and the ResNet18 model under identical hyperparameters. The results indicate that the TPR for the VGG16 model was 0.98, demonstrating superior performance compared to the TPR of 0.8 achieved by the ResNet18 model.

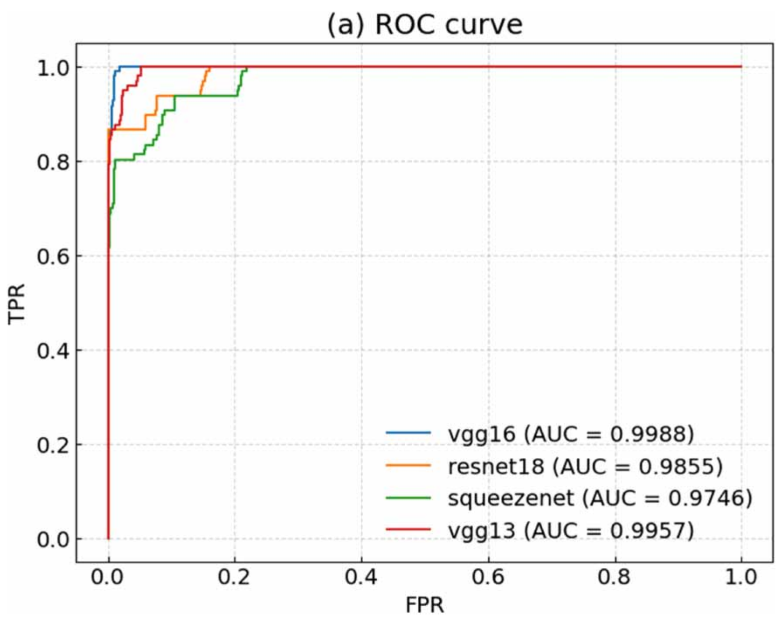

Figure 7 | Performance of four deep learning models in leak detection.

In addition to these comparisons, we also evaluated our deep learning model against other architectures such as SqueezeNet and VGG13. Our results consistently showed the superiority of the VGG16 model in leakage detection tasks, highlighting its robustness and reliability.

Despite the demonstrated advantages, our approach has not yet been tested on more complex or large-scale networks. Future work will focus on extending the application of this method to expansive urban networks and evaluating its performance under varied environmental conditions. This ongoing research will also explore the integration of real-time adaptive mechanisms to update the model according to seasonal changes in pressure patterns, thereby enhancing its accuracy and reliability.

In summary, the proposed deep learning-based leakage detection method not only outperforms traditional hydraulic models and other machine learning approaches but also offers significant potential for further optimization and application in diverse pipe network scenarios. Future research will aim to address current limitations and expand the applicability of this approach, ensuring its effectiveness in more complex and larger-scale WDNs.

5. CONCLUSION

The rapid advancement of SCADA systems has necessitated enhanced capabilities for the automatic detection and identification of pressure change signals at various monitoring points. However, traditional machine learning methods fall short in analyzing the deeper relationships within data, which curtails the accuracy of anomaly detection. In response to this limitation, this study proposes an advanced leakage detection approach for WDNs utilizing the VGG16 model. This method projects input spaces into progressively lower-dimensional latent representations hierarchically, eliminating the need for domain expertise or human supervision. Consequently, it can autonomously extract complex patterns and relationships, enhancing the detection process.

The essence of this method involves converting original pressure data into images and employing a deep learning model to accurately identify leakage events in the network. This approach effectively mitigates noise interference and demonstrates robustness. Additionally, the method for recognizing pipe network pressure images via deep learning models can be readily adapted to other large-scale pipe networks. Compared to VGG13, ResNet, and SqueezeNet, this method exhibits superior recognition performance.

Future research can focus on exploring the impact of varying the number and location of pressure sensors on data imaging and evaluating how different imaging results affect the accuracy of image-based anomaly identification. Furthermore, semantic segmentation could be utilized to visualize the decision-making process of the deep learning model, thereby facilitating the extraction of leakage location information in WDN.

ACKNOWLEDGEMENT

This work was supported by the National Key Research and Development Program of China (2023YFC3208102, 2022YFF0606905), and the Ningbo Science and Technology Plan Project (2023Z057).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Ang, W. K. & Jowitt, P. W. 2006 Solution for water distribution systems under pressure-deficient conditions. *Journal of Water Resources Planning and Management* **132** (3), 175–182. doi:10.1061/(ASCE)0733-9496(2006)132:3(175).

Cha, J.-J., Hong, S. J., Lim, S., Joo, H. J., Park, J. H., Yu, C. W., Ahn, T. H. & Lim, D.-S. 2022 An overview of near-infrared spectroscopy-intravascular ultrasound and its applications in coronary artery disease. *Journal of Cardiovascular Intervention* **1** (1), 1–6. doi:10.54912/jci.2021.0002.

Crowl, D. A. & Louvar, J. F. 2001 *Chemical Process Safety: Fundamentals with Applications*. Pearson Education, London, UK.

Fahmy, M. & Moselhi, O. 2010 Automated detection and location of leaks in water mains using infrared photography. *Journal of Performance of Constructed Facilities* **24** (3), 242–248. doi:10.1061/(ASCE)CF.1943-5509.0000094.

Fontanazza, C. M., Notaro, V., Puleo, V., Nicolosi, P. & Freni, G. 2015 Contaminant intrusion through leaks in water distribution system: Experimental analysis. *Procedia Engineering* **119**, 426–433.

Giustolisi, O., Savic, D. & Kapelan, Z. 2008 Pressure-driven demand and leakage simulation for water distribution networks. *Journal of Hydraulic Engineering* **134** (5), 626–635. doi:10.1061/(ASCE)0733-9429(2008)134:5(626).

Greyvenstein, B. & Van Zyl, J. E. 2007 An experimental investigation into the pressure – leakage relationship of some failed water pipes. *Journal of Water Supply: Research and Technology-Aqua* **56** (2), 117–124. doi:10.2166/aqua.2007.065.

Guo, S., Zhang, T., Shao, W., Zhu, D. Z. & Duan, Y. 2013 Two-dimensional pipe leakage through a line crack in water distribution systems. *Journal of Zhejiang University SCIENCE A* **14** (5), 371–376. doi:10.1631/jzus.A1200227.

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. *Journal of Water Resources Planning and Management* **147** (2), Artn 04020101. doi:10.1061/(Asce)Wr.1943-5452.0001317.

Ji, H., Kim, J., Hwang, S. & Park, E. 2022 Automated crack detection via semantic segmentation approaches using advanced U-net architecture. *Intelligent Automation & Soft Computing* **34** (1), 593–607. doi:10.32604/iasc.2022.024405.

Karim, M. R., Abbaszadegan, M. & Lechevallier, M. 2003 Potential for pathogen intrusion during pressure transients. *Journal AWWA* **95** (5), 134–146. doi:10.1002/j.1551-8833.2003.tb10368.x.

Karthikeyan, L. & Nagesh Kumar, D. 2013 Predictability of nonstationary time series using wavelet and EMD based ARMA models. *Journal of Hydrology* **502**, 103–119. doi:10.1016/j.jhydrol.2013.08.030.

Klisel, K. A., Murray, R. & Haxton, T. 2018 An Overview of the Water Network Tool for Resilience (WNTR), In: Proceedings of the 1st International WDSA/CCWI Joint Conference, Kingston, Ontario, Canada, July 23–25, 075, 8p.

LeChevallier, M. W., Gullick, R. W., Karim, M. R., Friedman, M. & Funk, J. E. 2003 The potential for health risks from intrusion of contaminants into the distribution system from pressure transients. *Journal of Water and Health* **1** (1), 3–14. doi:10.2166/wh.2003.0002.

Li, J. & Heap, A. D. 2008 *A Review of Spatial Interpolation Methods for Environmental Scientists*. Geoscience Australia, Australia.

Li, J. & Heap, A. D. 2011 A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* **6** (3–4), 228–241. doi:10/c8v843.

Li, R., Huang, H., Xin, K. & Tao, T. 2015 A review of methods for burst/leakage detection and location in water distribution systems. *Water Supply* **15** (3), 429–441. doi:10.2166/ws.2014.131.

Li, J., Zheng, W. & Lu, C. 2022 An accurate leakage localization method for water supply network based on deep learning network. *Water Resources Management* **36** (7), 2309–2325. doi:10.1007/s11269-022-03144-x.

Liemberger, R. & Wyatt, A. 2018 Quantifying the global non-revenue water problem. *Water Science and Technology: Water Supply* **19**, ws2018129. doi:10.2166/ws.2018.129.

Moors, J., Scholten, L., Van Der Hoek, J. P. & Den Besten, J. 2018 Automated leak localization performance without detailed demand distribution data. *Urban Water Journal* **15** (2), 116–123. doi:10.1080/1573062X.2017.1414272.

Muggleton, J. M., Brennan, M. J., Pinnington, R. J. & Gao, Y. 2006 A novel sensor for measuring the acoustic pressure in buried plastic water pipes. *Journal of Sound and Vibration* **295** (3–5), 1085–1098.

Pang, G., Shen, C., Cao, L. & Hengel, A. V. D. 2022 Deep learning for anomaly detection: A review. *ACM Computing Surveys* **54** (2), 1–38. doi:10.1145/3439950.

Rajabi, M. M., Komeilian, P., Wan, X. & Farmani, R. 2023 Leak detection and localization in water distribution networks using conditional deep convolutional generative adversarial networks. *Water Research* **238**, 120012. doi:10.1016/j.watres.2023.120012.

Romano, M., Kapelan, Z. & Savić, D. A. 2011 Burst detection and location in water distribution systems. In: *World Environmental and Water Resources Congress 2011*. American Society of Civil Engineers: Palm Springs, California, United States, pp. 1–10.

Romano, M., Kapelan, Z. & Savić, D. A. 2014 Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management* **140** (4), 457–467. doi:10.1061/(ASCE)WR.1943-5452.0000339.

Roshan, K. A., Tang, Z. & Guan, W. 2019 High fidelity moving Z-score based controlled breakdown fabrication of solid-state nanopore. *Nanotechnology* **30** (9), 095502.

Shao, Y., Li, X., Zhang, T., Chu, S. & Liu, X. 2019 Time-series-based leakage detection using multiple pressure sensors in water distribution systems. *Sensors* **19** (14), 3070. doi:10.3390/s19143070.

Simonyan, K. & Zisserman, A. 2015 Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015*, May 7–9, 2015, San Diego, CA, USA. Conference Track Proceedings.

Sun, C., Parellada, B., Puig, V. & Cembrano, G. 2020 Leak localization in water distribution networks using pressure and data-driven classifier approach. *Water* **12** (1), 54. doi:10.3390/w12010054.

Valizadeh, S., Moshiri, B. & Salahshoor, K. 2009 Leak Detection in Transportation Pipelines Using Feature Extraction and KNN Classification. In: Pipelines 2009, American Society of Civil Engineers, San Diego, California, United States, pp. 580–589. doi: 10.1061/41069%28360%2953.

Van Leuven, L. J. 2011 Water/Wastewater Infrastructure Security: Threats and Vulnerabilities. In RM Clark, S Hakim and A Ostfeld(Eds.), Handbook of Water and Wastewater Systems Protection, 27–46. New York, NY: Springer. doi：10.1007/978-1-4614-0189-6_2.

Vrachimis, S. G., Eliades, D. G., Taormina, R., Kapelan, Z., Ostfeld, A., Liu, S., Kyriakou, M., Pavlou, P., Qiu, M. & Polycarpou, M. M. 2022 Battle of the leakage detection and isolation methods. *Journal of Water Resources Planning and Management* **148** (12), 04022068. doi:10.1061/(ASCE)WR.1943-5452.0001601.

Wang, X., Li, J., Liu, S., Yu, X. & Ma, Z. 2022 Multiple leakage detection and isolation in district metering areas using a multistage approach. *Journal of Water Resources Planning and Management* **148** (6), 04022021. doi:10.1061/(ASCE)WR.1943-5452.0001558.

Yu, H., Zhou, H., Weng, X., Long, Z., Shao, Y. & Yu, T. 2023 An angle-based leak detection method using pressure sensors in water distribution networks. *AQUA – Water Infrastructure, Ecosystems and Society* **72** (12), 2216–2228. doi:10.2166/aqua.2023.202.

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K. & Fu, G. 2019 Deep learning identifies accurate burst locations in water distribution networks. *Water Research* **166**, 115058. doi:10.1016/j.watres.2019.115058.

Zhou, B., Lau, V. & Wang, X. 2020 Machine-learning-based leakage-event identification for smart water supply systems. *IEEE Internet of Things Journal* **7** (3), 2277–2292. doi:10.1109/JIOT.2019.2958920.

Zhou, X., Liu, S., Xu, W., Xin, K., Wu, Y. & Meng, F. 2022 Bridging hydraulics and graph signal processing: A new perspective to estimate water distribution network pressures. *Water Research* **217**, 118416. doi:10.1016/j.watres.2022.118416.

First received 24 April 2024; accepted in revised form 12 July 2024. Available online 5 August 2024