IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 September 2023, accepted 18 October 2023, date of publication 23 October 2023, date of current version 27 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3326470

RESEARCH ARTICLE

Leak Detection in Water Supply Network Using a Data-Driven Improved Graph Convolutional Network

SUISHENG CHEN[1], YUN WANG[2], WEI ZHANG[1], HAIRONG ZHANG[3], AND YUCHEN HE[3], (Member, IEEE)

[1]Hangzhou Vocational & Technical College, Fair Friend Institute of Intelligent Manufacturing, Hangzhou 310018, China
[2]Mechanical and Electrical Engineering Department, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311122, China
[3]Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University, Hangzhou 310018, China

Corresponding author: Yuchen He (yche@cjlu.edu.cn)

This work was supported in part by the Hangzhou Key Scientific Research Program of China under Grant 20212013B06, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant 2021YW18 and Grant 2021YW80.

ABSTRACT Due to the complex correlation within data collection, it is a challenging task to detect leakage in the water supply network. The Graph Convolutional Network (GCN) has recently gained significant attention in correlation research. However, most existing GCN-based models assumed that the topology of whole network should be derived from expert knowledge, which is always time-consuming and difficult to acquire. To tackle this problem, a data-driven improved graph convolutional network (IGCN) is proposed based on a self-learning fully connected association graph. Compared with traditional GCN, the proposed model can adaptively learn necessary data relationship information without accurate fixed undirected association graph. On this basis, a leakage-detection IGCN (LD-IGCN) is carried out for leakage detection. Finally, two case studies of water supply networks are utilized to demonstrate the efficiency of the proposed leakage detection method in this paper.

INDEX TERMS Leakage detection, deep learning, graph convolutional network, adaptive, correlation mining.

I. INTRODUCTION

Due to pipe aging and external stress changes, water leakage problem often occurs in water supply network, which results in additional energy and sources waste during water treatment and supply, and may cause other severe issues, such as bacterial pollution and poison contamination [1], [2]. Therefore, leakage detection for water supply network is of high significance and raises much attention [3], [4], [5].

The acoustic methods have been considered as means of detecting local leakages in water supply network [6], which are difficult for global water supply network due to the increasing complexity of the network. Furthermore, in the past decades, abundant data has been collected from water supply network, which promotes the growth of data-driven leakage detection methods [7], [8], [9]. For example, Zhang et al. [10] introduced a new method to identify the leakage zones of water distribution systems where the multiclass support vector machine was used as the leakage zoon identification model to detect possible leakages within observation data. Moreover, a real-time burst detection algorithm was proposed based on the patterns of water demand, in which supervised learning was introduced to detect the burst in the local areas of water supply distribution system [11]. Although the above algorithms can achieve good results in leakage detection, the manually designed feature extractors in those methods are hard to learn complex hidden representation [12].

Numerous deep learning (DL) algorithms have been developed and achieved great success with the advancements in computer technology. Due to the abilities of powerful nonlinear mapping and automatically discriminative feature

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

learning, the DL-based leakage detection methods have attracted much attention recently [13], [14], [15]. The most representative solution is based on AE [16] and CNN [17]. AE excels in capturing the most representative features among data through data reconstruction, while CNN can use graph convolution to learn local structural features. However, the former does not consider explicit variable relationship feature extraction, and the latter cannot learn the relationship information between non-connected variables.

It should be noted that the value of each node can be easily affected by the other nodes in water supply networks. The relationship among nodes can be depicted by the topology of the network. When a leakage occurs at one location, the value of other nodes will be changed correspondingly according to the data relationship. Sometimes, the data relationship will even be changed. Unfortunately, many existing DL methods usually concern with the measurement value information but fail in explicit relationship mining.

In recent years, as a feasible approach to learn network topology, the graph convolutional network (GCN) model is constructed based on a fixed and undirected association graph which can represent the relationship among variables. To update the value of each node, the graph convolutional operation is exploited to aggregate the features of the node and its neighbor nodes according to the association graph. The hidden layer of GCN can then be derived where the measurement value and correlation in the observation layer of all nodes are mapped into the hidden layer. Due to the ability of correlation learning, GCN and its variants have been applied in the leakage detection of industrial process [18], [19], [20]. For example, Li et al. proposed an effective intelligent fault diagnosis algorithm based on a multireceptive field graph convolutional network [21]. Chen et al. proposed a novel fault diagnosis approach that combines available measurements and prior knowledge by the GCN model [22]. Hamed et al. designed a GCN based solution for fault detection and isolation in industrial networks [23]. Moreover, Hu et al. also proposed a GCN-based scheme for complex urban distribution network fault localization [24].

Although the GCN model can learn complex data relationship through the graph convolutional operation, it still remains several problems. Firstly, it is assumed that the association graph should be fixed and undirected, which means the data relationship mining ability of the GCN model highly depends on the involved association graph. However, due to the structure of physical systems, the correlation among nodes cannot always be mutual and will be changed according to system status. Secondly, it is often difficult to acquire accurate association graph of a water supply network from its prior knowledge since knowledge accumulation and precise mechanism imitation are usually time-consuming and costly. The above problems highly limit the application of current GCN-based model in the leakage detection of water supply network.

To break the above technical bottleneck, a data-driven improved GCN model (IGCN) is presented in this paper. Compared with traditional GCN, the IGCN model is constructed based on a self-learning fully connected association graph. By this way, the structure of IGCN model is more flexible and can adaptively map necessary information of measurement value and data relationship into the hidden feature representation. On this basis, the leakage detection algorithm LD-IGCN is designed to take variable correlation into water supply network leakage detection. Finally, two water supply network experiments are conducted to demonstrate the effectiveness and superiority of the proposed algorithm. Experimental analysis reveals that the LD-IGCN outperforms traditional autoencoder (AE) [17], convolutional neural network (CNN) [18] and multi-layer perceptron graphical convolutional network (MLP-GCN) [24] model in the leakage detection of water supply network.

This paper presents the following main contributions.

1) Instead of using fixed undirected association graphs in traditional GCN, this paper introduces a new self-learning fully connected association graph in the proposed IGCN which makes it possible to adaptively explore the relationships between variables according to real situation.
2) A new fault detection model, LD-IGCN, is proposed. On the basis of the learnable association graph and back-propagation in the IGCN model, the LD-IGCN model takes the label information (normal or leakage) into the updating process of the proposed association graph, which can improve the leakage detection performance.
3) Two experiments of water supply networks are introduced to validate the effectiveness and stability of the proposed method. Experimental results show the superiority of the LD-IGCN model in the accuracy of water leakage detection.

The remainder of this paper is organized as follows. Traditional GCN model will be briefly reviewed in Section II, followed by the proposed IGCN model and corresponding leakage detection algorithm in Section III. In Section IV, the advantages of the proposed method are demonstrated, followed by a conclusion in Section V.

II. REVISIT OF GCN MODEL

As a kind of graph neural network [25], [26], [27], traditional GCN model can learn complex data relationship due to corresponding association graph and graph convolutional operation. Therefore, the association graph and graph convolutional operation in traditional GCN model will be briefly reviewed in this section.

A. THE ASSOCIATION GRAPH IN TRADITIONAL GCN MODEL

The association graph is the core component of traditional GCN model and can represent the correlation among nodes. In most of current applications, the association graph is assumed to be fixed and undirected, which considers that

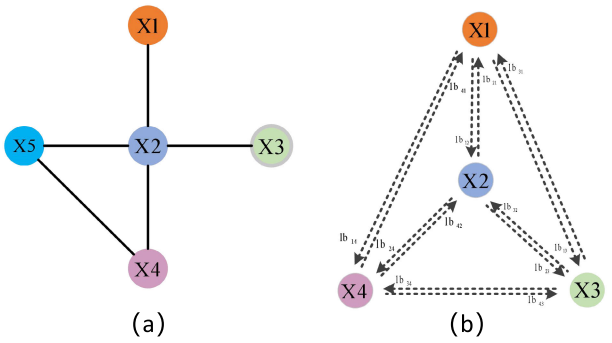


FIGURE 1. Structures of association graphs:(a) traditional fixed and undirected association graph; (b) proposed self-learning fully connected association graph.

the relationship between variables is always mutual and will never change.

The fixed and undirected association graph can be treated as $G = (W, F, B)$ where W and F represent the collection of all nodes and edges in graph G, respectively. $B \in R^{n\times n}$ is an adjacency matrix that describes the following relationships between each pair of nodes in one graph:

$$B = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix} \tag{1}$$

where n is the number of nodes. Define the edge between node i and node j as $\langle w_i, w_j \rangle$, and each element b_{ij} in matrix B can be denoted as:

$$b_{ij} = \begin{cases} 1, & \text{if edge } \langle w_i, w_j \rangle \text{ exists} \\ 0, & \text{if edge } \langle w_i, w_j \rangle \text{ does not exist} \end{cases} \tag{2}$$

For example, the adjacency matrix B of an association graph, shown in Fig.1(a), can be directly expressed as:

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{3}$$

In the above association graph, the relationship between nodes *X1* and *X2* is assumed as mutual and will never change. However, in real situation, the correlation between two nodes will change according to system operational status [28].

Moreover, it is still a challenge to establish an accurate association graph due to the lack of expert knowledge and increasing complexity of network situation. For more detail about the association graph in traditional GCN model, please refer to [29].

B. THE GRAPH CONVOLUTIONAL OPERATION IN TRADITIONAL GCN MODEL

The graph convolutional operation is another important component of traditional GCN model, which gives GCN models the capability of inter-variable correlation information learning.

The symmetric normalized graph Laplacian matrix is defined as follows:

$$L_{\text{normalize}} = K^{-\frac{1}{2}} L_{\text{original}} K^{-\frac{1}{2}} = I_N - K^{-\frac{1}{2}} B K^{-\frac{1}{2}} \tag{4}$$

where $K \in R^{n\times n}$ is the diagonal degree matrix of the graph and $K_{ii} = \sum_j b_{ij}$. I_N is an identity matrix. $L_{\text{original}} = K - B$ is the original Laplacian matrix.

The eigenvalue decomposition of $L_{\text{normalize}}$ can be expressed as:

$$L_{eign} = C \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} C = C \Lambda C^{-1} \tag{5}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix made up of the eigenvalues of L_{eign}. C is an orthonormal matrix consisting of eigenvectors of L_{eign}, and $C^{-1} = C^T$.

Then the feature updating function of one node can be defined as:

$$h = (o *_G f)_\theta = C\left(C^T o C^T f\right) \tag{6}$$

where o is the node feature. h represents the new node feature formed by the graph convolution. f is the eigenfunction of Λ. θ is a learnable coefficient. $*_G$ is denoted as the graph convolutional operation.

To simplify Equation (6), define the graph convolutional filter as $f_\theta = C^T f$, and the updating function can be rewritten as:

$$h = (o *_G f)_\theta = C\left(C^T o(Cf)\right) = C f_\theta C^T o \tag{7}$$

C. DISCUSSION

It should be noted that the performance of traditional GCN models highly depends on the association graph, which is determined by known expert experience and process mechanisms. However, the accumulation of the above two sources is always time-consuming and subjective. In addition, the traditional GCN methods always use fixed undirected association graphs. However, in practical application, the relationships between variables are much more complex. Therefore, the above issues greatly constrain the performance of traditional GCN models in water leakage detection.

III. PROPOSED METHODOLOGY

The main innovation of proposed methodology contains the following points:

1) During the training process of the IGCN model, a learnable association graph is used to capture variable relationships. Compared with traditional fixed association graph, the correlation between nodes in the proposed association graph will not be limited to non-binary values.

2) The LD-IGCN is proposed based on IGCN. The data correlation of different labels is introduced in the construction of association graph by the backpropagation, which further improve the leakage detection performance of the LD-IGCN model.

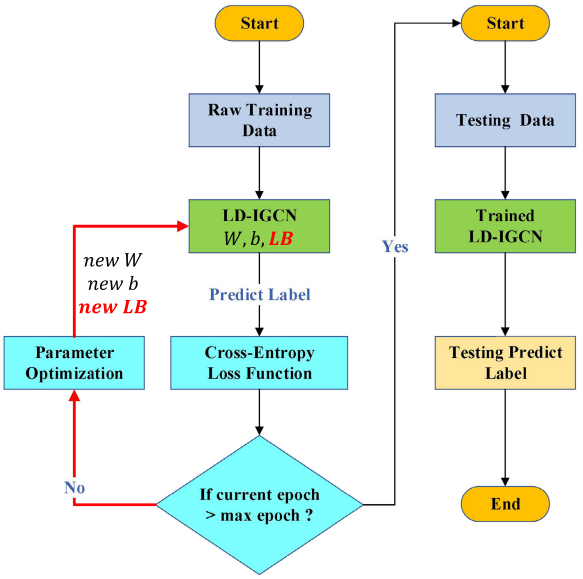


FIGURE 2. The training and testing flowchart of the LD-IGCN model.

The flowchart of the proposed methodology can be illustrated in Fig.2, which will be explained in the following subsections.

A. IMPROVED GCN MODEL (IGCN)

In this paper, to overcome the limitations in traditional GCN model caused by the fixed undirected association graph, a data-driven improved GCN model (IGCN) is proposed which is constructed on a self-learning fully connected association graph, as shown in Fig.1(b).

The proposed association graph in IGCN model is different from fixed and undirected association graphs in traditional GCN models. The advantages of the new association graph can be described as follows: Instead of one undirected edge, the IGCN contains two directed edges in opposite directions between each pair of nodes, which indicates that the model structure can be more flexible. For example, the edges between node $X1$ and node $X2$ can be defined as two directed edges $\langle w_1, w_2\rangle$ and $\langle w_2, w_1\rangle$ which represent the potential causal relationship from node $X1$ to node $X2$ and node $X2$ to node $X1$, respectively.

The definition of adjacency matrix for the new association graph can be represented by:

$$LB = \begin{bmatrix} lb_{1,1} & \cdots & lb_{1,n} \\ \vdots & \ddots & \vdots \\ lb_{n,1} & \cdots & lb_{n,n} \end{bmatrix} \tag{8}$$

where the value of $lb_{i,j}$ in Equation (8) is totally different from Equation (1). Compared with traditional graphs, the data relationship in the IGCN model is determined by data itself rather than expert knowledge. If parameter $lb_{1,2}$ or $lb_{2,1}$ is greater than zero, the edge $\langle w_1, w_2\rangle$ or $\langle w_2, w_1\rangle$ exists. Otherwise, the edge does not exist.

Based on the new association graph, the graph convolutional operation is also changed in the IGCN model. When the graph convolutional operation aggregates the node features, the parameters of the edges will be used to update the node features as follows:

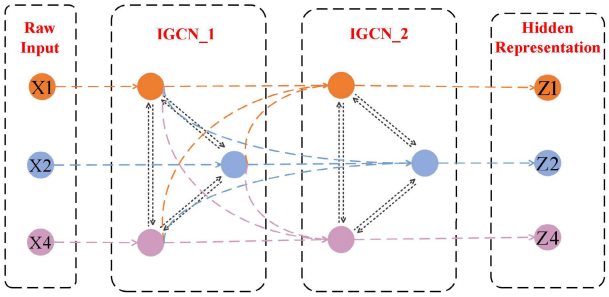


FIGURE 3. Structure of a standard two-layer IGCN model.

$$h_i^{(r+1)} = \begin{cases} \dfrac{1}{\left(\sum_{j\in N(i)} lb_{ji}\right)+1}\left[\sum_{j\in N(i)} lb_{ji} * h_j^r + h_i^r\right], r > 1 \\ x_i, r = 1 \end{cases} \tag{9}$$

where $h_i^{(r)}$ represents the i th node features of layer r and $h_i^{(r+1)}$ is the new node features in layer $r+1$. $N(i)$ is the neighbor node set of node i. x_i is the raw input data.

On this basis, the structure of a standard two-layer IGCN is given in Fig.3. The IGCN model serves as a feature extractor to learn the hidden feature, which can be defined as:

$$Z = h^{(2)}\left(\mathrm{ReLU}\left(h^{(1)}\left(X, LB^{(1)}\right)\right), LB^{(2)}\right) \tag{10}$$

where Z represents the hidden feature representation. It contains both measurement value information X and data relationship information LB. ReLU$(\cdot)$ is the activation function where $\mathrm{ReLU}(x) = \max(0, x)$.

B. LD-IGCN MODEL

1) THE STRUCTURE OF LD-IGCN

To enhance the efficiency of leak detection, we present the LD-IGCN algorithm in this section, which combines measurement values and data relationships to predict the state of the water supply system.

The main structure of LD-IGCN model is shown in Fig.4, which contains two important parts: the feature extraction network and the classification network. The input parameters of the proposed model are the water pressure data and the output of the model is the prediction label (normal or leakage) of network status.

The feature extraction network is composed of a stack of IGCN layers. When a sample is inputted into the LD-IGCN model, the feature extraction network will map the measurement value information and data relationship information into the hidden feature representation adaptively. The mapping process can be defined as:

$$Z^{(n)} = \begin{cases} X, n = 1 \\ \mathrm{Relu}\left(h^{(n-1)}\left(Z^{(n-1)}, L\,B^{(n-1)}\right)\right), n > 1 \end{cases} \tag{11}$$

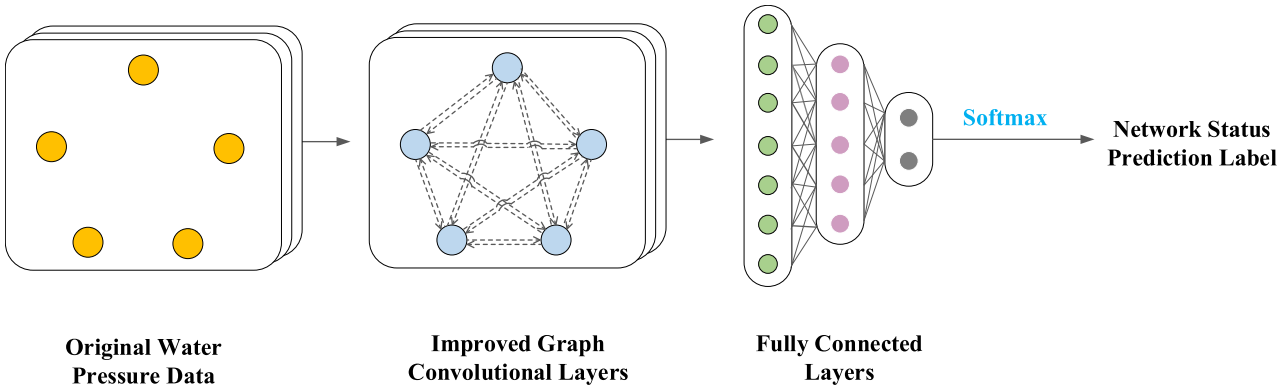


FIGURE 4. Structure of LD-IGCN.

where n is the number of layers in the feature extraction network. $Z^{(i)}$ represents the output feature of layer i and $Z^{(n+1)}$ is the output feature from the last layer of IGCN model.

The classification network is composed of a stack of fully connected layers which are typical two-class classification networks and widely used in detection situations. The extracted hidden features will be introduced into the classification network to obtain the following 2-dimensional vector $H^{(m+1)}$ indicating the score on the two types of labels (normal or leakage):

$$H^{(m)} = \begin{cases} Z^{(n+1)}, m = 1 \\ FC^{(m-1)}\left(H^{(m-1)}\right), m > 1 \end{cases} \tag{12}$$

where m represents the number of fully connected layers in the classification network. $FC^{(i)}$ represents the feature update function of fully connected layer i, and it is defined as $FC(X) = WX + b$ with input data X, weight parameter W, and bias parameter b. $H^{(i)}$ represents the output feature of the fully connected layer i in the classification network. It should be noted that $H^{(m+1)}$ will then be inputted into a softmax function to obtain the label of system status.

2) THE LD-IGCN MODEL TRAINING:

S1: Construct a self-learning fully connected association graph according to the number of input variables.

S2: Construct the LD-IGCN model based on the association graph which is constructed in S1.

S3: Input the original pressure data into the corresponding nodes of the LD-IGCN model to obtain the prediction label.

S4: Calculate the following loss value by cross-entropy function [30] based on the predicted labels and the real labels:

$$\begin{aligned} L &= \frac{1}{N}\sum_i L_i \\ &= \frac{1}{N}\sum_i -\left[y_i \cdot \log\left(p_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - p_i\right)\right] \end{aligned} \tag{13}$$

where y_i is the real label of the ith sample. p_i is the predicted label of the ith sample.

S5: Check if the current training epoch has reached the maximum preset epoch. If the maximum epoch is reached, stop training; otherwise, update the parameters of the LD-IGCN model and continue to the next epoch.

Finally, according to equation (13), it can be seen that the loss value L is related to the system label information and it will participate in the backpropagation process of model to update the model parameters layer by layer. As shown in Fig.2, adjacency matrix LB is one of the target parameters for the backpropagation process of the LD-IGCN model. Therefore, by relying on the learnability of the associated graph in IGCN and the backpropagation mechanism of the loss value, the LD-IGCN model successfully establishes the correlation between variable relationship features and system labels, thereby improving the performance of the leakage detection model. The process of updating matrix LB is as follows:

$$LB_{new} = LB_{old} - \alpha * \frac{dL}{dLB} \tag{14}$$

where α is the learning rate.

IV. CASE STUDY

In this section, the proposed method is programmed by Pytorch 1.3 machine learning framework. The programming environment is Windows 10 with Python 3.7.

Three conventional models, namely, the AE model [17], the CNN model [18], and the MLP-GCN model [24], are utilized for comparative analysis. The selected AE model contains two hidden layers and one leakage detection classifier, where each hidden layer will be pre-trained by stacked autoencoder (SAE). The CNN-based leakage detection model consists of two CNN blocks and a classifier. The selected MLP-GCN model is a traditional GCN-based model, which is established with two GCN layers and a classifier.

Moreover, to evaluate the performance of different models, accuracy(Accuracy) is selected as the indicator, which are defined as follow:

$$\text{Accuracy} = \frac{LL + NN}{LL + LN + NN + NL} \tag{15}$$

where LL and LN represent the number of leakage samples that are predicted as leakage and normal, respectively. NL

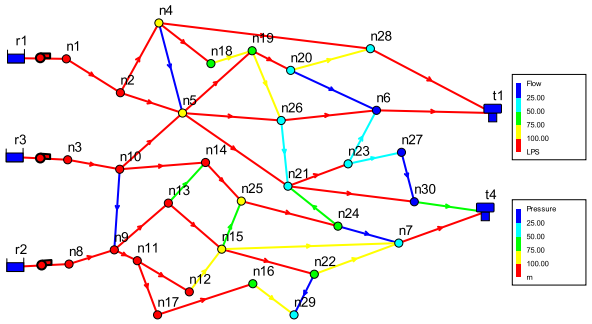


FIGURE 5. Structure of EPANET water supply network.

and NN represent the number of normal samples that are predicted as leakage and normal, respectively.

A. CASE A

The experimental data of this case is generated by the Environmental Protection Agency's Water Distribution Network Analysis Model (EPANET) benchmark simulation platform [31], which is an industrial analyzing software commonly utilized for the simulation of urban water supply systems. According to an actual water supply network in Hangzhou city, the structure of the hydraulic model is designed as Fig.5. The model comprises 30 user nodes, 48 water supply pipelines, 3 reservoirs, 3 pumps, and 2 tanks. The basic demand of the user nodes is assigned from 13.5 L/s to 99.1 L/s, while the diameter of pipelines between nodes ranges from 250 mm to 500 mm.

The node leakage in EPANET is simulated through the utilization of the hydraulic simulation tool library water network tool for resilience (WNTR) [32], [33]. The leakage model, which is provided by WNTR, is defined as follows:

$$\begin{aligned} e_{\text{leak}} &= D_d B p^{\alpha} \sqrt{\frac{2}{\rho}} \\ e_{\text{leak}} &= D_d B \sqrt{2gh} \quad \text{when } \alpha = 0.5 \end{aligned} \tag{16}$$

where e_{leak} is the leak demand $\left(\text{m}^3/\text{s}\right)$. B represents the area of the hole $\left(\text{m}^2\right)$. α represents the leakage degree coefficient (unitless). p is the gauge pressure (Pa). h, g and ρ represent the gauge head (m), acceleration of gravity $\left(\text{m}/\text{s}^2\right)$ and density of the fluid $\left(\text{kg}/\text{m}^3\right)$, respectively. D_d is the discharge coefficient (unitless).

It should be noted that the EPANET hydraulic model is driven by water pressure, where the leakage will be reflected in the pressure values of the nodes as well as the relationship with their adjacent nodes. We design three different leaks including the large, middle and small leaks, which are realized by setting the adjustable coefficient α to 0.75, 0.45 and 0.15, respectively. Fig.6 shows the changes in the pressure curve of the node n4 under normal and three leakage levels.

In addition, two leakage modes are designed including a single-node leakage (node n4) and multiple-node leakage (nodes n5, n13 and n21) with large, middle and small leakage degrees, which are shown in Table 1.

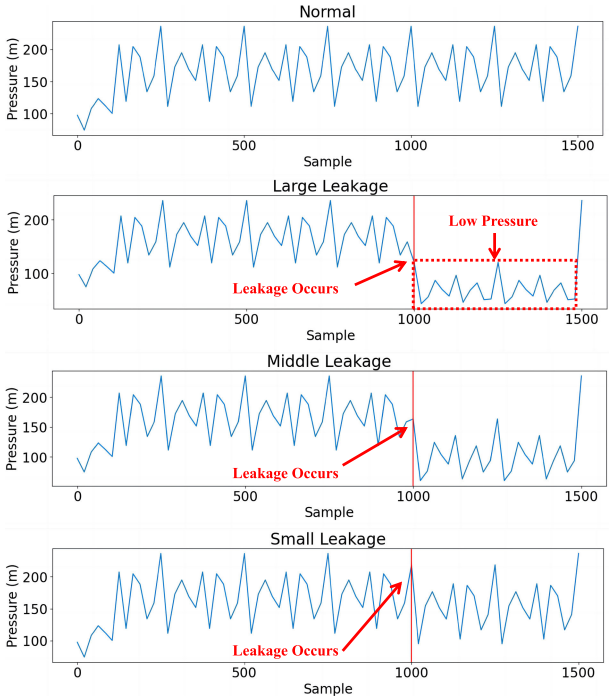


FIGURE 6. The pressure curves of nodes under normal and different leakage conditions.

TABLE 1. Different situations of leakage sources.

Leakage type	Leakage nodes	coefficient α / leakage degree
1	n4	0.75/ Large
2	n5, n13, 21	0.75/ Large
3	n4	0.45/ Middle
4	n5, n13, n21	0.45/ Middle
5	n4	0.15/ Small
6	n5, n13, n21	0.15/ Small

We independently gathered training and testing pressure data for each leakage type to substantiate the effectiveness and merits of our proposed model. In each leakage type, 1500 samples are collected for training of which the first 1000 are normal samples and the rests are leakage samples. Meanwhile, another 700 test samples are collected under the same condition, where the first 500 samples are normal and the remaining 200 samples are leakage samples.

An appropriate LD-IGCN structure is paramount for successful leakage detection. The over-smoothing problem caused by deep graph convolutions leads to indistinguishable node features. Consequently, we propose an LD-IGCN model for the current experiment, containing two IGCN layers, one fully connected (FC) layer, and a softmax layer. Because of the advantages in training speed, gradient vanishing problem and model generalization, the ReLU function is selected as the activation function which will be used in the hidden layers of LD-IGCN model.

The number of neurons in both IGCN layers and the softmax layer of the LD-IGCN model are determined based

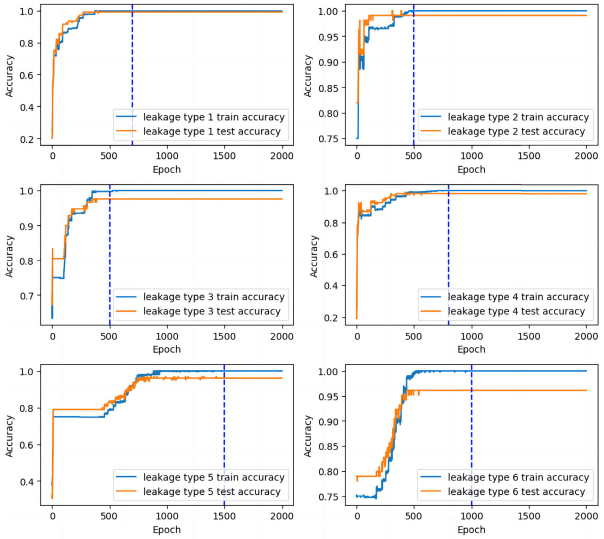


FIGURE 7. Accuracy of training and testing process in case A.

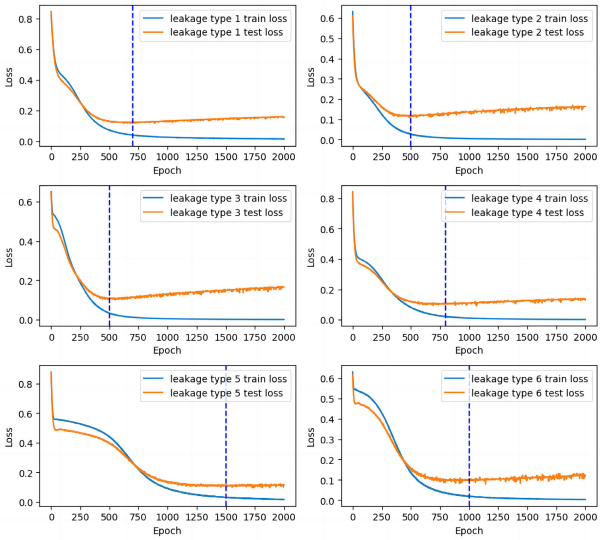


FIGURE 8. Loss of training and testing process in case A.

on the variable dimension of the input data (set as 30) and the number of classification categories (set as 2), respectively. By adjusting the number of neurons in the FC hidden layer and observing corresponding loss values, we find that the model can reach minimum loss results in all six sub-experiments when the number of neurons in the hidden layer is set to 25.

Finally, due to the capability of adaptive learning rate adjustment, the Adam algorithm is chosen as the optimizer for the LD-ICGN model, which can provide a better and more efficient parameter optimization path. The initial learning rate and training epochs are set to 0.01 and 2000, respectively. To avoid overfitting, it is necessary to choose an appropriate training epoch. Fig.7 and Fig.8 illustrate the training and testing process in six sub-experiments. The model is considered to be overfitting when the loss curve of the testing dataset starts to rise and the loss curve of the training set still goes down. According to Fig.8, the optimal training epochs for the six leakage types are 700, 500, 500, 800, 1500, and 1000, respectively. Therefore, the hyperparameters of the LD-IGCN model in this case can be represented in Table 2.

Table 3 and Fig.9 show the results of water leakage detection in case A. The large degree water leakage can be easily detected by all methods. However, when water leakage degree decreases, the LD-IGCN model outperforms the other methods in middle and small degree leakage. Compared with AE model, the proposed approach can explicitly mine the correlation between variables and map it to the hidden representation. Therefore, the model can comprehensively identify the changes of both variable values and their correlation to evaluate the network state, which benefits the water leakage detection performance. Compared to CNN, LD-IGCN is more suitable for non-Euclidean data such as water supply network data. Hence, the proposed method can exhibit superior performance on water leakage detection.

TABLE 2. Hyperparameters of LD-IGCN model in case A.

Description	Value
Neuron number in first IGCN layer	30
Neuron number in second IGCN layer	30
Neuron number in FC hidden layer	25
Neuron number in softmax layer	2
Activation function	Relu
Epochs of leakage type 1	700
Epochs of leakage type 2	500
Epochs of leakage type 3	500
Epochs of leakage type 4	800
Epochs of leakage type 5	1500
Epochs of leakage type 6	1000
Optimizer	Adam
Learning rate	0.01

Moreover, the MLP-GCN model is introduced for comparison which is established according to the fixed association graph in Fig.5. The experimental results of MLP-GCN model show inferior detection performance since the correlation among different variables is fixed, which is unsuitable for complex water supply network. In addition, water leakage will severely affect the correlation among nodes such as water pressure. However, the correlation in MLP-GCN is predetermined and cannot be modified according to real data collection. Even worse, the detection results will deteriorate as the number of leakage sources increases. Comparatively, the proposed method can adaptively learn the correlation among variables and thereby achieve better leakage detection performance.

To compare the computational efficiency of the proposed LD-IGCN model with the other models, The average time consumption of all models on the testing dataset is shown in Table 4. Due to the additional graph convolutional operations of GCN-based model, the computational time costs of the

TABLE 3. Detection results of case A.

Leakage type		1	2	3	4	5	6
AE	Training Accuracy (%)	100.0	100.0	99.1	99.4	98.6	98.9
	Testing Accuracy (%)	99.1	99.1	96.9	97.6	94.2	95.3
CNN	Training Accuracy (%)	100.0	100.0	99.3	99.5	99.2	99.3
	Testing Accuracy (%)	99.1	99.1	96.7	97.8	93.5	94.2
MLP-GCN	Training Accuracy (%)	100.0	100.0	99.4	99.6	98.9	99.2
	Testing Accuracy (%)	98.9	99.1	96.2	96.9	94.5	95.4
LD-IGCN	Training Accuracy (%)	100.0	100.0	99.6	99.8	99.4	99.7
	Testing Accuracy (%)	99.1	99.1	**97.3**	**98.1**	**96.1**	**96.7**

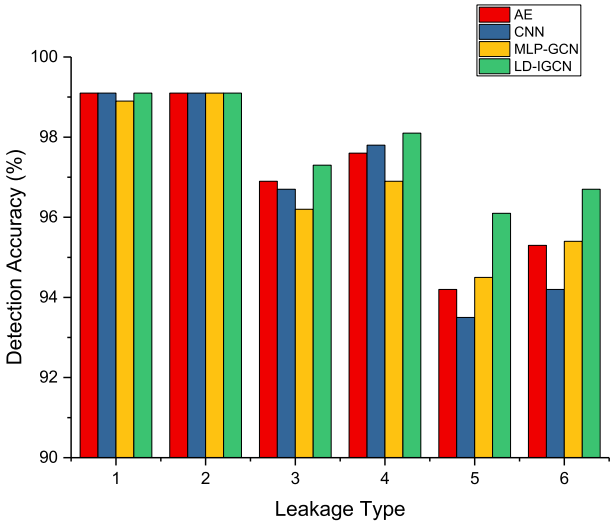


FIGURE 9. Leakage detection accuracy under different leakage types in case A.

MLP-GCN and LD-IGCN models are slightly higher than the other two models, which still satisfies the real-time demand of online leakage detection.

B. CASE B

To further validate the stability and reliability of the proposed method, the water pressure data, collected from the water supply network of a residential area in Taizhou city, is employed which includes 42 water pressure monitoring nodes. In the data collection of case B, a large leakage and a small leakage are found which are caused by pipe rupture and pipe corrosion, respectively. In leakage type, the training dataset consists of 1000 normal samples and 500 corresponding leakage samples while the testing dataset includes 500 normal samples and 200 corresponding leakage samples.

TABLE 4. Time cost used for different models in Case A.

Time(s)	AE	CNN	MLP-GCN	LD-IGCN
Leakage type 1	0.1527	0.1882	0.2158	0.2048
Leakage type 2	0.1543	0.1872	0.2133	0.2023
Leakage type 3	0.1530	0.1810	0.2165	0.2046
Leakage type 4	0.1512	0.1860	0.2173	0.2042
Leakage type 5	0.1552	0.1858	0.2156	0.2036
Leakage type 6	0.1566	0.1876	0.2177	0.2052

TABLE 5. Different leakage types in case B.

Leakage type	Description
Large	Pipe rupture
Small	Pipe corrosion

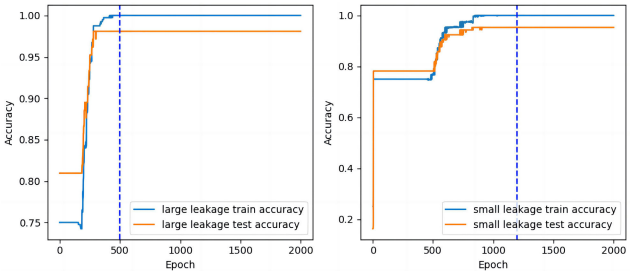


FIGURE 10. Accuracy of training and testing process in case B.

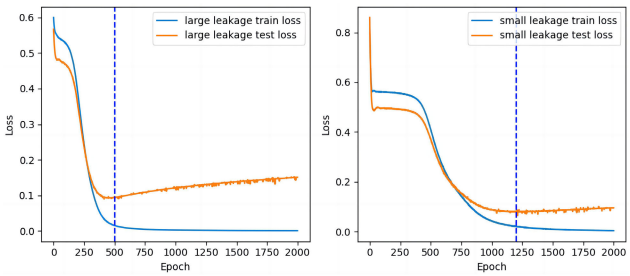


FIGURE 11. Loss of training and testing process in case B.

Similar to case A, the number of neurons in both IGCN layers is equal to the number of nodes and set to 42. The neuron number in the softmax layer is determined by the number of categories 2. Moreover, the optimum neuron number in the FC layer is 35.

The Adam algorithm is also chosen for model parameter optimization. The beginning learning rate and training epochs are still set as 0.01 and 2000, respectively. The training and testing process can be shown in Fig.10 and Fig.11 where the optimal training epochs for leakage type 1 and leakage type 2 are determined as 500 and 1200, respectively. Therefore, the hyperparameters of the proposed method in case B can be represented in Table 5.

The detection results of all models are shown in Table 7 and Fig.12. The influence of large leakages caused by pipe rupture on the water pressure values at the detection points is evident. Therefore, all models achieve excellent results in the large leakage detection experiment. Compared to the AE and

TABLE 6. Hyperparameters of LD-IGCN model in case B.

Description	Value
Neuron number in first IGCN layer	42
Neuron number in second IGCN layer	42
Neuron number in FC hidden layer	35
Neuron number in softmax layer	2
Activation function	Relu
Epochs of leakage type 1	500
Epochs of leakage type 2	1200
Optimizer	Adam
Learning rate	0.01

TABLE 7. Detection results of case B.

Leakage type		Large	Small
AE	Training Accuracy (%)	98.5	97.6
	Testing Accuracy (%)	98.1	95.4
CNN	Training Accuracy (%)	98.9	98.2
	Testing Accuracy (%)	97.8	94.8
MLP-GCN	Training Accuracy (%)	99.1	98.6
	Testing Accuracy (%)	97.6	94.3
LD-IGCN	Training Accuracy (%)	99.1	98.7
	Testing Accuracy (%)	**98.1**	**96.0**

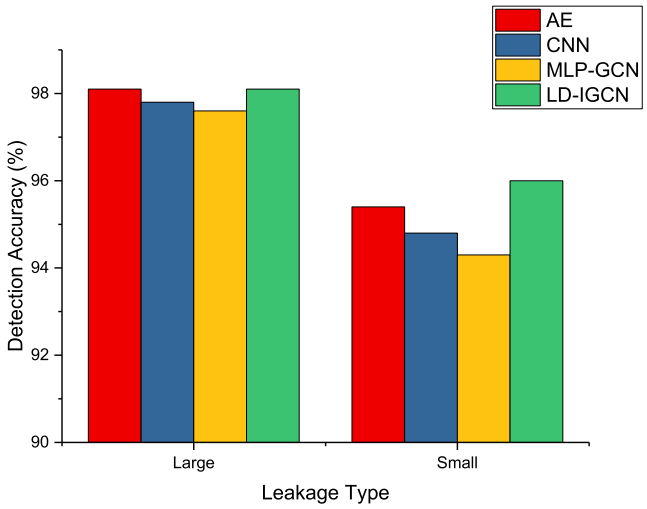


FIGURE 12. Leakage detection accuracy under different leakage types in case B.

TABLE 8. Time cost used for different models in Case B.

Time(s)	AE	CNN	MLP-GCN	LD-IGCN
Large	0.1757	0.2396	0.3291	0.3148
Small	0.1765	0.2348	0.3247	0.3172

CNN models, the proposed method is capable of discerning the differences in variable relationships between normal and leakage data, thus leading to superior performance. The learned relationships between variables in MLP-GCN reflect the system physical structure. However, the introduced graph prior only considers the physical structure and neglects the ruptures in the pipeline. Consequently, the performance of the MLP-GCN model is weaker than the proposed method in this experiment.

The impact of small leakage caused by corrosion is less pronounced compared to large leakage. Consequently, the detection accuracy of all methods decreases. It is worth highlighting that the LD-IGCN model still outperforms the other three methods. Table 8 shows the average time consumption of all models on the actual water pressure monitoring testing data. Compared with the other methods, the proposed method does not cost much extra time while improving the leakage detection performance.

V. CONCLUSION

This paper proposes a leakage detection algorithm for water supply network based on IGCN model. Compared with traditional GCN methods, both measurement value information and data relationship information are well considered in the proposed algorithm. Moreover, due to its unique self-learning fully connected association graph, the proposed method outperforms several state-of-the-art methods in leakage detection rate. It is noticed that the water demand of each node in the water supply network changes over time, which is not considered in this paper. Therefore, the data autocorrelation will be discussed in our feature work.

REFERENCES

[1] Z. Zhao, M. B. Ray, T. Mao, and W. Sun, "Impact of UV irradiation on disinfection by-product formation and speciation from post-chlorination of dissolved organic matter," *J. Water Supply, Res. Technol.-Aqua*, vol. 70, no. 8, pp. 1181–1191, Dec. 2021.

[2] O. Golovko, L. D. B. Anton, C. Cascone, L. Ahrens, E. Lavonen, and S. J. Köhler, "Sorption characteristics and removal efficiency of organic micropollutants in drinking water using granular activated carbon (GAC) in pilot-scale and full-scale tests," *Water*, vol. 12, no. 7, p. 2053, 2022.

[3] W. Nimri, Y. Wang, Z. Zhang, C. Deng, and K. Sellstrom, "Data-driven approaches and model-based methods for detecting and locating leaks in water distribution systems: A literature review," *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11611–11623, Jun. 2023.

[4] B. Bakhtawar and T. Zayed, "Review of water leak detection and localization methods through hydrophone technology," *J. Pipeline Syst. Eng. Pract.*, vol. 12, no. 4, Nov. 2021, Art. no. 03121002.

[5] H. Fan, S. Tariq, and T. Zayed, "Acoustic leak detection approaches for water pipelines," *Autom. Construct.*, vol. 138, Jun. 2022, Art. no. 104226.

[6] B. Van Hieu, S. Choi, Y. U. Kim, Y. Park, and T. Jeong, "Wireless transmission of acoustic emission signals for real-time monitoring of leakage in underground pipes," *KSCE J. Civil Eng.*, vol. 15, no. 5, pp. 805–812, May 2011.

[7] X. Liu, B. Cai, X. Yuan, X. Shao, Y. Liu, J. Akbar Khan, H. Fan, Y. Liu, Z. Liu, and G. Liu, "A hybrid multi-stage methodology for remaining useful life prediction of control system: Subsea Christmas tree as a case study," *Expert Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119335.

[8] C. Yang, B. Cai, Q. Wu, C. Wang, W. Ge, Z. Hu, W. Zhu, L. Zhang, and L. Wang, "Digital twin-driven fault diagnosis method for composite faults by combining virtual and real data," *J. Ind. Inf. Integr.*, vol. 33, Jun. 2023, Art. no. 100469.

[9] X. Kong, B. Cai, Y. Liu, H. Zhu, Y. Liu, H. Shao, C. Yang, H. Li, and T. Mo, "Optimal sensor placement methodology of hydraulic control system for fault diagnosis," *Mech. Syst. Signal Process.*, vol. 174, Jul. 2022, Art. no. 109069.

[10] Q. Zhang, Z. Y. Wu, M. Zhao, J. Qi, Y. Huang, and H. Zhao, "Leakage zone identification in large-scale water distribution systems using multiclass support vector machines," *J. Water Resour. Planning Manage.*, vol. 142, no. 11, Nov. 2016, Art. no. 04016042.

[11] P. Huang, N. Zhu, D. Hou, J. Chen, Y. Xiao, J. Yu, G. Zhang, and H. Zhang, "Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning," *Water*, vol. 10, no. 12, p. 1765, Dec. 2018.

[12] J. Li, W. Zheng, and C. Lu, "An accurate leakage localization method for water supply network based on deep learning network," *Water Resour. Manage.*, vol. 36, no. 7, pp. 2309–2325, May 2022.

[13] H. M. Tornyeviadzi and R. Seidu, "Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106062.

[14] M. Zhou, Y. Yang, Y. Xu, Y. Hu, Y. Cai, J. Lin, and H. Pan, "A pipeline leak detection and localization approach based on ensemble TL1DCNN," *IEEE Access*, vol. 9, pp. 47565–47578, 2021.

[15] C.-W. Lee and D.-G. Yoo, "Development of leakage detection model and its application for water distribution networks using RNN-LSTM," *Sustainability*, vol. 12, no. 16, p. 9262, Aug. 2021.

[16] Z. Yuan, Z. Ma, X. Li, and J. Li, "A multichannel MN-GCN for wheelset-bearing system fault diagnosis," *IEEE Sensors J.*, vol. 23, no. 3, pp. 2481–2494, Feb. 2023.

[17] Q. Jiang, X. Yan, and B. Huang, "Deep discriminative representation learning for nonlinear process fault detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1410–1419, Jul. 2020.

[18] Y.-L. Tsai, H.-C. Chang, S.-N. Lin, A.-H. Chiou, and T.-L. Lee, "Using convolutional neural networks in the development of a water pipe leakage and location identification system," *Appl. Sci.*, vol. 12, no. 16, p. 8034, Aug. 2022.

[19] J. Man, H. Dong, L. Jia, and Y. Qin, "AttGGCN model: A novel multi-sensor fault diagnosis method for high-speed train bogie," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19511–19522, Oct. 2022.

[20] H. Tong, R. C. Qiu, D. Zhang, H. Yang, Q. Ding, and X. Shi, "Detection and classification of transmission line transient faults based on graph convolutional neural network," *CSEE J. Power Energy Syst.*, vol. 7, no. 3, pp. 456–471, May 2021.

[21] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Multireceptive field graph convolutional networks for machine fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12739–12749, Dec. 2021.

[22] Z. Chen, J. Xu, T. Peng, and C. Yang, "Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9157–9169, Sep. 2022.

[23] J.-A. Miguel, H.-A. Javier, Y. M. Armando, and J. R. Matthew, "Embedded, real-time, and distributed traveling wave fault location method using graph convolutional neural networks," *Energies*, vol. 15, no. 20, p. 7785, Oct. 2022.

[24] Y. Hu, J. Chen, H. Ye, Y. Liu, J. Zhang, and Y. Li, "Fault location method of complex urban distribution network based on MLP-GCN," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (I&CPS Asia)*, Jul. 2022, pp. 1580–1585.

[25] Y. Zhou, H. Zheng, X. Huang, S. Hao, D. Li, and J. Zhao, "Graph neural networks: Taxonomy, advances, and trends," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 1, pp. 1–54, Feb. 2022.

[26] S. Xiao, S. Wang, Y. Dai, and W. Guo, "Graph neural networks in node classification: Survey and evaluation," *Mach. Vis. Appl.*, vol. 33, no. 1, p. 4, Jan. 2022.

[27] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[28] J. Yuan, M. Cao, H. Cheng, H. Yu, J. Xie, and C. Wang, "A unified structure learning framework for graph attention networks," *Neurocomputing*, vol. 495, pp. 194–205, Jul. 2022.

[29] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 119–131, Jan. 2020.

[30] J. Zhang and H. Matsuzoe, "Entropy, cross-entropy, relative entropy: Deformation theory (a)," *Europhys. Lett.*, vol. 134, no. 1, p. 18001, Apr. 2021.

[31] J. B. Burkhardt, H. Woo, J. Mason, F. Shang, S. Triantafyllidou, M. R. Schock, D. Lytle, and R. Murray, "Framework for modeling lead in premise plumbing systems using EPANET," *J. Water Resour. Planning Manage.*, vol. 146, no. 12, Dec. 2020, Art. no. 04020094.

[32] S. Kumar, S. K. Bharti, and N. Kumar, "Diurnal and seasonal variation in morphology and elemental composition of particulate matters," *J. Geolog. Soc. India*, vol. 99, no. 5, pp. 666–674, May 2023.

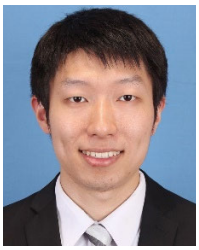
[33] K. A. Klise, M. Bynum, D. Moriarty, and R. Murray, "A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study," *Environ. Model. Softw.*, vol. 95, pp. 420–431, Sep. 2017.

SUISHENG CHEN is currently a Professor with the Hangzhou Vocational & Technical College, Fair Friend Institute of Intelligent Manufacturing. His research interest includes control and fault diagnosis in water supply networks.

YUN WANG is currently a Faculty Member of the Mechanical and Electrical Engineering Department, Zhejiang Tongji Vocational College of Science and Technology. She mainly focuses on the process monitoring for industrial networks.

WEI ZHANG received the Ph.D. degree from Zhejiang University, in 2014. He was with the Hangzhou Vocational & Technical College, where he is currently a Lecturer with the Fair Friend Institute of Intelligent Manufacturing. His research interests include data collection and analysis of industrial processes.

HAIRONG ZHANG is currently pursuing the master's degree with the School of Mechanical and Electrical Engineering, China Jiliang University. His research interests include deep learning algorithm, leakage detection, and leakage localization and applications in water supply networks.



YUCHEN HE (Member, IEEE) was with China Jiliang University, in 2017, where he is currently an Associate Professor with the College of Mechanical and Electrical Engineering. He mainly focuses on industrial process monitoring and soft sensing technique.