

Leakage Location Method of Water Supply Pipe Network Based on Integrated Neural Network

Xiaoxuan Wu*

School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601, China
Anhui Engineering Lab of Big Data Technology
Application for Urban Infrastructure, Hefei University, Hefei, 230601, China
kexinyufan@163.com

Chen Zhang

School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601, China
Anhui Engineering Lab of Big Data Technology
Application for Urban Infrastructure, Hefei University, Hefei, 230601, China

Abstract—Concerning the problem that selecting the monitoring points and locating the leakage points accurate in the water supply network, a method of locating leakage points based on integrated neural network model is proposed. Based on the similarity of water supply pipe network node pressure, a fuzzy C-means clustering fusion algorithm (FCMCF) was used to select the limited monitoring points, instead of randomly selecting monitoring points; on this basis, according to the distribution structure of water supply pipe network, the hydraulic simulation was carried out by using EPANET pipe network adjustment software, and the water pressure of all nodes before and after leakage was obtained 24hours, the integrated neural network combining the convolutional neural network (CNN) and long and short time memory network (LSTM) was used to locate the leakage points of water supply pipe network. In the experimental part, the application of a model LangXi County Xuan City, an area of underground water supply pipe network leakage problem, through FCMCF selected five monitoring stations, using EPANET software simulate the 396 kinds of fault state, through the integrated neural network model training and testing, finally realizes the accurate positioning of the leak, model accuracy reached 90.25% and loss rate of the model is up to 20.24%. Compared with the empirical mode decomposition (EMD) combined with LSTM structure, the accuracy of the model is significantly improved, while the loss rate is significantly reduced. The experimental results show that the leakage location method of water supply pipe network based on integrated neural network can effectively solve the problem of actual leakage location.

Keywords—CNN, LSTM, FCMCF, leakage location

I. INTRODUCTION

In the contemporary social production and life, the urban water supply network is one of the important infrastructure, which plays a pivotal role in ensuring the sustainable economic development and the normal life of residents. Because the urban water supply network is buried under the surface, the leakage failure is hidden. Therefore, it is particularly important to master the operation state of water supply pipe network system. Pipeline leakage is not necessarily caused by a single cause, but may be caused by a combination of many reasons (pipe quality problems, interface problems, construction problems, pipeline corrosion and external force damage). Therefore, pipeline leaks and its characteristic parameters change affected by various factors, in fact, the actual water supply network model is a typical nonlinear complex system, at the same time due to pipe water pressure, flow, often change and the effects of the medium in the pipeline, resulting in the operation of the network model

is set up when needed physical parameters are difficult to determine accurately, cause is difficult to establish accurate model to describe the actual network running state [1].

In the early days of China, the leakage location of water supply network mainly adopted the methods of "manual detection", such as acoustic listening method and negative pressure wave method [2] and so on. However, the cost of manual detection is high and the results are not satisfactory. Then, leakage detection and location based on mathematical model are proposed. Du et al [3] uses to flow monitoring points in the water supply network to obtain a large amount of flow data. On this basis, the feasibility and accuracy of t-test analysis method and mean difference analysis method of the practical application of leakage location is studied respectively. With the increasing amount of monitoring data onto water supply networks, machine learning techniques such as support vectors machines [4,5], deep learning [6], artificial neural networks [7] and other algorithms are also applied to the leakage location of water supply network. Liu [8] establishes a probabilistic neural network model and uses the probabilistic neural network to locate pipe network leakage of actual water plants. In [9], the BP neural network deep learning method was proposed to predict the location of leakage points and construct the leakage simulation model of water supply pipe network. Although some methods were used in the above study to detect and locate the leakage of water supply network, the positioning accuracy was not high, and the selection of monitoring points was random and not representative. In fact, the selection of different monitoring points has an impact on leakage location.

In reality, for the urban underground water supply network, on the one hand, because it is underground, real-time monitoring is very difficult, if the sensor equipment is installed to monitor the pipe network, the cost is too huge; On the other hand, the number of pipe segments and nodes is huge, and the flow of pipe network nodes changes with time, so the calculation amount of pipe network adjustment is too large. Therefore, it is very important to the supervision of urban underground water supply network to understand the operation condition of pipe network system scientifically. In order to solve the above problems, this paper makes the following contributions: (1) FCMCF is proposed based on the Fuzzy c-means Algorithm. By running the fuzzy C-means algorithm for several times, the results of multiple clustering are obtained, and then the weighted coco-matrix is used to fuse the members of the cluster to obtain the final clustering results. The representative nodes in the pipe network are

selected as the pressure monitoring points of limited positions by FCMCF. (2) An integrated neural network model based on deep learning is constructed, which combines the CNN with the LSTM to locate the leakage points of the water supply network. Considering the influencing factors of locating the leakage point, the pressure values of the monitoring points of different time points were constructed into time series data, and the time correlation between different time pressure values were mined by LSTM. Non-time series data onto pipe section length, pipe section diameter and water to demand were constructed, and the correlation between pipe sections of each pipe network was extracted by CNN.

II. MONITORING POINT SELECTION BASED ON FCMCF

Monitoring the water pressure (or flow) at every node in a network system is costly and difficult to process. Therefore, it is of certain practical significance to think whether the operation status of the whole pipe network can be understood by monitoring the pressure or flow of nodes at finite positions in the pipe network system. How to select finite nodes as monitoring points to represent the operation status of the whole pipe network is an urgent problem to be solved. Ideas of this article are from a node pressure in water supply pipe (or flow) similarity analysis, all nodes in the network according to its pressure (or flow) change degree, high degree of similarity is divided into the same class, in the same class similar or similar changes between the node, and pipe network pressure and flow of each node in the system software can be through the computer network adjustment is convenient, rapid and accurate calculation. The selection of limited monitoring points is realized by clustering results.

Clustering fusion refers to the fusion of existing clustering results from consensus function design, so as to maximize the shared information on existing clustering results, so as to obtain more accurate and stable mining results than a single clustering algorithm. Multiple clustering results is obtained by using the fuzzy C-means algorithm, and the members of the cluster are fused by the fusion algorithm to get the final clustering results.

Definition 1. (weighted coco-matrix [10]) defines the weighted coco-matrix Co , where each element $Co(i, j)$ is expressed in formula (1):

$$Co(i, j) = \sum_{p=1}^{H'} \frac{g(i, j, p) \omega_p}{H'} \quad (1)$$

$$g(i, j, p) = \begin{cases} 1, i \in C_p^m \text{ and } j \in C_p^m, m = 1, 2, \dots, k. \\ 0, i \notin C_p^m \text{ or } j \notin C_p^m \end{cases} \quad (2)$$

Where $g(i, j, p) = 1$ represents that data point i and j belong to a cluster C_p^m in a selected cluster member l_p' ; Otherwise, $g(i, j, p) = 0$. m represents the number of clusters, p represents the number of selected cluster members, and ω_p represents its corresponding weight.

The procedure for selecting pressure monitoring points of water supply network based on FCMCF is as follows:

Step 1. The pressure value (or flow value) of each node in the water supply network of different periods (peak period, low peak period and general average period) is calculated respectively. The pressure value is selected as the measurement standard of this paper, which is the same as the experimental result of flow values selection.

Step 2. Calculate the interaction matrix among each node according to formula (3):

$$X(i, j) = \frac{H_i - H'_i}{H_j - H'_j}, (i, j = 1, 2, \dots, n). \quad (3)$$

Where $X(i, j)$ is the influence degree of node i pressure change on node j pressure; H_i, H_j represents the water pressure of the node i and j under normal working condition; H'_i, H'_j represents the water pressure of the node i and j after the node j pressure changes.

Step 3. According to the calculation of each step of fuzzy clustering analysis, the truncation matrix of fuzzy similar equivalent matrix is obtained, and the results of fuzzy clustering of water consumption of the peak, low peak and general average period are obtained.

Step 4. Repeat Step1-Step3 for several times to carry out multiple fuzzy C-means clustering algorithm for pipe networks nodes to obtain multiple clustering results.

Step 5. The weighted coco-matrix defined in Definition 1 is used to achieve the fusion of multiple clustering results and obtain the final clustering result. According to the cluster results of pipe network nodes, nodes from different classes can be selected as representative points and pressure monitoring points.

III. LEAKAGE LOCATION BASED ON INTEGRATED NEURAL NETWORK

As the water flow (or pressure) in the water supply network is real-time, dynamic and collected in time sequence, which fully conforms to the characteristics of time series data, the water pressure on the monitoring point is also time series data. LSTM subtly controls the combination of short-term memory and long-term memory through "gate structure", and solves the problem of gradient disappearance to a certain extent. Compared with recurrent neural network, LSTM has better results from the analysis of time series data.

The effect of LSTM on stationary data is better than that on non-stationary data. Therefore, the stationary of time series data onto pipe network monitoring points obtained in this paper are judged first. Use Augmented Dickey Fuller (ADF) root test [11] to tests the stationary for pipe network pressure data. Through the test of ADF root test, if the time series data onto the pipe network monitoring point is unstable, the Empirical Mode Decomposition [12] (EMD) is adopted to decompositions and stabilizes the pressure time series data onto the pipe network monitoring point, decompositions the data into multiple intrinsic mode functions (IMFs) and residual terms, and the decomposed IMFs and residual terms are carried out in LSTM for characteristic analysis [13].

In addition to the time series data such as the water pressure on the monitoring point, the non-time series data

such as pipe section length, pipe section diameter, water demand and pipe friction is also considered in this paper. These data are also important influencing factors of pipe network leakage point monitoring. CNN is a kind of feed forward neural network algorithm with convolution computation and depth structure [14,15].

Taking advantage of the strong feature expression ability of CNN, this paper puts non-time series data onto CNN for feature extraction. After splicing the extracted features of those extracted by LSTM, the leakage point location is finally realized. The specific flow of the integrated neural network model adopted in this paper is shown in Figure 1.

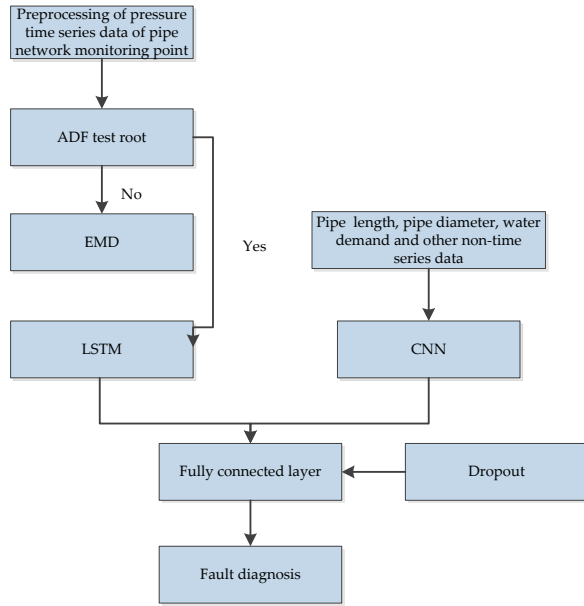


Fig. 1. Flow chart of integrated neural network model.

In this paper, LSTM [16,17] is used to effectively learn the temporal characteristics of data changes. First, the ADF test root is used to test the stationarity of the time series data. If the time series data is unstable, the original time series data of the monitoring point is stabilized by EDM; then the data is sent to the LSTM network for characteristic analysis, and the prediction results of LSTM are output through the full connection layer.

CNN is used to build a model adapted to non-time series data such as pipe length, pipe diameter and water demand. Considering that the actual data amount is much smaller than that in other application scenarios, too many network layers are not considered in the construction of CNN model. In order to prevent the loss of important features after the pooling layer, the pooling layer is connected after continuous convolution. After the pooling operation, the features are first drawn into one dimension, input to the full connection layer, and finally output results, as shown in Figure 2. Before the model training, in order to avoid the phenomenon of overfitting of the model training results due to the dispersion of input data, the normalized preprocessing is carried out for the data. The processed data is divided into training set and test set according to a certain proportion to realize the model training and model error analysis. In this paper, data preprocessing is realized according to formula (4):

$$y'_i = \frac{2(y_i - \min Y)}{(\max Y - \min Y)} - 1. \quad (4)$$

Where, Y is the data set, y_i is the i -th original data in the data set, y'_i is the normalized data, $\max Y$ is the maximum value in the data set, and $\min Y$ is the minimum value in the data set.

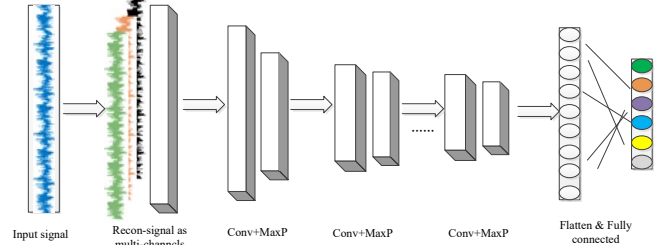


Fig. 2. CNN model architecture

IV. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

A. Experimental Environment

The proposed algorithm and model are applied to a real project case - leakage detection and location of underground water supply network in a region of Langxi County, Xuancheng City, Anhui Province. Based on the layout diagram of underground water supply pipe network in a certain area of Langxi County, the simulation topology diagram of water supply pipe network was constructed by using EPANETH software, as shown in Figure 3. It includes a reservoir, a pool, a pump, 42 pipe network nodes and 65 pipe sections. In the pipe network structure, nodes 19 and 22 represent water consumption in the industrial zone, the remaining nodes represent the water consumption of residential areas. The parameters of each node and pipe segment in the pipe network are shown in Figure 4 and Figure 5.

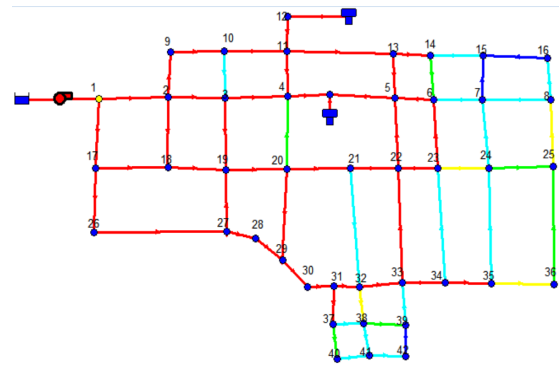


Fig. 3. Topological structure of water supply network

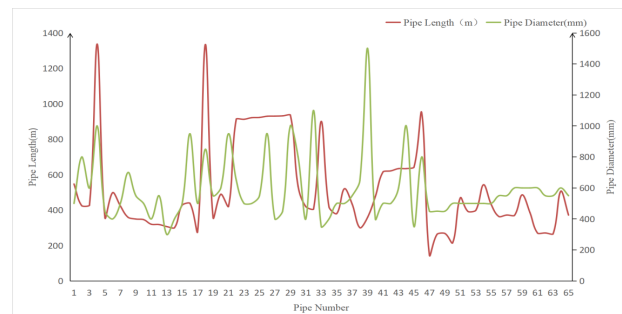


Fig. 4. Parameters of each section of the pipe network

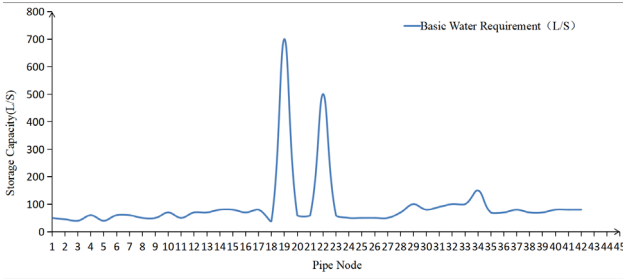


Fig. 5. Parameters of each node of the pipe network

B. Experimental Design

Based on the FCMCF-based pressure monitoring point selection method proposed in Part 2, five monitoring points are selected: 3, 11, 14, 23 and 32. Assuming that the leakage occurs in the center of each pipe segment, the pressure changes of five monitoring points 24 hours before and after the leakage of each pipe segment are analyzed, and the specific pipe segment where the leakage occurs and the specific leak location are determined according to the pressure changes of the five monitoring points. The specific steps are as follows:

Step 1. The EPANET software was used to simulate the topological structure of the water supply pipe network to obtain the node water pressure of all nodes of the network, including 5 monitoring points, 24 hours before the failure.

Step 2. Carry out fault simulation for the established water supply pipe network, and simulate a leakage at the center of each pipe section. Assume that the leakage amounts are 5L/s, 8L/s, 10L/s, 15L/s, 20L/s and 25L/s respectively.

Step 3. Build LSTM and CNN models respectively to learn and train the output of pressure changes at five monitoring points and the distance between fault locations under each group of fault states. By analyzing the corresponding relationship between real-time water pressure changes at monitoring points and leakage faults, the leakage faults of the water supply network are located. Among them, the water pressure time series data of five monitoring points 24 hours before and after the fault occurs are used as the input of the LSTM network, and the data in five cases (5L/s, 8L/s, 10L/s, 20L/s, 25L/s) are applied to the model training, and the trained model is used to locate the leakage fault of the water supply network. The failure data with a leakage amount of 15L/s was taken as the test set to verify the model error. The results were compared with the original data and the errors were analyzed to judge the accuracy of the trained model for fault location.

The CNN model adopts a structure of 5 layers: 1 input layer, 3 convolution layers and 1 output layer. The relevant parameter settings are shown in Table I below. Among them, the input layer is set as 4 neurons, which represent the non-sequential data of pipe length, pipe diameter, water demand and pipe friction. The number of neurons in each layer of the three convolution layers was 200, 80 and 30, respectively. The activation function was corrected Linear Unit (ReLU). The Pooling method was Max Pooling. The optimization algorithm adopts Stochastic Gradient Descent (SGD) algorithm. The above 4-dimensional data were input into CNN for feature extraction, and the extracted features were splicing with those extracted by LSTM network to jointly realize leakage location. The accuracy and loss of the training results of the integrated neural network model are shown in

Figure 6 and Figure 7. Two states are defined here: when no leakage occurs, the state is defined as normal; when a leak occurs, it is defined as a leak state.

TABLE I. PARAMETER SETTING OF CNN MODEL

Model parameters	Set the value
the input layer	four neurons
output layer	four neurons
the activation function	ReLU
pooling method	max Pooling
optimization algorithm	SGD
batch_size	20
momentum	0.9
learning rate	e-4
loss function	cross-entropy
dropout	0.5

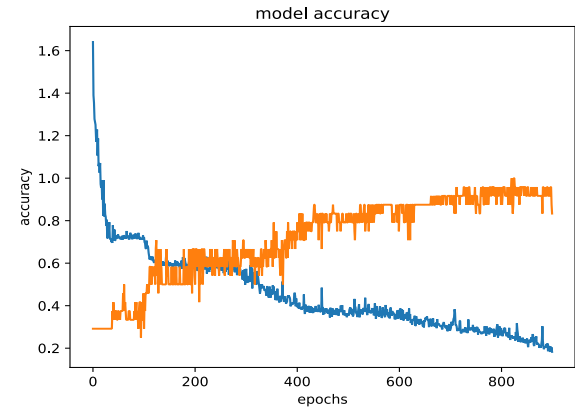


Fig.6. Model training accuracy

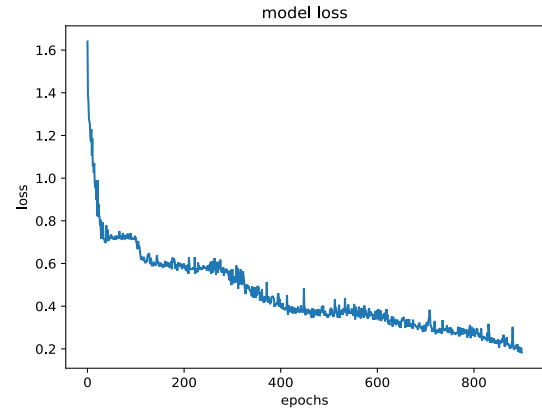
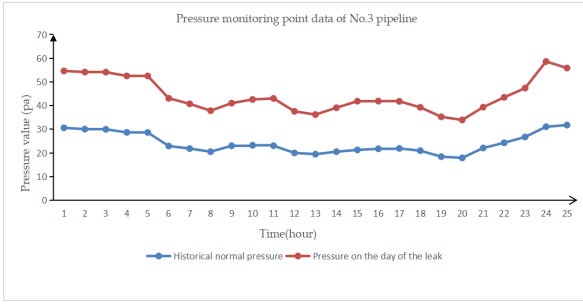
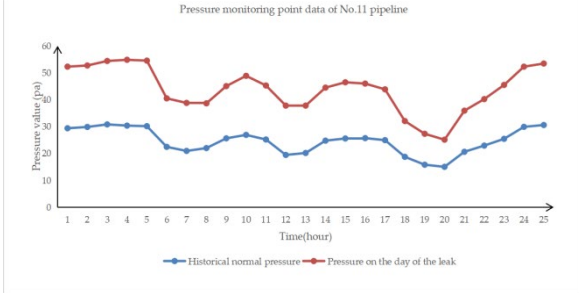


Fig. 7. Loss rate of model training

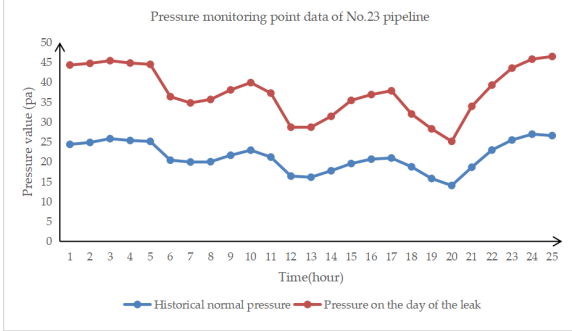
Now the trained model is applied to the occurrence of actual leakage. By observing the change of the pressure value of the five monitoring points for 24 hours under normal conditions and when leakage occurs, as shown in Figure 8, it can be clearly found that the water supply pipe network has leakage.



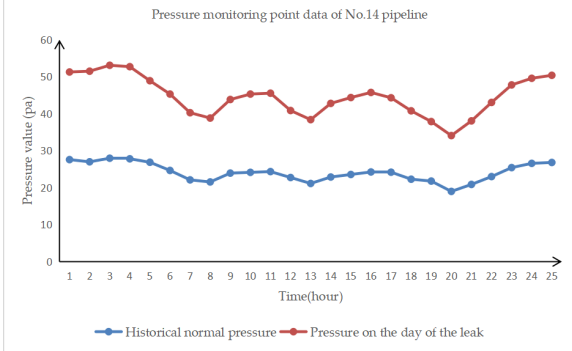
(a) Normal state and pressure change at the first test point after leakage



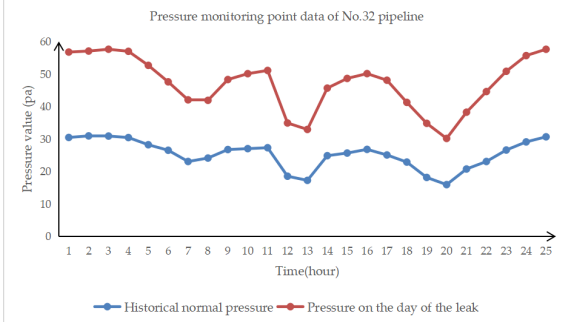
(b) Normal state and pressure change at the second test point after leakage



(c) Normal state and pressure change at the third test point after leakage



(d) Normal state and pressure change at the 4th test point after leakage



(e) Normal state and pressure change at the 5th test point after leakage

Fig. 8. Pressure changes at 5 test points in normal state and after leakage

C. Experimental Comparison and Result Analysis

Will be proposed in this paper the water supply pipe network leakage locating method (FCMCF+EMD+LSTM

+CNN), random monitoring of integrated neural network positioning method (EMD+LSTM+CNN), the integration of neural network based on Kmeans clustering locating method (Kmeans+EMD+LSTM+CNN) and two methods mentioned in ref. [15] (LSTM+ANN, EMD+LSTM+ANN) run on the above data sets, respectively. The accuracy and loss rate of the models in the training set and test set are shown in Table II below. Among them, ANN: neural network algorithm, train-loss: loss rate of training set, Val-loss: loss rate of test set, train-ACC: accuracy of training set, and train-ACC: accuracy of test set.

TABLE II. COMPARISON OF RESULTS OF DIFFERENT LEAKAGE LOCATION METHODS

Methods	Loss rate		Accuracy	
	Train-loss	Val-loss	Train-acc	Val-acc
FCMCF+EMD+LSTM+CNN	0.2024	0.2935	0.9025	0.8555
EMD+LSTM+CNN	0.3635	0.4059	0.8050	0.7557
Kmean+EMD+LSTM+CNN	0.3023	0.3367	0.8860	0.8252
LSTM+ANN	0.4962	0.3423	0.7803	0.8428
EMD+LSTM+ANN	0.5362	0.2438	0.8027	0.8562
FCMCF+CNN	0.5643	0.3339	0.8008	0.8053

As can be seen from Table II, when the monitoring points are randomly selected, the loss rates of the model on the training set and the test set are 0.3625 and 0.4059, respectively, which are higher than that of the proposed method (FCMCF+EMD+LSTM+CNN) on the training set and the test set. At the same time, when the monitoring points are randomly selected, the accuracy of the model on the training set and the test set is 0.8050 and 0.7557 respectively, both of which are lower than the accuracy of the proposed method on the training set and the test set, indicating that the selection of monitoring points does have a certain influence on the location of leakage points in the water supply network.

In order to verify the advantages of the integrated neural network model proposed in this paper, a comparative test was conducted with the two methods (LSTM+ANN, EMD+LSTM+ANN) mentioned in ref. [14]. As can be seen from Table II, FCMCF+EMD+LSTM+CNN has better loss rate and accuracy in the training set than LSTM+ANN and EMD+LSTM+ANN. However, in the test set, the test accuracy of FCMCF+EMD+LSTM+CNN is 0.8555, while that of EMD+LSTM+ANN is 0.8562. There is little difference in the test accuracy of the two methods. For one thing, only 66 groups of fault data are selected as the test set in this paper, which is too little data. For another, it is also related to the unsteady unknown dominance of pipe network flow and its data volatility. On the whole, the experimental results of FCMCF+EMD+LSTM+CNN perform relatively well.

V. CONCLUSION

In this paper, based on the selected monitoring points and the influential factors affecting the leakage points of the water supply network, the data were divided into time series data and non-time series data. The LSTM and CNN models were used to mine the data characteristics, and the neural network was integrated to locate the leakage points. The model accuracy reached 90.25% and the loss rate reached 20.24%. Experimental results verify the feasibility and accuracy of the proposed method.

The number of sections and nodes of urban underground water supply network is huge, so the amount of data

generated in practical application is very large, and a large amount of data is also a test for the model. If gaussian process, support vector machine, or ordinary BP neural network are used to detect leakage points, the accuracy of the model will decrease and the time complexity will increase. CNN and LSTM, as one of the deep learning technologies, have unique advantages for processing large-scale data. The larger the data amount is, the higher the accuracy of the model will be (on the premise of solving the overfitting problem), and the time complexity will also be reduced. This is also the advantage of the method proposed in this paper.

This paper has only been tested in the case of single-point leakage, but more complex cases (two-point or multi-point leakage) need to be further studied and tested. In this paper, only pipe section length, pipe section diameter, water demand and pipe friction are considered in the non-time series data in the experiment. However, in the actual large-scale water supply network, many factors need to be considered, at the same time, the data acquisition cost is high. Therefore, to build a more accurate hydraulic model, it is necessary to find the water use law through regular observation and comprehensively consider various influencing factors, which is more conducive to the accurate positioning of leakage points.

ACKNOWLEDGMENT

This research was funded by the Natural Science Academy of Education Department of Anhui Province KJ2019A0836, the Science and Technology Major Special Project of Anhui Province 17030901026, and the National Natural Science Foundation of China 61806068.

Author Contributions: Conceptualization, X.W.; methodology, X.W., and C.Z.; software, X.W., and C.Z.; validation, C.Z.; formal analysis, X.W., and C.Z.; project administration, X.W.; funding acquisition, X.W., and C.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

- [1] A. Rojik, Endroyono, A. N. Irfansyah, Water Pipe Leak Detection using the k-Nearest Neighbor Method. 2019 International Seminar on

Intelligent Technology and Its Applications (ISITIA), 2019, pp. 393-398.

- [2] Z. B. Wang, Brief introduction of leak detection in water supply network. *Sichuan Cement*, no. 02, 2016, pp. 33.
- [3] J. J. Du, S. Y. Hu, J. L. Gao, et al., Leakage location through flow data of water supply network. *China Water & Waste Water*, vol. 35, no. 19, 2019, pp. 46-49.
- [4] Q. Z. Zhang, Y. W. Zheng, M. Zhao, et al., Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *Journal of Water Resources Planning and Management*, vol. 142, no. 11, 2016, pp. 04016042.
- [5] J. Kang, Y. J. Park, J. Lee, et al., Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems. *IEEE Transactions on Industrial Electronics*, Oct. 2017, pp. (99)1-1.
- [6] X. Zhou, Z. H. Tang, W. R. Xu, et al., Deep learning identifies accurate burst locations in water distribution networks. *Water Research*, vol. 116, 2019, pp. 1-12.
- [7] C. C. D. Santos, A. J. P. Filho, Water demand forecasting model for the metropolitan area of san paulo, brazil. *Water Resources Manage*, vol. 28, no. 13, 2014, pp. 4401-4414.
- [8] M. C. Liu, Research on leakage location model of rural water supply pipeline network. Handan: Hebei University of Engineering, 2019, pp. 47-54.
- [9] L. H. Wang, H. W. Li, M. Lv, et al., Intelligent locating method of water pipeline leakage based on deep learning of BP neural network. *Water Resources and Power*, vol. 37, no. 5, 2019, pp. 61-64.
- [10] X. X. Wu, Z. W. Ni, L. P. Ni, et al., Research on selective clustering ensemble algorithm based on normalized mutual information and fractal dimension. *Pattern Recognition and Artificial Intelligence*, vol. 27, no. 9, 2014, pp. 847-855.
- [11] Y. Q. He, Research on corporate governance of southeast asian corporations. Lanzhou: Lanzhou University of Finance and Economics, 2015, pp. 25-39.
- [12] X. C. Yue, Research on Empirical Mode Decomposition Algorithm's Application. Xi'an: Xidian University, 2013, pp. 9-20.
- [13] X. L. Lv, Research on leakage location of water supply pipe network based on deep learning. Qingdao: Qingdao University of Technology, 2019, pp. 41-47.
- [14] J. Bao, Y. C. Dong, H. Z. Liu, A review of the development of convolutional neural networks// *Proceedings of the 24th Annual Conference on New Network Technologies and Applications*, Beijing: China Academic Journal Electronic publishing House, 2020, pp. 10-15.
- [15] A. Krizhevsky, I. Sutskever, G. E. Hinton, Image net classification with deep convolutional neural networks. *Communications of the ACM*, vol. 60, no. 6, 2017, pp. 84-90.
- [16] J. Liu, Q. Li, W. Chen, et al., Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks. *International Journal of Hydrogen Energy*, 2018, pp. 1-11.
- [17] K. Hu, C. Zheng, W. He, et al., Multi-channels LSTM networks for fence activity classification. *IEICE Transactions Information and Systems*, vol. 101, no. 8, 2018, pp. 2173-2177.