Received May 25, 2020, accepted May 30, 2020, date of publication June 8, 2020, date of current version June 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000960

A Novel PPA Method for Fluid Pipeline Leak Detection Based on OPELM and Bidirectional LSTM

LEI YANG AND QING ZHAO, (Member, IEEE)
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2N8, Canada
Corresponding author: Lei Yang (lyang2@ualberta.ca)

This work was supported by the Natural Science and Engineering Research Council (NSERC) of Canada under Grant 538630-19.

ABSTRACT Pipeline leak detection has attracted great research interests for years in the energy industry. Continuous pressure monitoring is one of the most straightforward approaches in leak detection which utilizes pressure point analysis (PPA) algorithms to exploit the transient pressure characteristics and identify leak events. However, a critical issue that jeopardizes the deployment of PPA based methods is the high false alarm rate. In this paper, a novel PPA based leak detection method is proposed which can accurately detect the leak events and dramatically decrease the number of false alarms compared to existing methods. Firstly, the proposed method takes advantage of the good approximation ability and fast learning speed of optimally-pruned extreme learning machine (OPELM) to produce a preliminary leak detection result. Then, the strong memorizing ability of bidirectional long-short term memory (BiLSTM) network is exploited to identify the true positive from the preliminary detection result, hence significantly decrease the number of false alarms. Furthermore, a feature extraction mechanism is also proposed to obtain both the dynamic and static characteristics from raw pressure wave. Experiments and verifications are performed on different real world data sets obtained from pipeline leak tests. It shows that the proposed method can achieve higher detection accuracy with significantly less false alarms. It enhances the practicality of pressure monitoring based leak detection schemes.

INDEX TERMS BiLSTM, feature extraction, leak detection, OPELM, pressure point analysis.

I. INTRODUCTION

Pipeline leak detection has been extensively studied for years due to its crucial importance in oil and gas industry. According to statistics [1], most of the products are transported through pipeline networks which benefit from the lower overhead cost and greater transportation capability, compared to other transportation methods. However, the rupture or leak issue may incur huge property loss and environmental hazard especially when the leak spot lacks monitoring. Hence, various techniques for pipeline leak detection and localization have been proposed. Pressure monitoring based leak detection is a commonly used internal method that continuously performs pressure point analysis (PPA) to detect leak events. The PPA based approach owns several advantages [2], [3], such as fast response, high sensitivity, continuous monitoring, and easy installation/maintenance.

Traditionally, model based PPA methods are dominant. In [4], a modified model analysis method is proposed to analyze the transient process for leak detection and localization. In [5], the detection method combining pipeline dynamic model and extended Kalman filter is proposed by which the detection accuracy is greatly improved. However, these model based methods require prior model knowledge, and have limited accuracy and flexibility in noisy and complex industrial situations.

In recent years, machine learning (ML) and data-driven techniques have been widely accepted and employed in industries with resources such as abundant sensors and big data. Successful applications of ML have been reported in plant wide system/process monitoring and fault diagnosis, [6], [7]. Although pipeline leak monitoring is distributed in nature, traditional leak detection techniques such as PPA

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

usually involves local and point-wise inspections. Recent research has shown promising applications of data-driven ML techniques in pipeline systems, which can not only achieve rapid and reliable local leak detection, but can also be extended to designing the distributed pipeline health monitoring system. In [8], [9], the application of back-propagation neural networks (BPNN) in leak detection is experimented and discussed. As one of the commonly applied ML methods, support vector machine (SVM) is also implemented in leak detection as shown in [10], [11]. In [12], [13], k-nearest neighbor (KNN) algorithm is adopted for pipeline leak detection and rupture size estimation. Its performance is validated through comparison study with some other ML methods. Extreme learning machine (ELM), which owns the merits of good approximation ability and fast learning speed, has been applied in leak detection in [14], [15]. It is shown that the time spent on model learning is greatly reduced. Applications of other ML methods in leak detection are also reported, including naive Bayesian (NB) based [16] and decision tree (DT) based classifier [17]. Comparing to the analytical model based method, the ML based can improve the detection accuracy and generalization performance when adapting to different industrial situations.

However, one challenge for deploying PPA based methods is the high rate of false alarms [18]. This is due to the fact that frequent pump or valve manipulations may also lead to pressure drops, which can be mistakenly detected as leaks. Due to such a drawback, PPA is usually taken as a supplement to other leak detection methods [19], adding complexity to a practical leak detection strategy. Thus, false alarm elimination becomes crucial for deploying PPA based methods in practice. In [20] and [21], the flow balance method is employed to assist the discrimination of false alarms through installation of flow meters at investigated points. In [22], a multi-sensor paring method is proposed to decide the genuineness of a leak by considering the feasibility of paired pressure drop time instants. The methods in [20]–[22] are conditioned on the prior knowledge such as multiple flow readings and installation topology of pressure sensors. Furthermore, a pattern matching method which compares the similarity of pressure drops between the real leak and normal adjustments is also proposed. In [23], a two-stage decision scheme is presented where the short-term and long-term models are trained respectively and a switching threshold is set to decide the proper model. This method intends to utilize the better fitting model to perform detection with respect to different lengths of pressure sequences, such that the number of false alarms can be reduced. Experiments show that the detection accuracy is improved while false alarms decrease if the model can be correctly selected. However, the appropriate value of the switching point for choosing suitable model is difficult to obtain which hinders its practical implementation.

In view of these drawbacks, in this paper a machine learning based PPA method is proposed which can accurately perform leak detection with significantly reduced false alarms. The proposed method is based on supervised optimally-pruned extreme learning machine (OPELM) combined with bidirectional long-short term memory (BiLSTM) network, which is shown to improve the performance and enhance the practicality of the pressure monitoring based leak detection. Main contributions of this paper are described as follows:

- An effective PPA leak detection method based on supervised OPELM combining BiLSTM is proposed to achieve higher detection accuracy and significantly less false alarms, compared to existing ML based PPA methods.
- The strong past and future memorizing ability of BiLSTM is firstly utilized to identify the true and false leaks by considering the ambient pressure status around the suspicious leaks.
- Several unique characterizations of leak features are proposed and the effectiveness is verified through experiments.
- Performance of the proposed method is assessed and compared with various ML based methods through multiple experiments on different industrial data sets.

The remainder of this paper is organized as follows. Section II briefly reviews the fundamental principles of OPELM and BiLSTM. Section III presents the main leak detection methodology. Section IV includes the experiment and comparison results with discussions. Finally, conclusion is drawn in Section V.

II. PRELIMINARIES

A. OPTIMAL-PRUNED EXTREME LEARNING MACHINE

ELM is a learning algorithm proposed by Huang *et al.* in [24]. It employs the structure of single-hidden layer feedforward neural networks (SLFNs) where hidden layer weights and biases can be randomly assigned while output weights are analytically obtained. Since it does not need time consuming back-propagation training process, ELM has extremely fast learning speed and satisfactory approximation capability.

OPELM is a variant of standard ELM with additional steps to make it more robust and generic [25]. ELM algorithm may not be very accurate when the input data set contains uncorrelated elements to the output, or the observations of different features are to some extent correlated. To overcome this, OPELM is proposed where a procedure is adopted to eliminate the uncorrelated variables via pruning of their associated hidden layer neurons. Main steps of OPELM [26] are shown in Fig. 1. Firstly, the ELM structure is established with relatively larger number of hidden neurons which may be pruned afterwards. Secondly, in the step of multiresponse sparse regression (MRSR), the forward feature selection scheme is applied to generate a list of most useful neurons

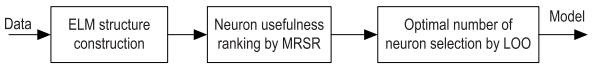


FIGURE 1. Block diagram of OPELM.

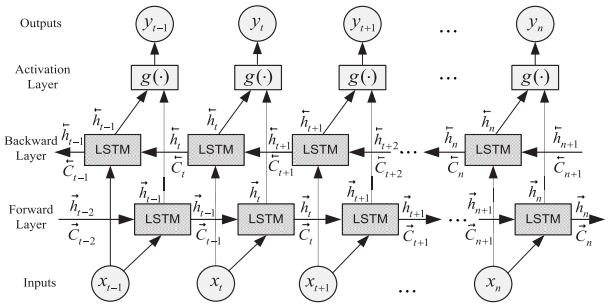


FIGURE 2. Bidirectional LSTM structure.

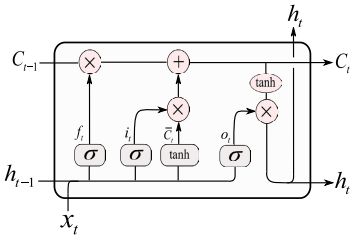


FIGURE 3. Memory cell of LSTM.

under different number of hidden neuron conditions. Finally, leave-one-out (LOO) criterion is employed to select the most appropriate number of neurons. Compared to ELM, OPELM requires longer training process due to the implementation of optimal-pruning step, but the model accuracy can be significantly improved.

B. BIDIRECTIONAL LONG-SHORT TERM MEMORY NETWORKS

LSTM is a type of recurrent neural network (RNN), which has been widely studied and implemented in image processing, sentiment analysis, language translation [27] and handwriting recognition, etc.. It is a promising technique in sequential data prediction and pattern recognition due to its ability of memorizing the previous states information.

Unlike the traditional LSTM which only has the forward layer, BiLSTM employs both forward and backward layers. The structure of BiLSTM is shown in Fig. 2. In BiLSTM [28], the current output y_t is dependent on both of the past and the future status. For example, the current output y_t is a function of both of the forward layer output $\vec{h}_t$ and the backward layer output $\overleftarrow{h}_t$. The forward layer output $\vec{h}_t$ is dependent on the current input x_t, the past forward layer output $\vec{h}_{t-1}$ and its cell state $\vec{C}_{t-1}$, hence the past status is taken into account. Similarly, the backward layer output $\overleftarrow{h}_t$ is dependent on the current input x_t, the future backward layer output $\overleftarrow{h}_{t+1}$ and its cell state $\overleftarrow{C}_{t+1}$, where the future status is also considered. Thus, by employing this structure, the ambient information can be exploited and utilized to decide the current output.

The core block in BiLSTM is the memory cell as shown in Fig. 3, the brief introduction of one memory cell is given as follows.

A memory cell contains three control gates, namely the input, forget and output gate. They are described as follows:

1) Input gate: It controls whether the memory cell is updated.

$$i_t = \sigma(W^i(h_{t-1}, x_t) + b^i) \tag{1}$$

2) Forget gate: It controls if the memory cell is reset.

$$f_t = \sigma(W^f(h_{t-1}, x_t) + b^f) \tag{2}$$

3) Output gate: It controls if the current cell state C_t is made visible.

$$o_t = \sigma(W^o(h_{t-1}, x_t) + b^o) \tag{3}$$

Besides, the cell state modification $\bar{C}$ is described as:

$$\bar{C}_t = tanh(W^c(h_{t-1}, x_t) + b^c) \tag{4}$$

It can be seen that the current gating effects i_t, f_t, o_t and the cell state modification $\bar{C}_t$ are all functions of the previous hidden state h_{t-1} and the current input x_t. Then the current cell state C_t and hidden state vector h_t are expressed as:

$$C_t = f_t C_{t-1} + i_t \bar{C}_t \tag{5}$$

$$h_t = tanh(C_t) \times o_t \tag{6}$$

The parameters in BiLSTM that can be obtained through training process are $[W^i, b^i]$, $[W^f, b^f]$, $[W^o, b^o]$ and $[W^c, b^c]$. They denote the weights and biases for input gate, forget gate, output gate and the cell state modification respectively. The training process is usually performed by the back-propagation through time (BPTT) algorithm [29].

III. MAIN METHODOLOGY

A. MOTIVATION OF THE PROPOSED METHOD

Although the high sensitivity to pressure changes owned by PPA can lead to fast leak detection response, it may also contribute to the high number of false alarms. Fig. 4 is given as an example to show how a false alarm can occur by closely comparing the real leak wave and the normal pressure fluctuating wave (that may be taken as a leak by mistake). The top plot in Fig. 4 displays the process of a real leak experiment where the red colored parts are corresponding to leak events. The bottom plot in Fig. 4 displays a process of non-leak normal working pressure fluctuation.

The current PPA based methods usually take the typical leaking pressure transient as a signature (e.g., the 2*nd* red colored portion from left in the top of Fig. 4) to detect leaks. When a section of pressure wave is deemed similar to the signature, it may be considered as a suspicious leak. For example, the wave sections between the narrow red dashed lines in the bottom of Fig. 4 may be considered as leaks because they bear certain similarity with the leak signature in the top plot. Thus, when applying the conventional PPA based method, many false alarms may be generated due to the existence of numerous similar transient pressure drops in the normal working process.

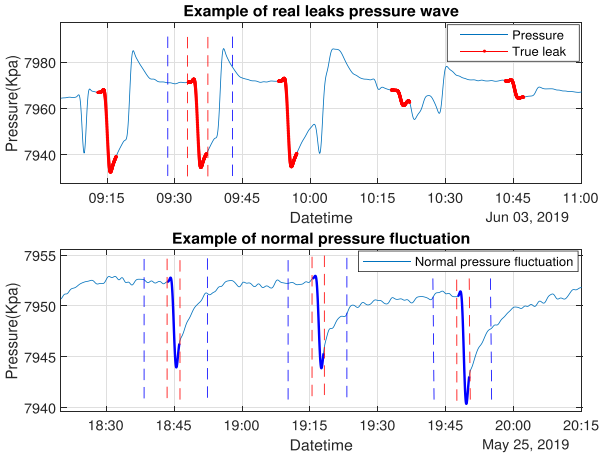


FIGURE 4. Example of true leaks and normal pressure fluctuation.

However, given the typical leak pressure wave shown in the top of Fig. 4, it is easy for the human to tell that the bottom plot does not show an actual leak process. This is because in human perception, both of the pressure drop contained in the narrow window (the gap between red dashed lines), and its ambient pressure in the wider temporal window (such as the gap between blue dashed lines) are under consideration. In other words, Fig. 4 shows similar sudden pressure drops in both leak and non-leak processes, however, if we observe a wider temporal range, their ambient pressure characteristics are obviously different.

If we merely choose a wider leak signature (i.e., expand the red-colored leak portion to a wider range), then the most prominent leak feature, usually manifested as a sudden pressure drop, may become indistinct due to the mixing of the ambient pressure status. Hence, a method that can imitate the above human perception is desirable, which can quickly capture a possible leak, and then pinpoint a true leak by excluding false alarms using ambient pressure information. The former can be achieved by proper feature extraction and classification, while the latter can be realized by BiLSTM, which has strong memorizing ability to treat the ambient pressure information.

Consequently, we propose a two-stage PPA leak detection scheme based on combined OPELM and BiLSTM networks. It exploits the fast learning and superior classification performance of OPELM to perform first-stage detection and then takes advantage of the strong memorizing ability of BiLSTM to broaden the temporal observation range, thus effectively eliminates false alarms.

B. STRUCTURE OF THE PROPOSED LEAK DETECTOR

For the pipeline leak detection studied in this paper, the training and detection phases of the proposed method are shown in Fig. 5 and Fig. 6. Since the method is combining OPELM and BiLSTM, both the training and detection processes can be conducted in two stages as marked in Fig. 5 and Fig. 6. In the first stage of training process, raw labeled pressure data P_r^t is passed to a low-pass filter (LPF) to remove high frequency noises. The superscript t in P_r^t represents the training process and the subscript r indicates raw pressure data. The output of LPF is the filtered pressure data, denoted as P_f^t. From P_f^t, the leak portion matrix $\boldsymbol{P}_l^t$ and non-leak portion matrix $\boldsymbol{P}_{nl}^t$ can be retrieved with the same length m according to the known label information. Apply feature extraction given in Section III-D on $\boldsymbol{P}_l^t$ and $\boldsymbol{P}_{nl}^t$ to obtain feature matrix for training, denoted as $\boldsymbol{F}_{tr}$, and feature matrix for testing, denoted as $\boldsymbol{F}_{te}$, for which the corresponding labels are written as Y_{tr} and Y_{te}, respectively. Thus, the OPELM network can be obtained through training.

In the second stage of training process, the testing result $\hat{Y}_{te}^t$ from the first stage, which contains suspicious leaks, is separated into true positive (TP) and false positive (FP) groups. The time instant vectors of the suspicious leaks T_{TP} and T_{FP} are taken as centers of the training sequences for BiLSTM. The training sequences of TP and FP, denoted as $\boldsymbol{S}_{TP}^t$ and $\boldsymbol{S}_{FP}^t$, respectively, are taken from $\boldsymbol{F}^t$, the feature matrix of the entire training pressure data. Thus, the training feature set for BiLSTM is established. After performing BiLSTM training, both the OPELM and BiLSTM networks are obtained.

Once OPELM and BiLSTM networks are constructed, leak detection can be performed as shown in Fig. 6. Denote the raw pressure data as P_r^d, where the superscript d represents the detection process. In the first stage of detection process, P_r^d is passed through LPF, feature extraction and OPELM detection. The output $\hat{Y}_{elm}^d$ is then fed to the second stage to further discriminate the TPs and FPs. The input sequences of BiLSTM, $\boldsymbol{S}^d$, are centered according to the suspicious leak time instants in $\hat{Y}_{elm}^d$ and selected from the extracted features in $\boldsymbol{F}^d$. Then, the output of BiLSTM are the final detection result $\hat{Y}_{lstm}^d$.

Details of the proposed method are further elaborated in the following subsections.

C. DATA PRE-PROCESSING

There are two main steps in the data pre-process, i.e., data filtering, data sectioning and labeling.

1) Data filtering: the raw training pressure data P_r^t is from the records of leak experiments and therefore well labeled. It contains high frequency noises which may jeopardize the accuracy of leak detection, thus, LPF is applied to eliminate the noise. In this paper, the low pass FIR filter of order 20 is employed with the sampling frequency $1Hz$ and the cutoff frequency of $0.1Hz$. Although the filter structure and parameters may vary from case to case depending on the noise condition, the filtered data should retain the characteristics of transient leak pressure drops. It can be observed in Fig. 7 that after filtering, the pressure wave is smoother with excessive noises eliminated.
2) Data sectioning and labeling: After raw data is filtered, to establish the training sets for OPELM, the typical leak portion matrix $\boldsymbol{P}_l^t$ and non-leak portion matrix

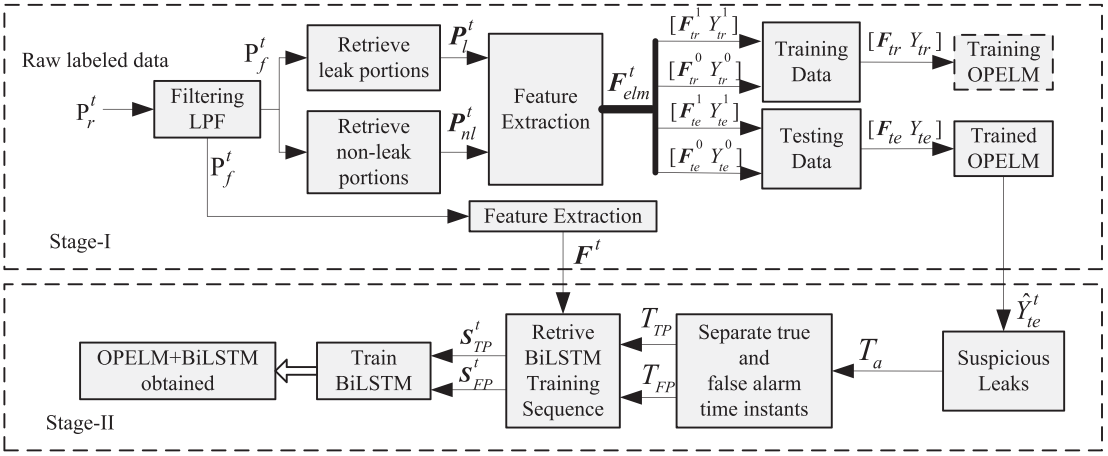


FIGURE 5. Block diagram of training process.

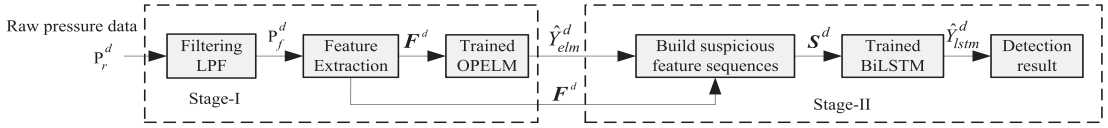


FIGURE 6. Block diagram of detection process.

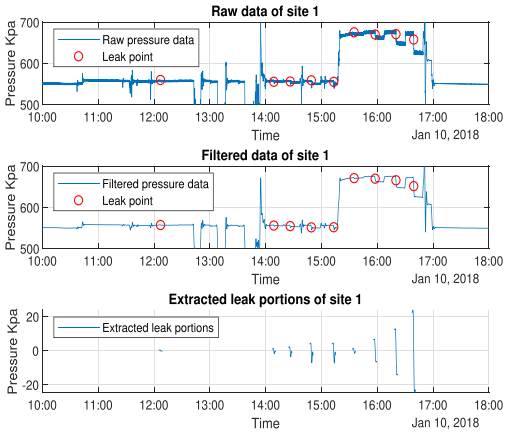


FIGURE 7. Filtering and leak portions of site 1.

$\boldsymbol{P}_{nl}^t$ are singled out from P_f^t. The $\boldsymbol{P}_l^t$ can be sectioned from P_f^t according to the provided leak information. For example, as shown in the bottom plot in Fig. 7, there are 9 leak portions retrieved from the data of Site-1. The length of the leak portion is set to be m seconds where m may vary with respect to different leak situations. In the case shown in Fig. 7, m is set to be 240 which appropriately covers the pressure impact duration caused by the rupture. The size of $\boldsymbol{P}_l^t$ is $r \times m$, where r is the number of leak portions. It can be observed that the negative pressure impacts caused by leaks usually have similarities in shape and dropping trend, but the amplitude may vary due to different leaking conditions such as leak size, pipeline pressure status, flow rate and pipe contents, etc.. Each row of the non-leak pressure matrix $\boldsymbol{P}_{nl}^t$ is randomly sectioned from P_f^t by avoiding the leak instants. Assume q non-leak portions are selected, the size of $\boldsymbol{P}_{nl}^t$ is $q \times m$. The length of a non-leak portion is also m seconds.

D. FEATURES EXTRACTION

A feature extraction scheme is proposed to represent leak characteristics, as shown in Fig. 8. For each pressure portion, 5 features are extracted including "Similarity", "Interception", "Slope rate", "Area" and "Variance". Thus, the size of the extracted feature matrix with n pressure portions is $n \times 5$.

Denote the feature matrix for training OPELM as $\boldsymbol{F}_{elm}^t$, which is divided into the training and testing parts by K fold training scheme, written as $\boldsymbol{F}_{elm}^t = [\boldsymbol{F}_{tr}^T, \boldsymbol{F}_{te}^T]^T$. Combining with the label vectors, the labeled training feature matrix can be written as $[\boldsymbol{F}_{tr}\ Y_{tr}]$, where $\boldsymbol{F}_{tr} = [\boldsymbol{F}_{tr}^{1\,T}\ \boldsymbol{F}_{tr}^{0\,T}]^T$ and $Y_{tr} = [Y_{tr}^{1\,T}\ Y_{tr}^{0\,T}]^T$. The superscript 1 and 0 are referring to the class 1 (leak) and class 0 (non-leak), respectively, and Y_{tr}^1, Y_{tr}^0 are the corresponding label vectors. $\boldsymbol{F}_{te}$ and Y_{te}, are defined similarly except that they are for testing the trained OPELM classifier.

In the following, the characterization of the leak signature and several features are calculated. It should be noted that the "Similarity" related calculations in III-D1 and III-D2 are performed on normalized $\boldsymbol{P}_l^t$ and $\boldsymbol{P}_{nl}^t$ with range of $[-1, 1]$.

1) LEAK SIGNATURE CHARACTERIZATION

As can be viewed in Fig. 7, multiple leak portions are retrieved from P_f^t, some of them vary in amplitudes and shapes. Therefore, the one which is the most similar to other leak portions but dissimilar to the non-leak portions should be chosen as the leak signature. The leak signature selection process is introduced as follows.

Assume one of the r normalized leak portions is written as $P_{li}^t \in \boldsymbol{P}_l^t$, $i = 1 \cdots r$, $P_{li}^t = [p_{i1} \cdots p_{im}]$ and $\boldsymbol{P}_l^t$ is a $r \times m$ matrix. Similarly assume one of the q normalized non-leak portion is written as $P_{nlj}^t \in \boldsymbol{P}_{nl}^t, j = 1 \cdots q$, $P_{nlj}^t = [p_{j1} \cdots p_{jm}]$ and $\boldsymbol{P}_{nl}^t$ is sized at $q \times m$. The subscripts "l" and "nl" are corresponding to the "leak" and "non-leak" respectively. Define the total number of selected leak and non-leak portions as $n = r + q$.

Apply Gaussian method to calculate the similarity matrix $\boldsymbol{W}$ of the concatenated matrix $\boldsymbol{P}^t = \left[\boldsymbol{P}^{tT}_l, \boldsymbol{P}^{tT}_{nl}\right]^T$, where $\boldsymbol{P}^t$ is sized at $n \times m$. The element w_{ij} representing the similarity between the i-th row P_i^t and the j-th row P_j^t in $\boldsymbol{P}^t$ is written as,

$$w_{ij} = e^{-\frac{||P_i^t - P_j^t||}{2}} \quad i, j = 1 \cdots n. \tag{7}$$

Thus, the similarity matrix $\boldsymbol{W}$ of $\boldsymbol{P}^t$ is obtained as,

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1r} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ w_{i1} & \cdots & w_{ir} & \cdots & w_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ w_{n1} & \cdots & w_{nr} & \cdots & w_{nn} \end{bmatrix} \tag{8}$$

When the Euclidean distance between P_i^t and P_j^t is small, the value of w_{ij} is close to 1; Oppositely, if the Euclidean distance is big which means the two vectors are obviously different, the value of w_{ij} is approaching 0.

The leak signature is chosen as the i-th leak portion P_{li}^t in $\boldsymbol{P}_l^t$ from the following

$$\underset{i}{\text{argmax}} \left(\sum_{j=1}^{r} w_{ij} - \sum_{j=r+1}^{n} w_{ij} \right), \quad i \in [1 \cdots r]. \tag{9}$$

The first sum in (9) indicates the similarity of i-th leak portion with other leak portions including itself, and the second sum indicates the similarity of the i-th leak with the non-leak portions. It is to choose the leak portion which has the highest similarity sum value with other leak portions and lowest similarity sum value with non-leak portions. Denote the chosen leak signature P_{li}^t as the template P_s^t, where the subscript s means signature.

2) SIMILARITY

Take P_s^t as the template, the feature f_1 representing similarity is calculated by (10). As shown in Fig. 8, the "similarity" is calculated between a normalized pressure portion P_i^t in green and the chosen template P_s^t in blue. Given a pressure portion

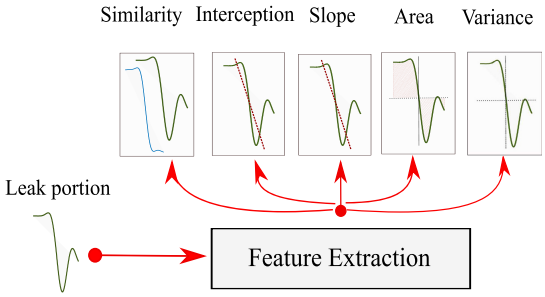


FIGURE 8. Diagram of feature extraction.

$P_i^t \in \boldsymbol{P}^t, i = [1, \cdots, n]$, the similarity feature of the i-th pressure portion is obtained as:

$$f_{i1} = e^{-\frac{||P_i^t - P_s^t||}{2}} \tag{10}$$

Although f_{i1} can reflect the similarity between the normalized pressure vectors and the chosen template, it still has drawbacks. For example, when the amplitude of a pressure vector is obviously different from the template, after normalization, the difference is ignored due to the consistent normalizing range $[-1, 1]$. It may result in high similarity value in f_{i1} as long as the two shapes are alike after normalization. In view of this drawback, other factors are also considered.

3) SLOPE RATE AND INTERCEPTION

The feature "similarity" represents the overall shape similarity after normalization. However, the dynamic characteristics such as "slope rate" and "interception" reflecting the intensity of a pressure dropping also need to be considered. Instead of implementing normalization, the pressure portion is treated with mean centering.

Apply minimum mean squared error (MMSE) based linear fitting on the selected portion P_i^t after mean centering, the two features can be found as:

$$[\alpha, \beta] = \underset{\alpha,\beta}{\text{argmin}} MSE(P_i^t - \hat{P}_i^t) \tag{11}$$

$$\hat{P}_i^t = \alpha t + \beta \tag{12}$$

where $\hat{P}_i^t$ is the linear fitted vector of P_i^t when slope and interception are α and β respectively.

Features of slope rate and interception for P_i^t can be written as:

$$f_{i2} = \alpha \tag{13}$$

$$f_{i3} = \beta \tag{14}$$

4) AREA

The factor representing the feature of amplitude is considered in f_{i4} by calculating the area. The amplitude indicating the distance between the maximal and minimal values is a commonly used feature to evaluate the pressure drop. This feature may be effective when the pressure wave is smooth and no outlier exists. However, the sharp spikes or outliers usually exist, hence the amplitude may not faithfully represent the

pressure drop. It is necessary to use an alternate measure which is less sensitive to spikes and outliers to represent the essential pressure drop. For the selected data vector of length m, after mean centering, the area formed by P_i^t and time axis is calculated by discrete integration method, which is always positive and not sensitive to outliers. So the "Area" is taken as the 4th feature to represent the general pressure drop.

$$f_{i4} = Area, \tag{15}$$

5) VARIANCE

The variance representing the extent of variation of a pressure portion P_i^t is also taken as a feature, which is calculated as:

$$f_{i5} = \sum_{j=1}^{m} p_j^2, \quad p_j \in P_i^t \tag{16}$$

6) CONCATENATION OF FEATURES

After the features for class 1 (Leak) and class 0 (Non-leak) are extracted, the training set for OPELM can be established by concatenation of $\boldsymbol{F}_{tr}^0$ and $\boldsymbol{F}_{tr}^1$, for which the corresponding label values are Y_{tr}^1 and Y_{tr}^0, where

$$\boldsymbol{F}_{tr}^1 = \begin{bmatrix} f_{11}^1 & \cdots & f_{15}^1 \\ \vdots & \cdots & \vdots \\ f_{r1}^1 & \cdots & f_{r5}^1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_1^{1T} \\ \vdots \\ \boldsymbol{f}_r^{1T} \end{bmatrix} Y_{tr}^1 = \begin{bmatrix} +1 \\ \vdots \\ +1 \end{bmatrix} = \begin{bmatrix} y_1^1 \\ \vdots \\ y_r^1 \end{bmatrix} \tag{17}$$

$$\boldsymbol{F}_{tr}^0 = \begin{bmatrix} f_{11}^0 & \cdots & f_{15}^0 \\ \vdots & \cdots & \vdots \\ f_{q1}^0 & \cdots & f_{q5}^0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_1^{0T} \\ \vdots \\ \boldsymbol{f}_q^{0T} \end{bmatrix} Y_{tr}^0 = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} = \begin{bmatrix} y_1^0 \\ \vdots \\ y_q^0 \end{bmatrix} \tag{18}$$

In practice, the data acquired during leak events are very rare and valuable. It is usually difficult to have enough class 1 training samples from real leak events. In this case, a training sample generation scheme is applied.

Assume there are l real leak events, and the r training vectors of class 1 are generated based on the l real leak vectors. The training features are generated following a Gaussian distribution with mean values corresponding to the l real leak feature vectors and the user defined standard deviations. Theoretically, if more real leak events can be recorded to analyze the distribution of features, the feature generation scheme may achieve better approximation by following the analyzed distribution instead of Gaussian. On the other hand, the original data for training class 0 is plenty and in our case the q samples are randomly selected.

7) DISTRIBUTION OF EXTRACTED FEATURES

The feature extraction maps the pressure data to a higher dimensional feature space to enhance the feature representation. As shown in Fig. 9, the distribution of extracted features can be easily classified. The three axes are chosen as the first three principle components (PCs) from the principal component analysis (PCA) method. The separable leak and non-leak

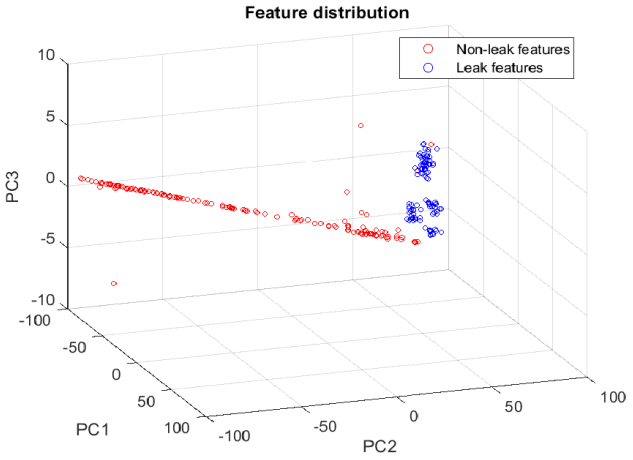


FIGURE 9. Distribution of feature extraction.

distributions and experiment results given in Section IV verify the effectiveness of proposed feature extraction scheme.

E. THE FIRST-STAGE OPELM TRAINING

1) ELM STRUCTURE

OPELM is a variant of ELM which can optimally prune the number of hidden neurons. It shares the same structure with ELM, which is depicted in Fig. 10 with details given as follows.

- Input layer: it contains 5 nodes corresponding to the 5 elements of the i-th extracted feature vector $\boldsymbol{f}_i = [f_{i1}, \cdots, f_{i5}]^T$, where $i \in [1, n]$.
- Hidden layer: it contains L hidden nodes with activation function $g(x) : R^5 \rightarrow R$. The output of the l-th hidden node is given by

$$h_l = g(a_l{}^T \cdot \boldsymbol{f}_i + b_l), \tag{19}$$

  where $l = 1, \cdots, L$ and a_l is the 5×1 column vector of the weights connecting the l-th hidden node and the 5 input nodes. b_l is the bias of the l-th hidden node. Here a_l and b_l are randomly assigned following the uniform distribution over $[-0.2 \quad 0.2]$ and kept fixed afterwards.
- Output layer: it has only 1 node corresponding to the label y_i when input is $\boldsymbol{f}_i$. Thus, the output of an ELM network with L hidden nodes can be written by

$$y_i(\boldsymbol{f}_i) = \sum_{l=1}^{L} h_l \boldsymbol{\beta}_l = \boldsymbol{h}_i \boldsymbol{B}, \quad \boldsymbol{\beta}_l \in R, \tag{20}$$

  $\boldsymbol{h}_i$ is a row vector of the L hidden nodes outputs. The output weight matrix $\boldsymbol{B}$ includes L output weights connecting the hidden layer and output layer. It can be written as $\boldsymbol{B} = \left[\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_L\right]^T$, and can be obtained by training the network, which is explained in the following.

2) ELM TRAINING

To train ELM network to obtain the output weight matrix $\boldsymbol{B}$, the extracted features in Section.III-D and their label

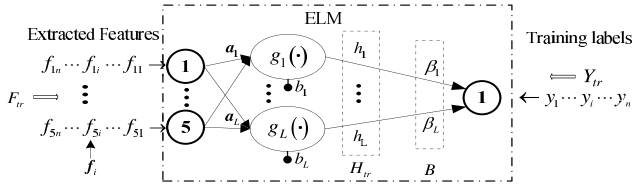


FIGURE 10. ELM training diagram.

vectors are used as training samples, which are denoted as $\boldsymbol{F}_{tr} = [\boldsymbol{F}_{tr}^{1\,T}, \boldsymbol{F}_{tr}^{0\,T}]^T = [\boldsymbol{f}_1^1, \cdots, \boldsymbol{f}_r^1, \boldsymbol{f}_1^0, \cdots, \boldsymbol{f}_q^0]^T$, and $Y_{tr} = [Y_{tr}^{1\,T}, Y_{tr}^{0\,T}]^T = [y_1^1, \cdots, y_r^1, y_1^0, \cdots, y_q^0]^T$, where $n = r + q$.

Let

$$\boldsymbol{H}_{tr} = \begin{bmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_i \\ \vdots \\ \boldsymbol{h}_n \end{bmatrix} = \begin{bmatrix} g(a_1^T \boldsymbol{f}_1 + b_1) & \cdots & g(a_L^T \boldsymbol{f}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(a_1^T \boldsymbol{f}_i + b_1) & \cdots & g(a_L^T \boldsymbol{f}_i + b_L) \\ \vdots & \cdots & \vdots \\ g(a_1^T \boldsymbol{f}_n + b_1) & \cdots & g(a_L^T \boldsymbol{f}_n + b_L) \end{bmatrix} \quad (21)$$

$\boldsymbol{H}_{tr}$ denotes the $n \times L$ hidden layer outputs matrix in the training phase and $\boldsymbol{f}_i$ is the ith column of $\boldsymbol{F}_{tr}$. The n rows of $\boldsymbol{H}_{tr}$ are the hidden layer output vectors corresponding to the n input feature vectors during training.

Establish the cost function considering the empirical and structural risk of the proposed ELM and obtain the following:

$$Min : \mathcal{L}_{ELM} = \frac{1}{2}||\boldsymbol{B}||^2 + \frac{C}{2}||\boldsymbol{Y}_{tr} - \boldsymbol{H}_{tr}\boldsymbol{B}||^2 \quad (22)$$

Solving the regularized least squares optimization problem in (22), we have

$$\boldsymbol{B} = \left(\boldsymbol{H}_{tr}^T \boldsymbol{H}_{tr} + \frac{\boldsymbol{I}_L}{C}\right)^{-1} \boldsymbol{H}_{tr}^T \boldsymbol{Y}_{tr}, \quad (23)$$

where C is the coefficient for balancing the empirical and structural risks, I_L is the L dimensional identity matrix.

In a nutshell, the overall procedures of OPELM are described separately as the optimal-pruning in Algorithm 1 and ELM training in Algorithm 2.

F. THE SECOND-STAGE BiLSTM NETWORK

The objective of BiLSTM is to further identify false alarms (FP) and true alarms (TP). As the raw data is already labeled during the first-stage training, the feature sequences centered with true leak instants are chosen as TP training sequence. On the other hand, the feature sequences centered with the false alarm instants obtained in OPELM are taken as FP training sequences. Thus, the BiLSTM can be trained to flag false alarms. Fig. 11 gives an example of training sequences of TP and FP respectively. BiLSTM can memorize the characteristics over time for the two sets of sequential features, which can then be classified correspondingly. The details can be found as follows.

Algorithm 1 Optimal-Pruning Process of OPELM

1: Set the ELM structure as 5 input nodes, 1 output node and the maximum permitted hidden nodes number N;
2: Train ELM network (Algorithm 2) with 1 hidden node and compare the usefulness among all the N possible selections, then the hidden node corresponding to least residual is selected, denoted as h_1;
3: Add one more hidden node together with h_1 to train ELM network (Algorithm 2) with 2 hidden nodes. The newly added node is selected from the hidden nodes other than h_1. Evaluate the usefulness of all the networks with 2 hidden nodes and choose the best one. The corresponding selected hidden nodes are denoted as $[h_1, h_2]$;
4: Repeat with k hidden nodes, $k = [3, 4, \cdots, N]$, until the maximum hidden nodes number N is reached.
5: Perform LOO validation on the N trained ELM networks, then the one with least LOO error is chosen and the corresponding L hidden nodes are selected.
6: The hidden nodes other than the L chosen ones are pruned and the best ELM structure is achieved.

Algorithm 2 First-Stage ELM Training Process

1: Set the ELM structure as 5 input nodes, 1 output node and L hidden nodes;
2: Randomly assign the input weights a_l and hidden nodes biases b_l, and keep them fixed afterwards;
3: Calculate the hidden layer output matrix $\boldsymbol{H}_{tr}$ based on Eq.(21) using training data input;
4: Calculate the output weight $\boldsymbol{B}$ by Eq.(23).
5: Take the test feature F_{te} as input to the trained ELM in Eq.(20) to find the classified result $\hat{Y}_{te}^t$ for all the test features.
6: Pass the test result $\hat{Y}_{te}^t$ to the next step to build the training data set for BiLSTM.

1) BiLSTM TRAINING SEQUENCES SELECTION AND TRAINING

The time span of a BiLSTM memory depends on the temporal length of the training sequence. Define the length of BiLSTM training sequence as s, then the memory time of a BiLSTM is $(s-1) \times \delta$, where δ is the step size of the selected section of pressure wave. With $s = 20$ and the step size $\delta = 30s$, the pressure variation characteristics within the time span of 570 seconds can be memorized by the BiLSTM to identify if a reported alarm is true or false.

As shown in Fig. 5, the sequences for training BiLSTM are chosen from the extracted feature $\boldsymbol{F}^t$ according to the first-stage testing result $\hat{Y}_{te}^t$, and $\boldsymbol{F}^t$ is obtained from the entire labeled pressure data set. Assume the total length of labeled pressure wave as N, then there will be $K = \lfloor (N - m)/\delta \rfloor$ sectioned portions, where $\lfloor \cdot \rfloor$ takes the nearest lower integer. Hence, the corresponding feature matrix $\boldsymbol{F}^t$ has the dimension of $K \times 5$, the median time instants of sectioned portions are denoted as T, where $T = [t_1, \cdots, t_K]$.

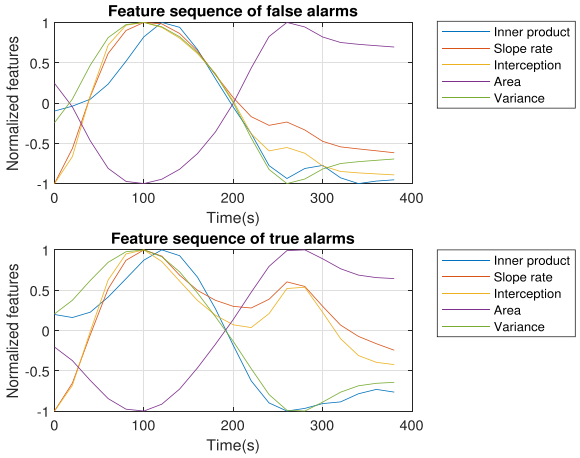


FIGURE 11. Example of BiLSTM training sequence.

The suspicious alarm time instants, denoted as T_a, are selected from T according to the first-stage result $\hat{Y}_{te}^t$. T_a contains the time instants with the corresponding test result in $\hat{Y}_{te}^t$ are in class 1. Since the label information is known, the suspicious alarm time instants T_a can be separated as true positive T_{TP} and false positive T_{FP}. Take the instants in T_{TP} and T_{FP} as centers, choose sequences from $\boldsymbol{F}^t$ with length s, the BiLSTM training sequences for true positive $\boldsymbol{S}_{TP}^t$ and false positive $\boldsymbol{S}_{FP}^t$ are established.

The training process of BiLSTM is usually performed by back-propagation through time (BPTT) algorithm. The overall procedures of the BiLSTM training is summarized in Algorithm 3.

Algorithm 3 Second-Stage BiLSTM Training Process

1: Extract the time instants of the true and false alarms T_{TP} and T_{FP} according to the first-stage test result $\hat{Y}_{te}^t$;
2: Retrieve the feature sequences $\boldsymbol{S}_{TP}^t$ and $\boldsymbol{S}_{FP}^t$ from $\boldsymbol{F}^t$, from which the center time instants of retrieved sequences are T_{TP} and T_{FP}, and the length of retrieved sequences is s;
3: Normalize $\boldsymbol{S}_{TP}^t$ and $\boldsymbol{S}_{FP}^t$ within $[-1, 1]$ and set BiLSTM structural parameters.
4: Train the BiLSTM with normalized $\boldsymbol{S}_{TP}^t$ and $\boldsymbol{S}_{FP}^t$ as a classifier.

G. DETECTION

After the OPELM and BiLSTM networks are trained, the detection process can be applied as shown in Fig. 6. It can be separated into two stages similarly as the training process.

1) FIRST-STAGE OPELM DETECTION

In the first-stage of detection, the raw pressure data P_r^d is fed to LPF to remove high frequency noise and obtain the filtered pressure data P_f^d, where the superscript d represents the detection process. Then, the filtered pressure data is transformed to feature space denoted as $\boldsymbol{F}^d$ via the feature extraction procedure, given in Section.III-D. $\boldsymbol{F}^d$ is then passed to the trained OPELM networks to obtain the first-stage detection result $\hat{Y}_{elm}^d$.

2) SECOND-STAGE BiLSTM DETECTION

Provided with $\hat{Y}_{elm}^d$ from the previous stage, the feature sequences for the second-stage detection can be constructed following the same procedure as in the training phase. Take the sequences according to the suspicious leak instants in $\hat{Y}_{elm}^d$ as input, denoted as $\boldsymbol{S}^d$, the output of BiLSTM networks $\hat{Y}_{lstm}^d$ is the final detection result.

IV. EXPERIMENT

In this section, the proposed two-stage PPA leak detection method is validated on data sets collected from leak experiments on four different industrial sites. In addition, a thorough comparison study is performed to verify the effectiveness of the proposed method. The raw data is measured by the pressure sensors installed along the pipelines. The first pipeline's content is oil/gas while the other three pipelines contain salt water. Besides, the pressure fluctuations on Site-1 and Site-2 are less than that on Site-3 and Site-4.

A. EVALUATION CRITERIA

To evaluate the detection performance, some of the commonly used measures are employed. Denote the counts of positive events (i.e. leaks) as P and negative events (i.e. normal operations) as N. From the detection result, denote the counts of true positive as TP, true negative as TN, false positive as FP and false negative as FN.

- TPR: true positive rate which is the percentage of leaks that are correctly detected.

$$TPR = \frac{TP}{P} \times 100\% \tag{24}$$

- FDR: false discovery rate which indicates the ratio between false positives and all the detected positives. It can be calculated as the the percentage of false leak alarms in all the leak alarms.

$$FDR = \frac{FP}{TP + FP} \times 100\% \tag{25}$$

- ACC: the detection accuracy which indicates the ratio between the number of correct detection counts and all events. It is used to assess the overall detection accuracy including leak events and normal operation events.

$$ACC = \frac{TP + TN}{P + N} \times 100\% \tag{26}$$

B. RESULTS

The proposed OPELM+BiLSTM method employs single hidden layer feedforward neural network structure for OPELM, and 5-layer structure for BiLSTM, which are 'Sequence input layer', 'BiLSTM layer', 'Fully connected layer', 'Softmax layer' and 'Classification output layer'. The corresponding parameters and their values are listed in Table.1.

The overall leak detection results on 4 industrial sites are summarized in Table. 2. The results displayed are averaged

TABLE 1. Related parameters of proposed method.

Model	*Parameter*	*Value*	*Parameter*	*Value*
OPELM	Input Nodes	5	Kernel	Gaussian
	Hidden Nodes	31	Output Nodes	1
BiLSTM	Input Size	5	Output Size	2
	Hidden Units	200	Gate Activation	$Sigmoid$
	State Activation	$tanh$	Softmax Layer	$Softmax$

TABLE 2. Experiment results on different industrial sites.

Site	*Stage*	*TP*	*FP*	*P*	*N*	*TPR*	*FDR*	*ACC*	m
1	1	8.7	426.4	9	71698	96.6%	98%	99.42%	$120s$
	2	8.5	5	9	71698	94.44%	37.04%	99.99%	
2	1	9.7	92.4	10	116150	97%	90.49%	99.92%	$160s$
	2	9	2.9	10	116150	90%	24.37%	99.99%	
3	1	15.2	409	18	120272	84.44%	96.42%	99.66%	$300s$
	2	15.1	5.9	18	120272	83.89%	28.09%	99.99%	
4	1	17.9	261.1	19	120655	94.21%	93.58%	99.78%	$260s$
	2	17.6	11.4	19	120655	92.63%	39.31%	99.98%	

TABLE 3. Experiment environment of Site-1.

Index	*Pipe Flow*	*Leak Flow*	*Leak Percentage*	*Detected*
1	$100m^3/day$	$1.44m^3/day$	1.44%	Yes
2	$100m^3/day$	$5.74m^3/day$	5.74%	Yes
3	$100m^3/day$	$10.25m^3/day$	10.25%	Yes
4	$100m^3/day$	$20.04m^3/day$	20%	Yes
5	$100m^3/day$	$20.04^3/day$	20%	Yes
6	$300m^3/day$	$3m^3/day$	1%	Yes
7	$300m^3/day$	$15.26m^3/day$	5.08%	Yes
8	$300m^3/day$	$1.25m^3/day$	10%	Yes
9	$300m^3/day$	$2.5m^3/day$	20%	Yes
Data Length: 399 $hours$		Leak Event: 9	Content: Oil, Gas	

over 100 detection results. From *TPR* which represents the percentage of detected leaks among all leak events, it can be seen that most of the leak events are successfully detected. Furthermore, by comparing the *FP* results of the two stages, it shows that the second stage can greatly decreases the number of *FP*. Moreover, it should be noted that the detection performance is related to the choice of m, the length of pressure template. The appropriate value of m is dependent on multiple factors regarding the intensity of pressure fluctuations. The relation between detection performance and m is further investigated and discussed in subsection IV-E and the results listed in Table.2 are based on the chosen values.

More detailed detection results for each site are given as follows.

1) EXPERIMENT ON DATA FROM SITE-1

The leak events description of Site-1 is listed in Table.3. There are 9 leak events experimented on an oil and gas pipeline at different time instants. All the 9 leaks are successfully detected even for the small leak in the first event. Details of the first and second stage results are depicted in Fig. 12.

- First stage result of Site-1: The top plot in Fig. 12 shows the overview of the first-stage detection result. It can be observed that many suspicious leaks are detected. There are totally 435 alarms in the first-stage result where only 9 of them are TP. Considering the number of total events is over 71700, even with the number of 426 false alarms, the *Accuracy* (ACC) is still higher than 99%. However, more than 400 false alarms may be troublesome in practice.
- Second stage result of Site-1: The middle and bottom plots in Fig. 12 demonstrate the second-stage detection result of Site-1. It shows that the alarms number has been dramatically decreased from 435 to 13 where the 9 cases of TP are all detected and the false ones are mostly eliminated. Referring to Table.3, even the $1st$ and $6th$ leak events with small leak sizes are successfully detected.

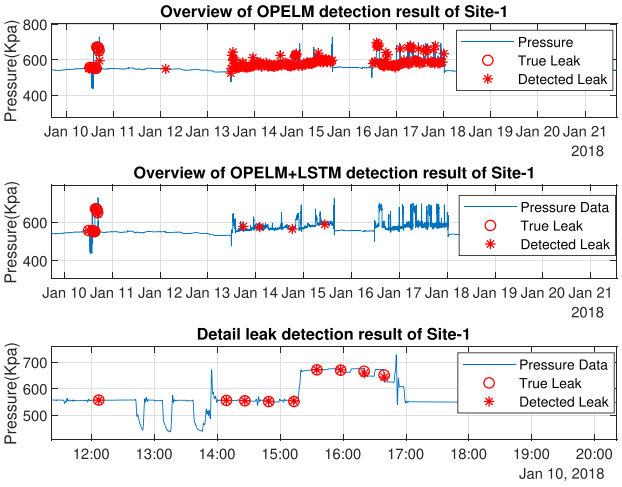


FIGURE 12. Detection result on Site-1.

TABLE 4. Experiment environment of Site-2.

Index	*Pipe Flow*	*Leak Flow*	*Leak Percentage*	*Detected*
$1-3$	$100m^3/day$	$20m^3/day$	20%	Yes
$4-7$	$20m^3/day$	$1m^3/day$	5%	No
$8-10$	$10m^3/day$	$0.25m^3/day$	2.5%	Yes
Data Length: 487 $hours$		Leak Event: 10	Content:Salt water	

2) EXPERIMENT ON DATA FROM SITE-2

The leak events description of Site-2 is listed in Table.4. There are 10 leaks occurred under different conditions in terms of flow rate and leak size.

- First stage result of Site-2: The top plot in Fig. 13 shows the first-stage detection result of Site-2. It can be observed that similar to the result of Site-1, many alarms have been reported including the true and false ones. In this case, totally 138 alarms are detected where only 10 TP events exist. The suspicious leaks mainly appear when there is a sudden drop of the pressure wave. Although the FP number is less than that in Site-1,

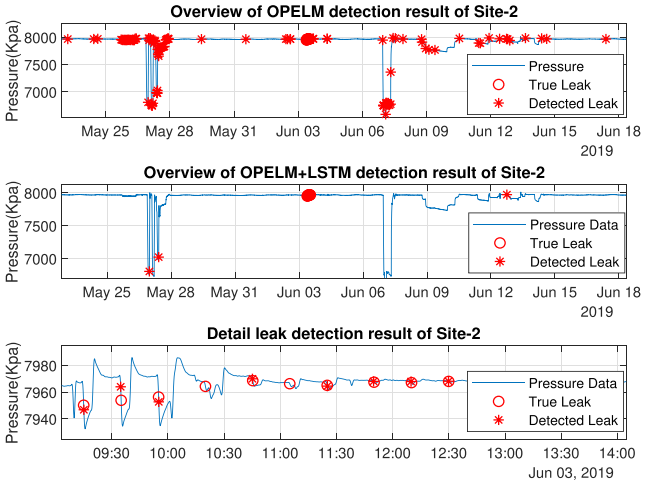


FIGURE 13. Detection result on Site-2.

the false alarms appearing every few hours may still jeopardize the entire detection efficiency.

- Second stage result of Site-2: The middle and bottom plots in Fig. 13 show the final detection result of Site-2. The number of false alarms is significantly decreased from 138 to 13 in the 487 hours long data set. In the detailed view of the second-stage result, 2 leak events are missed. In fact, it should be noted that the second-stage BiLSTM is adopted to mainly remove the large number of false alarms. However it may induce a slight increase of missed leak detection at the same time. By a close inspection of the two missed leaks, it can be found that their dropping trends are not smoothly downward like other leak events. There exist dropping rate variations in the pressure wave. The proposed method mistakenly take these two events as noises instead of TP. In general, most of the leak events are accurately detected with very few false alarms in the final result. Even for the small leaks $8-10$ listed in Table.4, they are successfully detected.

3) EXPERIMENT ON DATA FROM SITE-3

Table 5 shows the leak status of Site-3 experiment. There are 18 leak events under multiple pressure and leak flow conditions. It can be observed from Fig. 14 that the pressure wave is fluctuating continuously in a large amplitude. Such pressure fluctuation is the normal working state caused by pumping or other pipeline operations. However, the sudden and continuous pressure change may lead to numerous FP and deteriorate the detection performance.

- First stage result of Site-3: The top plot in Fig. 14 shows the first-stage detection result. In this case, there are 460 suspicious leak events detected in the 668 hours long experiment data. Comparing to the working conditions of Site-1 and Site-2, the pressure wave fluctuates continuously which may increase the amount of false alarms.

TABLE 5. Experiment environment of Site-3.

Index	*Pipe Flow*	*Leak Flow*	*Leak Percentage*	*Detected*
$1-3$	$347.5m^3/day$	$55m^3/day$	16%	Yes
$4-6$	$347.5m^3/day$	$15.6m^3/day$	4.5%	Yes
$7-9$	$347.5m^3/day$	$8.7m^3/day$	2.5%	Yes
$10-12$	$82.5m^3/day$	$20.6m^3/day$	25%	Yes
$13-15$	$82.5m^3/day$	$3.7m^3/day$	4.5%	Yes
$16-18$	$82.5m^3/day$	$2.0m^3/day$	2.5%	No
Data Length: 668 $hours$	Leak Event: 18	Content:Salt water		

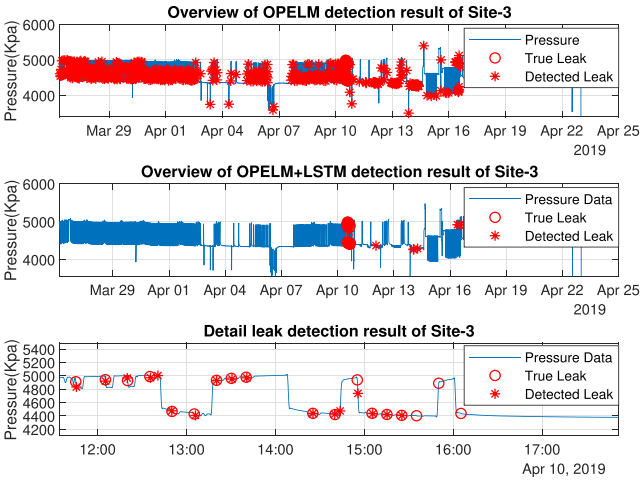


FIGURE 14. Detection result on Site-3.

TABLE 6. Experiment environment of Site-4.

Index	*Pipe Flow*	*Leak Flow*	*Leak Percentage*	*Detected*
$1-4$	$200m^3/day$	$28.4m^3/day$	14.4%	Yes
$5-7$	$200m^3/day$	$10m^3/day$	5%	Yes
$8-10$	$200m^3/day$	$5m^3/day$	2.5%	No
$11-13$	$150m^3/day$	$34m^3/day$	22.66%	Yes
$14-15$	$150m^3/day$	$7.5m^3/day$	5%	Yes
$16-17$	$150m^3/day$	$30m^3/day$	20%	Yes
$18-19$	$150m^3/day$	$7.5m^3/day$	5%	No
Data Length: 670 $hours$	Leak Event: 19	Content:Salt water		

- Second stage of Site-3: On the second stage as shown in the middle and bottom plot in Fig. 14, the number of alarms decreased from 460 to 25 where 15 alarms are TP. The last three leak events are missed. It can be seen from Table.5 that the last three leaks are relatively smaller than the others, at the same time the flow rate is also lower. It indicates that under the conditions of low pipeline flow rate and noisy pressure environment, small leaks may not be successfully detected.

4) EXPERIMENT ON DATA FROM SITE-4

Table 6 shows the leak status of Site-4 experiment. There are 19 leak events under various pressure and leak flow conditions in a noisy environment. The $8th$, $9th$ and $19th$ leak events are failed to be detected. The details are shown below.

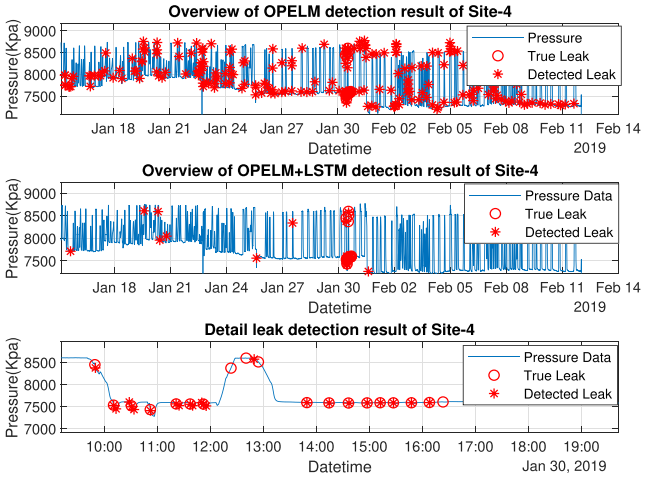


FIGURE 15. Detection result on Site-4.

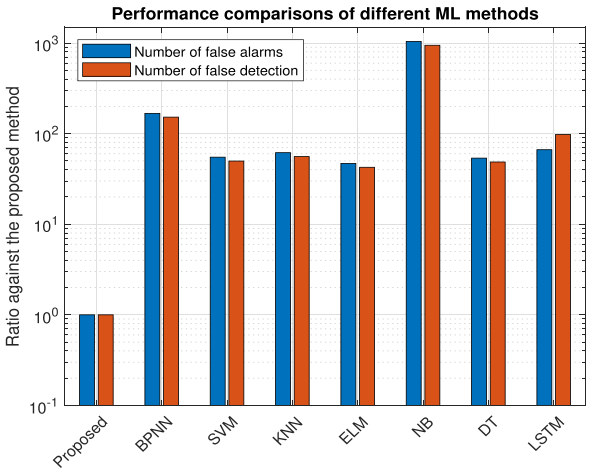


FIGURE 16. Performance comparison among different ML methods.

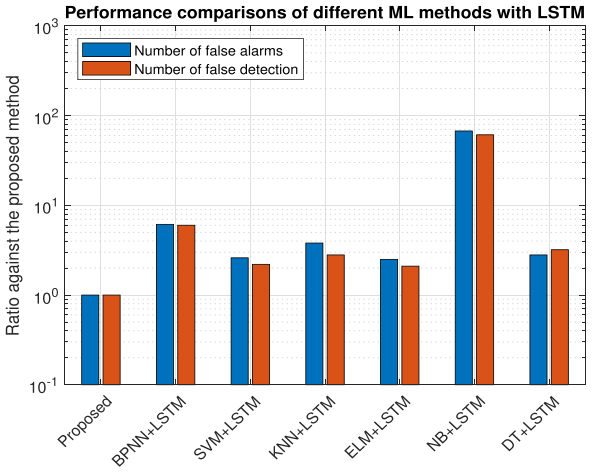


FIGURE 17. Performance of improved ML detection methods.

- First stage result of Site-4: As shown in the top plot in Fig. 15, in the first stage, 280 suspicious leaks are alarmed. Most of them appear at the edges of the dropping pressure wave. Obviously, the frequently appearing pressure drops are not leaks. By taking into account of its ambient pressure characteristics, false alarms can be discriminated in the next step.
- Second stage result of Site-4: In this stage, the number of alarms is reduced from 280 to 29. As shown in the middle and bottom plots of Fig. 15, the 8*th*, 9*th* and the 19*th* leak events are not successfully detected. From Table.6, their leak sizes are relatively smaller than the others. Moreover, by viewing the pressure wave of the 8*th* leak, it exists in a rising pressure trend which is opposite to the typical dropping trend of a leak. It can be seen that when a leak occurs in the uprising part of a pressure wave, the leak may not be successfully detected.

C. DETECTION PERFORMANCE COMPARISON

In this section, several conventional ML based leak detection methods such as the BPNN in [9], SVM in [10], KNN in [13], ELM in [14], NB in [16] and DT in [17] are compared with the proposed method, and BiLSTM is also tested in the first-stage detection. Furthermore, to verify the effectiveness of BiLSTM in eliminating false alarms, the aforementioned ML methods with BiLSTM added as the second stage are also implemented. The performances are evaluated on 100 rounds of experiments on data from Site-2.

1) PERFORMANCE COMPARISON BETWEEN VARIOUS ML METHODS AND THE PROPOSED METHOD

In this section, two key factors including the number of false alarms (FP) and the number of false detection (false alarm and missed alarm) are adopted for comparison among various ML methods. The ratio values between results from other investigated methods and the proposed one are shown in Fig. 16.

It can be observed that the proposed method achieves the least number of false alarms and false detection. The second best result is achieved by ELM. Among all other ML methods, Naive Bayes method has the worst performance where the amount of false alarms and false detection are almost 1000 times greater than the proposed method. Notably, the performance of applying BiLSTM only without the first stage detection is not satisfactory, in fact it may be even worse than other conventional methods such as SVM, KNN and ELM. Although the strong memorizing ability of BiLSTM can be utilized to effectively identify long-short term sequence patterns, its sensitivity to the transient change may not be comparable to other methods.

2) PERFORMANCE COMPARISON BETWEEN VARIOUS ML METHODS AFTER ATTACHING BiLSTM

Fig. 17 shows the performance comparison between the proposed OPELM+LSTM method and various ML methods

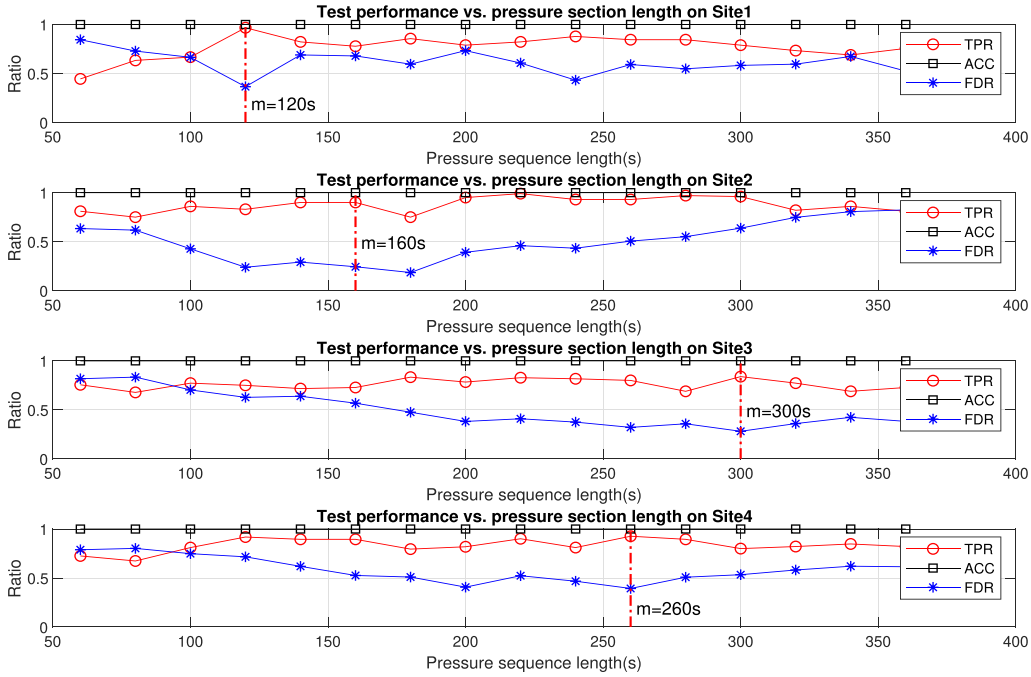


FIGURE 18. Detection performance against different pressure section length *m*.

combined with the BiLSTM to improve the detection performance.

It can be seen that all the investigated methods achieve much less false alarms and false detection when attaching BiLSTM as the second stage to identify false alarms. Take the SVM+LSTM as an example, the false alarm number decreases from 190 to 4 and the false detection number drops from 4800 to 130 out of 116160 samples. It verifies that by considering a broader temporal range of observations around the suspicious leaks, the false alarms can be effectively eliminated.

Among all the ML methods combined with BiLSTM, the proposed method still achieves the best performance, owing to the excellent classification performance of the OPELM.

D. FIRST-STAGE CLASSIFIER SELECTION

To choose the most appropriate classifier to perform the first stage detection, some comparisons in terms of learning speed and test accuracy are included. The learning speed and accuracy among aforementioned methods are shown in Table. 7. The experiment is based on the training set sized at 20000 within which the amounts of class0 and class1 samples are equal. The numbers shown are the averaged value of 100 experiments. For methods such as BPNN, ELM and OPELM, the number of hidden layer neurons is set to be 40. The kernels in SVM and OPELM are chosen as radial basis functions (RBF).

It can be seen that KNN and NB have the fast learning speed, but the test accuracy is obviously lower than OPELM and SVM. The three classifiers with test accuracy higher than 0.9 are SVM, ELM and OPELM, however, SVM takes much longer in learning. The accuracy of ELM is obviously lower than OPELM although the learning speed is faster. Therefore, considering both of the learning speed and test accuracy, OPELM is deemed the most effective one among the investigated methods to perform the first-stage leak detection in this study.

TABLE 7. Learning speed and accuracy comparison.

Methods	*BPNN*	*SVM*	*KNN*	*NB*	*ELM*	*OPELM*
Time(s)	9.54	3.26	0.06	0.02	0.16	0.81
Accuracy	0.854	0.972	0.863	0.832	0.922	0.974

E. EFFECTS OF TEMPLATE LENGTH

The length of pressure template *m* defines the time span of the pressure portion used for feature extraction. The value of *m* is closely related to the quality of feature representation. When *m* is too small, the template may not be able to represent the entire leak pressure characteristics, and too large *m* may lead to adverse influence by non-leak pressure wave, thus deteriorates the feature representation. The appropriate value of *m* is dependent on multiple factors affecting waveform fluctuations such as pipe flow rate, leak size, pipe content, pump or valve operations etc.. It usually varies from case to case.

Fig. 18 demonstrates the detection performance with respect to different template length *m* of the four investigated data sets. All ACC results of the 4 sites indicating the overall detection accuracy are close to 1, which show that ACC

is not sensitive to m since the detection error is generally small.

Hence, to choose a proper value of m, both of TPR and FDR are considered. Higher TPR means higher percentage of leak events are to be successfully detected. However, FDR is expected to be small because it indicates the percentage of false alarms counted in all alarms. Therefore, the value of m is chosen at which the difference between TPR and FDR is maximum. As can be observed in Fig. 18, the first two m values are 120 and 160 respectively. When the intensity of fluctuations increases, it will need longer templates to achieve a satisfied detection performance. The fluctuations in Site-3 and Site-4 are more dramatic than that in the first two cases, therefore m is chosen as 300s and 260s respectively.

V. CONCLUSION

In this paper, a novel PPA leak detection method is proposed based on supervised OPELM combining BiLSTM for continuous pressure monitoring leak detection. The contributions of this paper are summarized as follows.

Firstly, the proposed method can achieve higher detection accuracy with significantly lower false alarm rate than existing ML based PPA leak detection methods.

Secondly, the advantage of strong memorizing capability of BiLSTM is firstly utilized to discriminate the true and false alarms. The effectiveness is verified by experiments on different real-world data sets.

Finally, the proposed feature extraction scheme can effectively represent the characteristics of leak pressure transient, thus, enhance the detection performance. Furthermore, a thorough comparison study is performed.

However, when minor leaks occur resulting in very subtle pressure variations, or when leaks coincide with abrupt up-trend (i.e., increasing) pressure changes, they may not be effectively detected by the currently proposed approach. In this case, other detection methods with higher sensitivity can be investigated, but the trade-off between sensitivity to minor leaks and false alarm rate needs to be carefully considered.

In the future research, more deep learning based methods are to be explored in the pressure wave analysis for distributed pipeline leak detection. The feature extraction and leak spot localization will be investigated using observations from multiple distributed wireless pressure sensors.

REFERENCES

[1] J. Conca. Pick Your Poison for Crude—Pipeline, Rail, Truck or Boat. Forbes. Accessed: Apr. 26, 2014. [Online]. Available: https://www.forbes.com/sites/jamesconca/2014/04/26/pick-your-poison-for-crude-pipeline-rail-truck-or-boat/#6b7f23a717ac

[2] A. Abdulshaheed, F. Mustapha, and A. Ghavamian, "A pressure-based method for monitoring leaks in a pipe distribution system: A review," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 902–911, Mar. 2017.

[3] U. Baroudi, A. A. Al-Roubaiey, and A. Devendiran, "Pipeline leak detection systems and data fusion: A survey," *IEEE Access*, vol. 7, pp. 97426–97439, 2019.

[4] X. Diao, G. Shen, J. Jiang, Q. Chen, Z. Wang, L. Ni, A. Mebarki, and Z. Dou, "Leak detection and location in liquid pipelines by analyzing the first transient pressure wave with unsteady friction," *J. Loss Prevention Process Industries*, vol. 60, pp. 303–310, Jul. 2019.

[5] I. Santos-Ruiz, J. R. Bermúdez, F. R. López-Estrada, V. Puig, L. Torres, and J. A. Delgado-Aguiñaga, "Online leak diagnosis in pipelines using an EKF-based and steady-state mixed approach," *Control Eng. Pract.*, vol. 81, pp. 55–64, Dec. 2018.

[6] Q. Jiang, S. Yan, H. Cheng, and X. Yan, "Local-global modeling and distributed computing framework for nonlinear plant-wide process monitoring with industrial big data," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 21, 2020, doi: 10.1109/TNNLS.2020.2985223.

[7] Q. Jiang, X. Yan, and B. Huang, "Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes," *Ind. Eng. Chem. Res.*, vol. 58, no. 29, pp. 12899–12912, Jul. 2019.

[8] M. B. Abdulla, R. O. Herzallah, and M. A. Hammad, "Pipeline leak detection using artificial neural network: Experimental study," in *Proc. Int. Conf. Modelling, Identificat. Control (ICMIC)*, Cairo, Egypt, Aug. 2013, pp. 328–332.

[9] R. B. Santos, M. Rupp, S. J. Bonzi, and A. M. F. Fileti, "Comparison between multilayer feedforward neural networks and a radial basis function network to detect and locate leaks in pipelines transporting gas," *Chem. Eng. Trans.*, vol. 32, pp. 1375–1380, Jun. 2013.

[10] H. Chen, H. Ye, C. Lv, and H. Su, "Application of support vector machine learning to leak detection and location in pipelines," in *Proc. 21st IEEE Instrum. Meas. Technol. Conf.*, Como, Italy, May 2004, pp. 2273–2277.

[11] R. Xiao, Q. Hu, and J. Li, "Leak detection of gas pipelines using acoustic signals based on wavelet transform and support vector machine," *Measurement*, vol. 146, pp. 479–489, Nov. 2019.

[12] S. Rashid, U. Akram, and S. A. Khan, "WML: Wireless sensor network based machine learning for leakage detection and size estimation," *Procedia Comput. Sci.*, vol. 63, pp. 171–176, Sep. 2015.

[13] A. Rojik, Endroyono, and A. N. Irfansyah, "Water pipe leak detection using the k-nearest neighbor method," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Aug. 2019, pp. 393–398.

[14] H. Zhang, Q. Li, X. Zhang, and W. Ba, "Industrial oil pipeline leakage detection based on extreme learning machine method," in *Advances in Neural Networks—ISNN*, F. Cong, A. Leung, and Q. Wei, Eds. Hokkaido, Japan: Springer, 2017, pp. 380–387.

[15] A. E. U. Salam, M. Tola, M. Selintung, and F. Maricar, "Application of SVM and ELM methods to predict location and magnitude leakage of pipelines on water distribution network," *Int. J. Adv. Comput. Res.*, vol. 5, no. 19, pp. 140–144, Jun. 2015.

[16] A. Soldevila, R. M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig, "Leak localization in water distribution networks using Bayesian classifiers," *J. Process Control*, vol. 55, pp. 1–9, Jul. 2017.

[17] Z. Chen, X. Xu, X. Du, J. Zhang, and M. Yu, "Leakage detection in pipelines using decision tree and multi-support vector machine," in *Proc. 2nd Int. Conf. Electr., Control Autom. Eng. (ECAE)*, Xiamen, China, Dec. 2018, pp. 327–331.

[18] T. R. Sheltami, A. Bala, and E. M. Shakshuki, "Wireless sensor networks for leak detection in pipelines: A survey," *J. Ambient Intell. Humanized Comput.*, vol. 7, no. 3, pp. 347–356, Jun. 2016.

[19] J. Reynolds and A. Kam, "An evaluation of negative pressure wave leak detection: Challenges, limitations, and use cases," in *Proc. Pipeline Simulation Interest Group Annu. Meeting*, London, U.K., May 2019, pp. 1–28.

[20] C. Ma, S. Yu, and J. Huo, "Negative pressure wave-flow testing gas pipeline leak based on wavelet transform," in *Proc. Int. Conf. Comput., Mechatronics, Control Electron. Eng.(CMCE)*, vol. 5, Aug. 2010, pp. 306–308.

[21] Z. Peng, J. Wang, and X. Han, "A study of negative pressure wave method based on Haar wavelet transform in ship piping leakage detection system," in *Proc. IEEE 2nd Int. Conf. Comput., Control Ind. Eng. (CCIE)*, vol. 2, Aug. 2011, pp. 306–308.

[22] C. H. Tian, J. C. Yan, J. Huang, Y. Wang, D.-S. Kim, and T. Yi, "Negative pressure wave based pipeline leak detection: Challenges and algorithms," in *Proc. IEEE Int. Conf. Service Oper. Logistics, Informat. (SOLI)*, Suzhou, China, Jul. 2012, pp. 372–376.

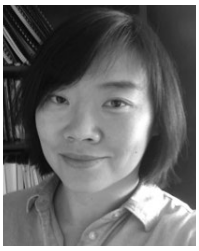
[23] J. Liu, D. Zang, C. Liu, Y. Ma, and M. Fu, "A leak detection method for oil pipeline based on Markov feature and two-stage decision scheme," *Measurement*, vol. 138, pp. 433–445, May 2019.

[24] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[25] H.-J. Rong, Y.-S. Ong, A.-H. Tan, and Z. Zhu, "A fast pruned-extreme learning machine for classification problem," *Neurocomputing*, vol. 72, nos. 1–3, pp. 359–366, Dec. 2008.
[26] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, "OP-ELM: Optimally pruned extreme learning machine," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 158–162, Jan. 2010.
[27] M. Sundermeyer, H. Ney, and R. Schluter, "From feedforward to recurrent LSTM neural networks for language modeling," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 3, pp. 517–529, Mar. 2015.
[28] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018.
[29] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

LEI YANG received the B.Sc. degree in electrical engineering from Jilin University, Changchun, China, in 2001, and the M.S. degree in information theory and engineering from the Henan University of Science and Technology, Luoyang, China, in 2006. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada.

He has been working in the field of designing and developing measurement/test systems in various industrial scenarios, from 2006 to 2016. His research interests include machine learning in wireless sensor networks, data fusion, localization, communication and fault diagnosis, and so on.



QING ZHAO (Member, IEEE) received the B.Sc. degree in control engineering from Northeastern University (NEU), China, and the Ph.D. degree in electrical engineering from the University of Western, London, ON, Canada.

During her sabbatical leave, she was a Visiting Professor with the Control Engineering Department, Université Libre de Bruxelles, Brussels, Belgium, from August 2008 to December 2008. She was a Visiting Professor with the University of Duisburg-Essen, Duisburg, Germany, from January to July 2009. She is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Canada. Her current research interests include fault diagnosis, fault tolerant control, robust control, stochastic control and machine condition monitoring. She received the A. V. Humboldt Research Fellowship, in 2009.