Interpretable PCA and SVM-Based Leak Detection Algorithm for Identifying Water Leakage Using SAR-Derived Moisture Content and InSAR Closure Phase

Yan Yan, *Member, IEEE*, Xujie Le, *Member, IEEE*, Taoli Yang, *Senior Member, IEEE*, and Hanwen Yu, *Senior Member, IEEE*

***Abstract*—The sewer/water leakage has become a worldwide concern. Conventional field measurement method is cost-consuming and time-inefficient, and deep learning (DL)-based method is not readily to be interpreted. Under this condition, we propose an interpretable principal component analysis (PCA) and support vector machine (SVM)-based leak detection algorithm (PSLDA) for identifying the sewer/water leakage. The PSLDA takes synthetic aperture radar (SAR)-derived moisture content and SAR interferometry closure phase as inputs, conducts feature extraction and dimensionality reduction with PCA, and implements binary classification with SVM, finally identifying leaks or nonleaks. The main advantage of PSLDA lies in that it, respectively, adopts the mathematically derived PCA and SVM instead of the convolutional and pooling layer and the loss-function layer in the DL network, ensuring the validity and interpretability simultaneously. With 1222 in situ leak points and 1206 nonleak points in Beijing and Tianjin, China, the PSLDA is trained. The well-trained PSLDA is subsequently applied to detect leak/nonleaks in real-world scenarios, and achieves a leak-detection accuracy of 91.86% in the fifth ring road of Beijing, a nonleak-detection accuracy of 88.10% in the Huizhou City, Guangdong Province. The proposed PSLDA demonstrates to be efficacious and credible, potentially offering explicit guidance for pipeline maintenance.**

***Index Terms*—Leak/nonleak detection, principal component analysis (PCA) and support vector machine (SVM) (PCA-SVM)-based leak detection algorithm (PSLDA), principal component analysis (PCA), synthetic aperture radar (SAR) interferometry (InSAR), support vector machine (SVM), SAR.**

I. Introduction

THE underground pipeline is widely used in water supply, sewage, and heating, etc. In China, the total length of water supply pipelines has been increasing with the urbanization to nearly 1 102 975.660 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-water-pipeline/cn-length-of-water-pipeline-city) and drainage pipelines to 913 507.590 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-drain-pipeline) until 2022. However, the pipelines have a high risk of leakages, which may be attributed to management negligence, improper operations, material aging, and mechanical forces, etc. [1]. In recent years, water pipeline leakages have become a worldwide concern since it is closely related to three main problems, i.e., water scarcity, human health, and secondary disasters. It is reported that around 27% of the global population has been suffering from water scarcity since the middle of 2010, and the proportion is predicted to be over 42% until 2050 [2], [3]. Around 50% of the world's population inhabits areas suffer from water scarcity at least one month a year [3], [4]. Continuous leakage will lead to the loss of clean water, thus exacerbating the water scarcity situation. It also contradicts to the global advocacy of green sustainable development. In addition, in the presence of leak, the pipe-water, which is originally safe, tends to be contaminated during transportation, posing threatens to human lives. Furthermore, the sewer and/or water pipeline leakage can induce secondary environmental hazards, such as groundwater pollution, infrastructure damage, soil erosion, and ground subsidence, resulting in tremendous financial losses [1], [5]. For example, the leakage has been one of driving factors of man-made sinkholes. From 2006 to 2010, the Florida Office of Insurance Regulation reported 24 671 cases of sinkholes that caused financial damage of around 1.4 billion USA dollars [6]. To minimize the damages, it is imperative to initiate leak detection, which can provide timely decision-making assistance for the repair and maintenance of underground pipelines [1], [5].

Traditionally, the leakage acoustic-based method is frequently used for leak detection [3], [7], [8]. The method relies on in situ devices (e.g., geophones, hydrophones, pressure sensors, piezoelectric transducers, and accelerometers) and has defects as listed in the following.

1) The devices are not only cost-consuming but also inconvenient to deploy. The water pipeline is buried underground and is above the groundwater level [9]. The depth of

Manuscript received 2 July 2024; revised 25 July 2024; accepted 3 August 2024. Date of publication 13 August 2024; date of current version 5 September 2024. This work was supported by the National Natural Science Foundation of China under Grant 42271456. *(Corresponding author: Taoli Yang.)*

The authors are with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: yanyan_river@163.com; lexujie123@163.com; yangtl@uestc.edu.cn; yuhanwenxd@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2024.3443127

shallow groundwater roughly varies from 0 to ∼5 m, making it hard to install devices.

2) The detection accuracy is dependent on pipeline materials, i.e., metal or nonmetal. Tariq et al. [3] developed separate models for metal and nonmetal pipes and revealed that the detection accuracy of metal pipeline is obviously higher than that of nonmetal pipeline.
3) The method is inefficient and impractical for long-range detection, whereas the pipelines are basically constructed with considerable lengths.

Under the circumstances, it is of necessity and urgency to develop other effective and efficient leak detection method.

Straightforwardly, the post-leakage moisture should be higher than the preleakage moisture. Thus, the soil moisture content would facilitate identifying leakage if it is properly derived. The synthetic aperture radar (SAR) transmits microwaves and receives backscatter signals of ground targets with large-spatial extent [10], [11]. The microwaves can penetrate asphalt and concrete of urban surfaces and probe into the underground condition [12], [13]. The longer the wavelength, the deeper the penetration depth. The L-band satellite (radar wavelength $\lambda = 24$ cm) proves to be able to penetrate the Earth's surface with a depth of approximately 10 feet [12]. In the presence of water leakage, the dielectric permittivity of underground soil varies [14]. Such variation further influences the radar backscattering signal, thus making SAR capable of deriving moisture content. Currently, SAR-based moisture content derivation algorithm is mature and prevalent [15], [16]. For example, Guo et al. integrated the $H/A/\alpha$ polarmetric decomposition with support vector regression (SVR) to derive the surface moisture content, and achieved a high accuracy of 1.50% per volume (vol%) [16]. In a word, the SAR-derived moisture content provides an opportunity for detecting large-scale leakage without distinguishing the pipe materials.

In reality, SAR-derived moisture content alone is insufficient for locating leakage because there distribute targets naturally characterized by high moisture content, such as artificial pond, shallow vegetation, and farmland. Hence, extra information reflecting the variation of moisture content is needed. The closure phase of SAR interferometry (InSAR) demonstrates to be an indicator of moisture content variation without being distracted by surface deformation and atmospheric disturbance [17]. As will be detailedly described later, it can be reasoned from a mathematical perspective that pixel having closure phase deviated from an integer multiple of 2π is featured with apparent moisture content variation. In brief, the moisture content derived from SAR backscattering coefficient and the moisture variation reflected by SAR phase should be complementary to each other, and their conjunction could, thus, provide more information for accurate leakage detection.

However, how to utilize the moisture content feature and closure phase feature for detecting leak is still ambiguous. For instance, to what extent the moisture content increase or to what extent the closure phase varies, a pixel should be identified as a leakage? In this case, a neural network (NN) [1], [18] algorithm may help since it can detect a binary change, i.e., leak or nonleak, if adequately trained. The SAR observations could offer the needed inputs for training and implementing the NN algorithm to detect the pipeline leak. As a matter of fact, pipeline leak-detection methods based on deep learning (DL) emerge one after another [19], [20], [21], [22]. For example, considering that the closed circuit television (CCTV) inspection is expertise-intensive and time-consuming, Li et al. [19] proposed a method based on the deep convolutional neural network (CNN) to detect and classify defects from CCTV inspections. A detection accuracy of 83.2% was obtained. Sinha and Fieguth [20] proposed a new neuro-fuzzy classifier that combines NNs and fuzzy concepts for the classification of objects in segmented underground pipe image, with classification accuracies around 90% on real concrete pipe images. Kang et al. [21] presented a leakage monitoring architecture by using the ensemble one-dimensional (1-D)-CNN-support vector machine (SVM) and a graph-based optimal localization algorithm. During CNN network operation, the few actual leak samples could give rise to the imbalanced dataset problem. To address this issue, Hu et al. [22] proposed a minor class-based status detection method using enhanced generative adversarial networks. In a word, the DL-based leak detection method is more efficient than the conventional method, and is of relatively high accuracy.

Unfortunately, the CNN or DL network basically constitutes the convolutional layer, pooling layer, and loss-function layer. The convolutional layer is used for feature extraction. The pooling layer aims to reduce the feature dimension extracted by convolutional layer. The loss-function realizes classification. The convolution and pooling, and loss-function optimization processes in the network are complicated and with the closed box fashion. The weights of these layers are obtained through iterative training with samples, making it unexplainable. Currently, the researchers are increasingly devoted to exploring the possibility of unraveling complicated processes with explainable procedure and expecting that the closed box would be ultimately opened. Under this condition, the leak-detection method which not only shares high efficiency and accuracy with DL-based method, but also has interpretability, deserves further exploration. It is accepted that the principal component analysis (PCA) [23], [24] can obtain the proper information or extract features and realize dimensionality reduction, making it possible to replace the convolutional layer and pooling layer of DL network. Endowed with an optimization component, the mathematically interpretable SVM [25], [26], [27] could potentially replace the loss-function optimization layer in the DL model to implement binary classification. Based on the discussions above, the definition of the research problem in this article can be given as follows.

Problem definition: Based on soil moisture content and InSAR closure phase, we aim at developing an interpretable PCA-SVM-based leak detection algorithm (PSLDA) for identifying sewer/water pipeline leaks/nonleaks.

The proposed algorithm is innovative in two aspects. First, PSLDA can detect leaks/no-leaks with high accuracy and large spatial extent in comparison with conventional method. Second, the algorithm adopts PCA to conduct feature extraction and dimensionality reduction, applies SVM to conduct classification, ensuring its interpretability.

The rest of this article is organized as follows. In Section II, the basic principle of PSLDA is introduced, and the framework is designed. Following the framework, linear SVM training of PSLDA is implemented in Section III. In Section IV, the performance of PSLDA is scrutinized through real leak detection and nonleak detection scenarios. It is worth mentioning that the performance of PSLDA is not compared with DL-based methods because this work aims for interpretability but the DL network is unexplainable. In Section V, the performance of nonlinear SVM is discussed and compared with linear SVM. Finally, Section VI concludes this article.

II. Principle and Design of PSLDNET

Under the circumstances of interpretable binary classification by using SAR-derived data, the PSLDA is designed to constitute four modules, i.e., the input data module, the interpretable feature-extraction module, the interpretable classification module, and the output module. Since the wanted output is unambiguous, i.e., leak or nonleak distribution map, the other three modules are elaborately introduced as follows.

A. Input Module

The input module aims to prepare input data for the subsequent module. The input data include SAR-derived moisture content and closure phase.

The moisture content could be obtained using the algorithm proposed in [16]. The algorithm integrates the *H/A/α* decomposition and the SVR model to retrieve the moisture content. Wherein, the *H/A/α* decomposition is used to weaken vegetation disturbance and SVR is used to address the ill-posed moisture retrieval problem, finally making the integrated algorithm characterized by vegetation-independence, agreeable transferability, and high accuracy of 1.5 vol%. The algorithm has gained successful applications [12], [16], [28]. Given N SAR single look complex (SLC) images relative to the same area and acquired at the ordered times (t_1, t_2,..., t_N), N moisture content images are obtained.

Provided that $N \geq 3$, it is feasible to compute the closure phase. For simplicity, we consider $N = 3$ to depict the basic principle of closure phase, and the principle is also applicable to the case $N > 3$. With three SAR images, one can obtain three differential interferograms, that is, $\varphi_{i,j}$ ($i < j$ and $i, j = 1, 2, 3$) between SAR acquisitions at t_i and t_j, respectively. The term "differential" means that the flat-Earth phase and the topographic phase have been successfully removed. Then, the closure phase is defined as [17]

$$\Phi(s) = \phi_{12}(s) + \phi_{23}(s) - \phi_{13}(s) \tag{1}$$

where $\Phi(s)$ denotes the closure phase. $\phi_{ij}(s)$ ($i < j$ and $i, j = 1, 2, 3$) is the unwrapped phases of $\varphi_{ij}(s)$. The relationship between $\phi_{ij}(s)$ and $\varphi_{ij}(s)$ can be expressed as

$$\phi_{ij}(s) = \varphi_{ij}(s) + 2k_{ij}(s)\pi \quad (k_{i,j} \in \text{integer}). \tag{2}$$

Assume that that $d_{i,j}$ ($i < j$ and $i, j = 1, 2, 3$) represents the ground surface displacement along the LOS direction. Notably, $d_{12} + d_{23} = d_{13}$. Because the deformation-related absolute phase is only dependent on surface displacement and radar wavelength, i.e., $4\pi d/\lambda$, the deformation-related closure phase is naturally zero. Similarly, the atmospheric delay phase is induced by different atmospheric condition at acquisitions t_1, t_2, and t_3. The atmospheric delay phase between acquisitions t_1 and t_2 plus that between t_2 and t_3 is equal to the atmospheric delay phase between t_1 and t_3, making the atmosphere-related closure phase intrinsically zero. In summary, the closure phase is inherently distracted from the Earth's surface deformation and atmospheric disturbance [29].

De Zan et al. [17] and Zwieback et al. [30] observed that changes in soil moisture and in the water content of vegetation could lead to the closure phase excess and deficit, or called phase inconsistencies. From the perspective of mathematics, we have

$$\begin{cases} \Phi(s) \neq 2k(s)\pi, & \text{If moisture changes} \\ \Phi(s) = 2k(s)\pi, & \text{If no moisture changes} \end{cases} \quad (k \in \text{integer}). \tag{3}$$

However, the reliability of (1) relies on the unwrapped results. Improper or erroneous phase unwrapping (PU) results would impair the effectiveness of Φ. The situation deteriorates in urban areas where the underground pipelines are mainly distributed. The reason lies in that the edges of buildings are generally characterized by abrupt interferometric fringe changes, disobeying the phase continuity assumption [31], [32]. The phase continuity assumption is the foundation of widely used single-baseline PU algorithms. Then, whether is it possible to improve the closure phase defined in (1) without performing PU procedure? The answer is yes. Specifically, if we integrate (2) into (1), we will have

$$\begin{aligned} \Phi(s) = &(\varphi_{12}(s) + 2k_{12}(s)\pi) + (\varphi_{23}(s) + 2k_{23}(s)\pi) \\ &- (\varphi_{13}(s) + 2k_{13}(s)\pi). \end{aligned} \tag{4}$$

Further integrate (4) into (3), and conduct a simple transposition of terms, we will have

$$\begin{cases} \Phi'(s) \neq 2k'(s)\pi, & \text{If moisture changes} \\ \Phi'(s) = 2k'(s)\pi, & \text{If no moisture changes} \end{cases} \quad (k' \in \text{integer})$$

$$\begin{aligned} \text{where} \quad \Phi'(s) &= \varphi_{12}(s) + \varphi_{23}(s) - \varphi_{13}(s) \\ k'(s) &= k(s) - (k_{12}(s) + k_{23}(s) - k_{13}(s)). \end{aligned} \tag{5}$$

If we let $\Phi'(s)$ subtract $2k'(s)\pi$, and still name the result as $\Phi'(s)$ (for convenience), we will have

$$\begin{cases} \Phi'(s) \neq 0, & \text{If moisture changes} \\ \Phi'(s) = 0, & \text{If no moisture changes} \end{cases} \tag{6}$$

where Φ' is the improved closure phase and $\Phi' \in (-\pi, \pi]$. Equation (6) indicates that one can compute the closure phase by using differential interferograms directly, and the pixels with water leakage should be characterized by nonzeros values. In addition, the absolute closure phase $|\Phi'|$ is often adopted since its magnitude directly reflects the variation degree of moisture contents.

In summary, given the SAR data, 2-D moisture content and 2-D closure phase images (or matrices) can be obtained in terms of the principles above. We stretch each 2-D image (matrix) into

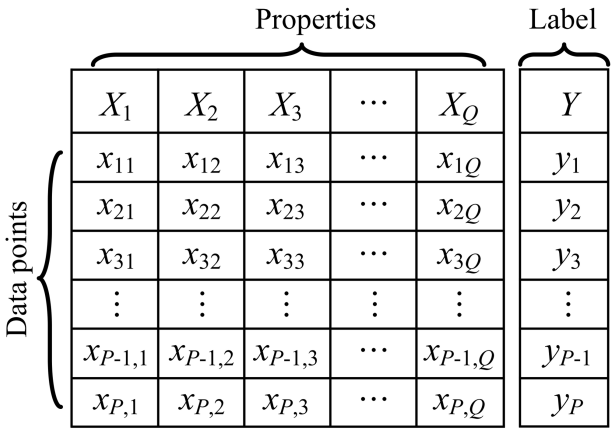


Fig. 1. Structure of the input data including *InputData* and Labels (in the case of model training). In the case of prediction, only the *InputData* is required.

a 1-D vector by column, and stack all vectors into a new matrix column-by-column. The new matrix is named as *InputData* and its structure is illustrated in Fig. 1. If there are P pixels in each image and Q different images, the *InputData* matrix should have P rows and Q columns. Each row represents an independent point, and each column corresponds to one property X_i (i =1, 2,..., Q) (i.e., moisture content or closure phase). The *InputData* has Q properties, which is high-dimensional. It is worth noting that the organization format of *InputData* is identical for irregular grid points as long as all properties for a certain pixel are aligned. In the case of model training, the labels are simultaneously required in the input module. As shown in Fig. 1, the leak point is labeled as 1 while nonleak point is labeled as 0. So far, the input data are well prepared and matrix *InputData* is provided for the subsequent feature extraction module.

B. Feature Extraction Module: PCA

As depicted above, the *InputData* originally has Q properties. There may exist correlated or redundant information among these properties, but the patterns are initially ambiguous, making feature extraction and feature dimensionality reduction inevitable. It is well-known that the convolutional and pooling layers are powerful in mining the pertinent feature information from inputs and reducing feature dimensionality. However, such module in DL is not readily to be interpreted.

Contrastively, PCA is a mature tool for analyzing and identifying the feature patterns of high-dimensional data to highlight their similarities and difference. Meanwhile, PCA has advantages in reducing data dimensions without much loss of key information. To be specific, PCA computes the eigenvalues and eigenvectors of the covariance matrix related to inputs, generating principal components (PCs) or features. The process is called feature extraction. The eigenvalue indicates the significance of PCs. PCs with high eigenvalues contain dominating features and are reserved, while other components are excluded, accomplishing feature dimensionality reduction. More mathematical introduction can be found in [23] and [24].

In brief, instead of the convolutional and pooling layers, the interpretable PCA is used as the feature extraction module of PSLDA. Given the *InputData*, the specific implementing steps of PCA is as follows.

1) Preprocess InputData: The columns in *InputData* correspond to different physical properties, i.e., moisture content or absolute closure phase. To eliminates the influence of different measurement units, it is essential to normalize and centralize the columns, respectively, and finally generate a $P \times Q$ matrix whose values are between $[-1, 1]$ with mean value of zero.

2) Calculate Covariance Matrix of Preprocessed InputData: To figure out whether the columns of *InputData* are correlated, the covariance matrix needs to be calculated. The covariance matrix consists of covariance, where positive (negative) value implies that two columns are positively (negatively) correlated while zero-value implies the columns are independent of each other. Considering the dimensionality of *InputData*, the covariance matrix should be $Q \times Q$. The matrix element at the ith row and jth column is the covariance of the ith column and jth column of preprocessed *InputData*. The element at the ith row and jth column is equivalent to that at the jth row and ith column, i.e., the matrix is symmetrical along the main diagonal.

3) Calculate the Eigenvalues and Eigenvectors of the Covariance Matrix: By taking the eigenvalues and eigenvectors of the covariance matrix, we are able to extract features of *InputData*. Since the covariance matrix is $Q \times Q$ and is real symmetric, there should be Q nonnegative eigenvalues and Q pair-wise orthogonal eigenvectors. One eigenvalue corresponds to one eigenvector. If sorted from the highest to the lowest, the eigenvalues can be obtained as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_Q$ (λ_i is the ith eigenvalue, i = 1, 2,. . ., Q). The corresponding eigenvectors are PCs, i.e., extracted features, with significance ranging from high to low. The directions of eigenvectors represent directions of PCs. Since the eigenvectors are perpendicular to each other, the extracted features can now be distinguishable along the directions of PCs.

4) Generate Dimensionality-Reduced InputData: In reference to the contributions of PCs, insignificant components are excluded. If the first K eigenvectors contain most of the information (e.g., $>$ 90%), the K eigenvectors are reserved while other (Q–K) eigenvectors are excluded. We arrange the K eigenvectors column-by-column into a matrix, then right-multiply the *InputData* by this matrix, and could finally generate a feature-dimensionality reduced *InputData*. The *InputData* is now K-dimensional, where only dominating features remain.

C. Classification Module: SVM

The dimensionality-reduced *InputData* (i.e., main features) together with labels are fed into this classification module. In the DL network, features extracted after the convoluting and pooling are fed into the loss-function layer, where a leak or nonleak decision is made with the optimization. Intrinsically, the loss-function optimization is similar to a classification algorithm, which optimizes properly-selected input features. Then, a machine-learning algorithm consisting of an optimization component can be a valid candidate for interpreting the loss-function optimization and generating the wanted output (i.e., leak or nonleak) location by location. The SVM is a suitable choice.

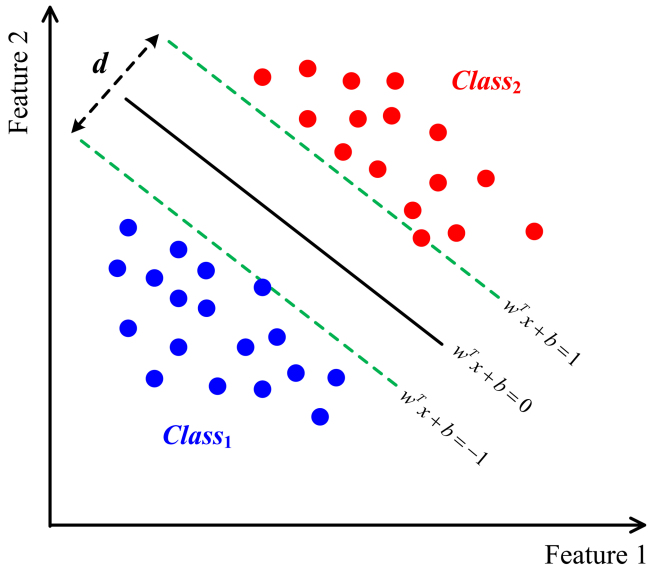


Fig. 2. Illustration of SVM hyperplane and the margins of two classes in 2-D plane.

SVM first applies the kernel function to map the nonlinear data to a higher dimensional feature space where linear separation may be possible, and then separates the two classes with a hyperplane defined as

$$w^T x + b = 0 \tag{7}$$

where w is a P-dimensional vector (determined by the row number of *InputData*) and b is a bias term. w and b are unknowns and needs to be determined by training data $x_{i,j}$ ($i = 1, 2, 3, \ldots, P, j = 1, 2, 3, \ldots, K$). If the hyperplane is determined properly, the margin distance between two classes should be maximal such that data of each class can only be in the left ($y = -1$) or in the right ($y = 1$) sides of the hyperplane. Fig. 2 illustrates the hyperplane and the margins of two classes in 2-D plane. The margins are defined as

$$w^T x + b = \begin{cases} \leq -1 \text{ for } y_i = -1 \\ \geq \quad 1 \text{ for } y_i = \quad 1. \end{cases} \tag{8}$$

Therefore, the key of separability lies in determining the optimal hyperplane [25]. However, the hyperplane theoretically can be anywhere between $w^T x + b = -1$ and $w^T x + b = 1$. To find the optimal one, SVM in turns seeks to maximize the margin distance d between two classes. The distance d could be measured with [25]

$$d(w, b; x) = \frac{\left|(w^T x + b - 1) - (w^T x + b + 1)\right|}{\|w\|} = \frac{2}{\|w\|}. \tag{9}$$

In (9), maximizing the margin distance is equivalent to minimizing the length of vector w, which can be also expressed as $w^T w/2$ [27]. Ultimately, the general convex problem to determine the optimal hyperplane can be addressed as

$$\min_{w,b} \frac{1}{2} w^T w$$

$$\text{s.t. } y_i(w^T x + b) \geq 1. \tag{10}$$

Equation (10) is a quadratic programming problem whose convex objective function can always be minimized efficiently under the given constraints [33]. In other words, SVM can always be trained to the achieve the global minimum solution in polynomial time, i.e., time-efficiency. After the optimal hyperplane has been determined, it is feasible to predict the leak or nonleak category for unseen data by determining which side of line the given point lies in.

So far, the entire framework of the PSLDA has been explicitly introduced. Following the framework, the PSLDA is subsequently trained and applied to detect the sewer/water leak or nonleak.

III. Training of PSLDA

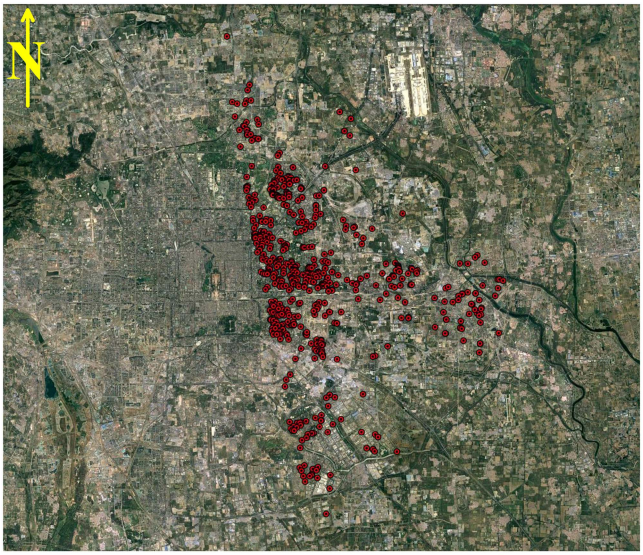
A. Training Area

In China, the average leakage ratio is relatively high for the majority of large and medium-sized cities, which resulted in an expected total loss of 5×10^6 m^3 water per year [34]. As the capital of China, Beijing is one of the eight largest cities of the world. During the period of 2011–2022, the permanent population of Beijing grew to about 21.84 million (https://data.stats.gov.cn/easyquery.htm?cn=E0105&zb=A02&reg=110000&sj=2023). The growing population pose severe pressure on demands for water supply and drainage. By the year of 2022, the total length of supply water pipeline was 19 553.00 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-water-pipeline/cn-length-of-water-pipeline-city-beijing), and the sewer pipeline length extended to 20 137.13 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-drain-pipeline) in Beijing. In addition, the pipeline distribution in Beijing is very complex with an interlaced ring-shaped pattern [34]. How to take effective and efficient measures to detect, and thus, maintain the leak locations has posed a big challenge for local government. The testing and application of the PSLDA could provide new insights into addressing this problem.

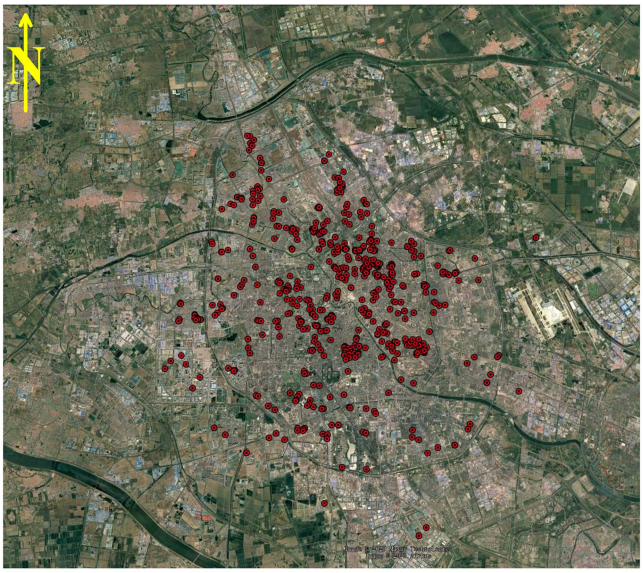
To enhance the robustness and transferability of PSLDA, we also selected the water/sewer leakage in the main urban area of Tianjin as our research area. Until 2022, the length of supply water pipeline has continuously increased to 22574.16 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-water-pipeline/cn-length-of-water-pipeline-city-tianjin), and the sewage pipeline reached 23 909.54 km (https://www.ceicdata.com.cn/zh-hans/china/length-of-drain-pipeline). Thus, the water pipeline distribution system of Tianjin also provides an opportunity to obtain in situ leak information. In summary, Beijing and Tianjin, China, were chosen as the training areas.

B. Training Dataset

1) In Situ Leak Data: Over the two training areas, the in situ leak locations were detected via the listening stick in the year of 2021. In the eastern part of Beijing, 668 leak locations were collected. Their spatial distribution is superimposed on the GoogleEarth image and shown in Fig. 3(a). Similarly, 556 leak locations were compiled in the main urban area of Tianjin. Their



(a)



(b)

Fig. 3. Spatial distribution of in situ leak points (red) superimposed on the GoogleEarth image. These points are measured in (a) Beijing and (b) Tianjin, China.

spatial distribution is superimposed on the GoogleEarth image and shown in Fig. 3(b). In total, there exist 1224 in situ leak data, which are potentially applicable for data training.

2) SAR Data: Because the pipelines are buried underground, a long wavelength (e.g., an L-band or a wavelength of about 24 cm) SAR system is preferred due to its deep penetration ability. Thus, L-band data should have priority over a short wavelength SAR system. Since pipeline leaking is typically not instantaneous, the multitemporal SAR observations should be considered if possible. From the perspective of closure phase, multitemporal SAR data are also required. However, long-wavelength radar data (e.g., the SAOCOM and ALOS-2/PALSAR-2) are generally charged, making it costly to obtain multitemporal SAR data. Fortunately, the C-band Sentinel-1 data are freely downloadable, and could provide sufficient SAR data.

TABLE I
SAR DATASET USED FOR PSLDA TRAINING

Satellite	Acquisition date (DD/MM/YYYY)		Polarization
	Beijing	Tianjin	
SAOCOM (L-band)	7 November 2021	23 September 2021	Quad
Sentinel-1 (C-band)	26 October 2021	8 September2021	Dual
	7 November 2021	20 September2021	
	19 November 2021	2 October2021	

Over the training area, both L-band SAOCOM data and C-band Sentinel-1 data in 2021 are used (see Table I). In Beijing, One SAOCOM SLC image on 7 November 2021, and three Sentinel-1 SLC images separately on October 26, November 7, and November 19 of 2021 were downloaded. In Tianjin, one SAOCOM SLC image on 23 September 2021, three Sentinel-1 SLC images on September 8, September 20, and October 2 of 2021 were acquired. Since the polarimetric information is used for moisture content retrieval, all SAOCOM data are in quad-polarization mode (HH+HV+VH+VV) and all Sentinel-1 data are in dual-polarization (VV+VH). The copolarized mode (VV) of Sentinel-1 was adopted in interferometry for calculating the closure phase. It is also worth noting that the precise orbit determination (POD) precise orbit ephemerides (POE) for each Sentinel-1 acquisition was downloaded to correct orbital errors. Moreover, the National Aeronautics and Space Administration (NASA) digital elevation model (DEM) with a spatial resolution of 30×30 m was utilized to remove topographic phases in the interferometric analysis.

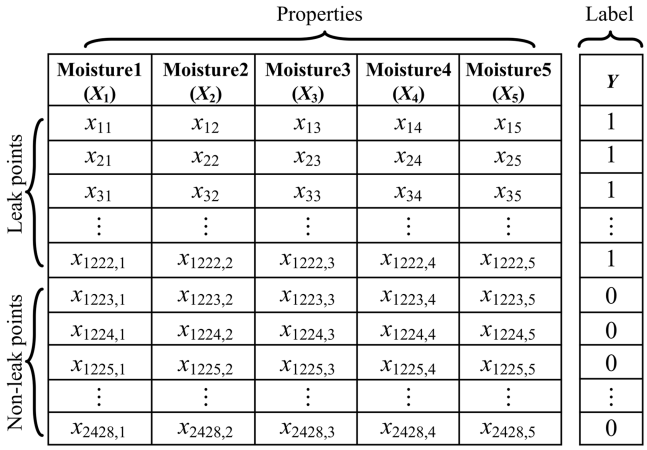
3) Input Data Preparation: Corresponding to each SAR SLC image, the moisture content image was derived with a resolution of 30×30 m. That is, four moisture content images on 7 November 2021 from SAOCOM, on 26 October 2021, 7 November 2021, and 19 November 2021 from Sentinel-1 were obtained for Beijing. Four moisture content images on 23 September 2021 from SAOCOM, on 8 September 2021, 20 September 2021, and 2 October 2021 from Sentinel-1 were obtained for Tianjin.

In addition, three C-band Sentinel-1 images were interferometrically processed. The procedures include coregistration, interferogram generation, topographic phase removal, phase filtering, and geocoding, finally generating filtered differential interferograms in the geographic coordinate system. For convenience, the result is referred to as differential interferogram. The differential interferogram has a spatial resolution of 30×30 m. On the basis of three differential interferograms, the absolute closure phase ranging between [0, π] was calculated, resulting in an absolute closure phase image.

Up to now, each known leak location in Fig. 3(a) and (b) corresponds to five images, i.e., four moisture content images and one closure phase image. Generally speaking, the four moisture content values and one closure phase value of each leak location could be collected pixel-by-pixel. However, the water is permeable. Thus, the water leakage influences not an isolated point but a continuous surface horizontally and vertically. As

a result, pixels around leak location may exhibit quasi-leak characteristics, i.e., high moisture content and closure phase deviated from zero. Under this condition, it is much more meaningful to determine an accredited spatial range that water/sewer leakage may potentially exist than an isolated location. Within the range, the pixel with the highest moisture content value or most deviated closure phase from zero is of high possibility to be a leak location. Thus, given a leak location in Fig. 3(a) and (b), we determine its moisture content values and closure phase value as follows. First of all, a 3×3 window was placed around the given leak location. The window size is generally chosen as an odd number. As will be depicted later, a circular range with a radius of 120 m is used during in situ verification. Thus, a window size of 3 was adopted here such that the detection range is just within the verification range. Within the window, the maximum moisture content value or the closure phase value deviated from zero most was assigned to the current leak point (at the window center). The operation was repeated 5 times until four corresponding moisture content values and one closure phase value were all determined. In addition, pixels with moisture contents and absolute closure phase obviously lower than the mean level of the entire image were randomly selected as nonleak points for training. In reference to the number of in situ leak samples, 668 nonleak locations were randomly selected in Beijing area and 556 nonleak locations for Tianjin area. Consequently, there are a total of 1224 leak candidates and 1224 nonleak candidates for PSLDA training.

Subsequently, all leak and nonleak points can be arranged row by row, while their four moisture contents and one closure phase arranged column by column. For example, the first row corresponds to leak point 1, the second row to leak point 2,. . ., until the 1224th row to leak point 1224. The 1225th row corresponds to nonleak point 1, and so on, until the 2448th row to nonleak point 1224. The first column corresponds to the moisture1 of SAR acquisition 1, the second column to moisture2 of SAR acquisition 2, the third column to moisture3 of SAR acquisition 3, the fourth column to moisture4 of SAR acquisition 4, and the fifth column to the absolute closure phase. It is worth mentioning that the arrangement order among rows or among columns exerts no influence as long as all five values for a certain point are aligned. To avoid the disturbance of unreliable training points, each column was standardized such that the mean value is zero, and the standard deviation is 1. Based on the standardized results, points with value beyond $[-3, 3]$ were considered as outliers and eliminated in the light of 3σ principle (σ is the standard deviation). Consequently, a total of 1222 leak points and 1206 nonleak points remained. Now, the input data *InputData* including moisture contents and closure phase, and labels are prepared and organized as Fig. 4.



	Properties					Label
	Moisture1 (X_1)	**Moisture2 (X_2)**	**Moisture3 (X_3)**	**Moisture4 (X_4)**	**Moisture5 (X_5)**	***Y***
Leak points	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	1
	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	1
	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	1
	⋮	⋮	⋮	⋮	⋮	⋮
	$x_{1222,1}$	$x_{1222,2}$	$x_{1222,3}$	$x_{1222,4}$	$x_{1222,5}$	1
Non-leak points	$x_{1223,1}$	$x_{1223,2}$	$x_{1223,3}$	$x_{1223,4}$	$x_{1223,5}$	0
	$x_{1224,1}$	$x_{1224,2}$	$x_{1224,3}$	$x_{1224,4}$	$x_{1224,5}$	0
	$x_{1225,1}$	$x_{1225,2}$	$x_{1225,3}$	$x_{1225,4}$	$x_{1225,5}$	0
	⋮	⋮	⋮	⋮	⋮	⋮
	$x_{2428,1}$	$x_{2428,2}$	$x_{2428,3}$	$x_{2428,4}$	$x_{2428,5}$	0

Fig. 4. Final arrangement of input data *InputData* for PSLDA training. The arrangement order among rows or among columns exerts no influence.

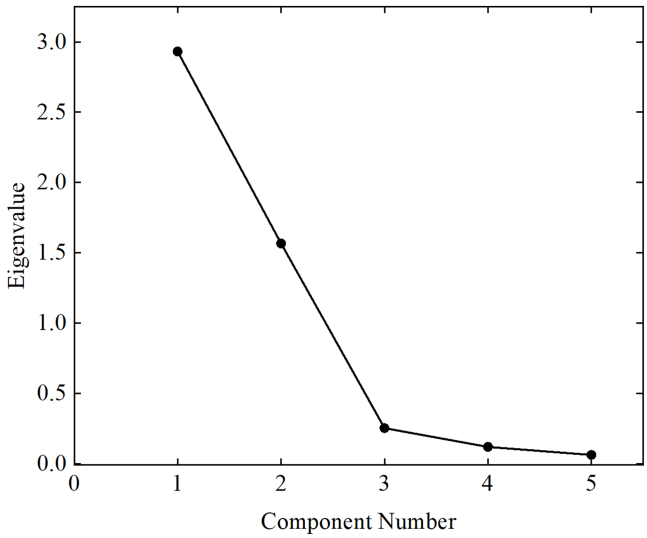


Fig. 5. Eigenvalues of each PC. The eigenvalues have been sorted in descending order.

C. Training Implementation

Upon inputting the training data *InputData* into the PCA module of PSLDA, five PCs were obtained. The five eigenvalues were sorted from the highest to the lowest (i.e., $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5$) and illustrated in Fig. 5. It is apparent that the first two eigenvalues (λ_1 and λ_2) exceed 1. PCs with eigenvalues surpassing 1 are typically considered significant as they account for a substantial portion of the data variance. These crucial PCs can be utilized for generating data visualizations, constructing predictive models, and conducting various data analysis tasks. Quantitatively, the contribution of each variance to the total variance were calculated in terms of the following equation:

$$\text{ratio}_k = \frac{\lambda_k}{\sum_{i=1}^{5} \lambda_i} \tag{11}$$

where λ_k is the eigenvalue of the kth PC. $\sum_{i=1}^{5} \lambda_i$ is the sum of all eigenvalues.

The results are presented in Table II. Notably, the first two PCs explain more than 90% of the variations in the dataset. In other words, PC1 and PC2 already contain main features or most useful information. Therefore, the first two PCs are retained

TABLE II
PROPORTION OF EXPLAINED VARIANCE FOR EACH PC

PC	Eigenvalue	Total variance (%)	Cumulative variance (%)
PC1	2.93	59.44	59.44
PC2	1.57	31.75	91.19
PC3	0.25	5.13	96.32
PC4	0.12	2.42	98.74
PC5	0.06	1.26	100

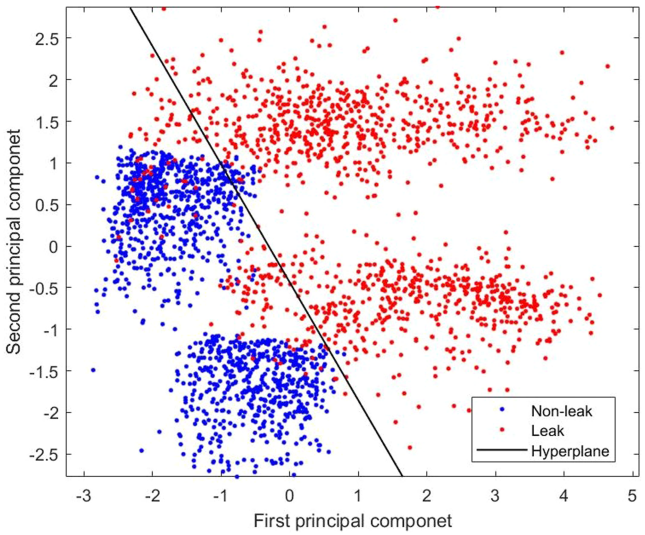


Fig. 6. Scatterplot of the nonleak and leak points in the training dataset. The SVM hyperplane line (PC2 $= -1.4151\times$PC1-0.4672) is black.

while others are excluded. Such operation reduces the subsequent computational costs, enhances the model interpretability and facilitates the visualization of high-dimensional data.

Following the PCA dimensionality reduction of the training dataset, the resulting two PCs were inputted into the SVM module of PSLDA, allowing for the generation of classification models for leak and nonleak points. Fig. 6 presents scatter plots depicting the distribution of leak and nonleak points. Notably, two types of points are distinguishable. Leak points (red) primarily occupy the right and upper positions, while nonleak points (blue) are situated in the left and lower positions. The SVM hyperplane is represented by the black line in the figure.

To evaluate the detection accuracy, we consider the misclassification rate (MR), or

$$\mathrm{MR} = \frac{\mathrm{FA} + \mathrm{MD}}{2} \tag{12}$$

where FA is the false alarm probability. It is the ratio of the number of nonleak locations identified as leak positions divided by the total leak locations. MD denotes the missed detection probability or the ratio of water leakage locations delineated as nonleak positions divided by the total nonleak locations. The smaller the MR, the better the detection. Statistically, there are 38 false alarm locations, and 154 missed detection locations. Considering that there are 1222 leak training points and 1206 nonleak training points, the FA is calculated to be 3.11% (38/1222) and MR 12.78 % (154/1206). As a result, the MR of PSLDA is 7.94%, signifying that the proposed PSLDA is adept at distinguishing between water supply pipe leak and nonleak points with a relatively low error rate.

TABLE III
SAR DATASETS USED FOR SEWER/WATER LEAK DETECTION IN BEIJING

Satellite	Acquisition date (DD/MM/YYYY)	Polarization
SAOCOM (L-band)	3 May 2023	Quad
Sentinel-1 (C-band)	7 April 2023	Dual
	19 April 2023	
	1 May 2023	

IV. PERFORMANCE AND ANALYSIS

The validity and practical applicability of the PSLDA are evaluated through leak detection case and nonleak detection case in real-world scenarios. The validation procedure incorporates a blend of satellite image-based leak detection method (i.e., the PSLDA) and acoustic detection method.

A. Leak Detection in the Fifth Ring Road of Beijing

We designated the region within the fifth ring road of Beijing as the detection area. The original SAR SLC data consist of one quad-polarized SAOCOM image on 3 May 2023, and three dual-polarized Sentinel-1 images on 7 April 2023, 19 April 2023, and 1 May 2023, respectively (see Table III). Using these SAR data, five features—comprising four soil moisture content and one absolute closure phase—were computed and entered into the PSLDA for predicting the location of leak points. These features align with the same type used in the training process.

Through PSLDA, potential leak points were obtained, where those situated within the proximity of the water supply pipe range were selected as points of interest (POIs). As reported by the Beijing Water Supply Group Company, Ltd., a radius of (125$\pm$25) m is universally applicable in domestic cities under the premise of both number and efficiency of dark leak detection [35]. Therefore, a circular area with a radius of 120 m was drawn around each POI, and the circular area was designated as the region of interest (ROI). In the fifth ring of Beijing, 150 ROIs were initially selected. However, during the field investigation, some areas within the ROIs were found to be inaccessible (e.g., military regions and schools). After excluding these inaccessible areas, 86 ROIs remain. They are overlapped on the GoogleEarth image and depicted as red circles in Fig. 7. The ROIs are dispersed overall, but clustered locally. The phenomena should be reasonable because it is of low possibility that the water pipelines burst evenly around the city. Contrastively, it is very likely that the spatially proximate water pipelines are featured with analogous installation timelines, installation materials, and installation environments, etc., thus experiencing malfunctions at nearly the same time.

It is accepted that the detected leak would be valid once there indeed exist water leak locations within the ROI. Therefore, professional leak detectors were subsequently dispatched to each

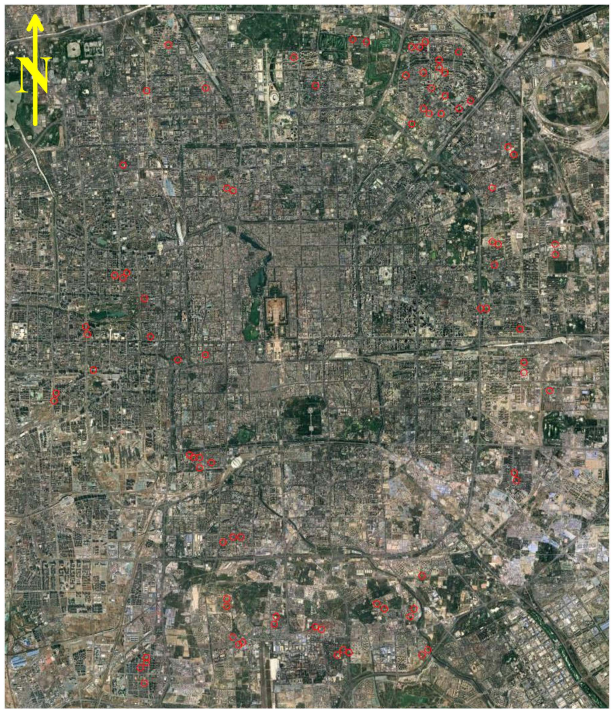


Fig. 7. Leak detection areas in the water pipeline system within the fifth ring road of Beijing. The red circles signify the designated detection area.

Fig. 8. Real scenarios of the leaking water supply pipes at (a) site 1, (b) site 2, (c) site 3, and (d) site 4 in Beijing.

ROI area to execute acoustic leak detection. The listening sticks were utilized to discern any unusual sounds indicative of water leaks on the water supply pipes, aiding in the determination of whether the pipes were indeed leaking. For example, Fig. 8 illustrates four scenarios of water pipe leaks detected by the leak detectors, where standing water is observed. Such phenomenon coincides perfectly with the experimental thoughts in this article (i.e., moisture content and closure phase), thus indirectly confirming the feasibility of PSLDA. The field investigations over 86 ROIs were ultimately accomplished by 17 June 2023. The results were compiled and overlapped on the GoogleEarth image in Fig. 9, where actual leak locations are marked as red pins. Qualitatively, almost all ROIs contain leak location. Quantitatively, only 7 ROIs remain where no leak points exist. This signifies a remarkable accuracy rate of 91.86% for the PSLDA in discerning water supply pipe leaks across Beijing. The successful identification of leak points within the majority of the designated ROIs underscores the algorithm's robustness and practical utility in real-world scenarios.

B. Nonleak Detection in Huizhou, Guangdong Province

The training data of PSLDA include leak points and nonleak points. In other words, the algorithm learns the leak feature and nonleak feature simultaneously. Hence, the PSLDA should be able to detect nonleak locations. Under the circumstances, the Huizhou city in Guangdong Province, China, is showcased. The original SAR SLC data consist of one quad-polarized SAOCOM image on 9 August 2023, and three dual-polarized Sentinel-1 images 10 July 2023, 22 July 2023, and 3 August 2023, respectively (see Table IV). Using these data, five features—comprising four soil moisture content and one closure phase—were computed and entered into the PSLDA for predicting the location of nonleak points.

TABLE IV
SAR DATASETS USED FOR NONLEAK DETECTION IN THE HUIZHOU CITY

Satellite	Acquisition date (DD/MM/YYYY)	Polarization
SAOCOM (L-band)	9 August 2023	Quad
Sentinel-1 (C-band)	10 July 2023	Dual
	22 July 2023	
	3 August 2023	

With the proposed algorithm, nonleak points were detected. Those points situated within the proximity of the water supply pipe range were selected as POIs. Around POIs, a total of 42 nonleak ROIs are outlined and depicted as red circles in Fig. 10. Similarly, professional leak detectors were dispatched to each ROI area to implement acoustic leak detection and check whether the pipelines within the ROIs are indeed not leaking. Consequently, only 5 ROIs contained leaks, and no leaks were found in remaining 37 ROIs. That is to say, the proposed PSLDA achieves an accuracy of 88.10% in detecting nonleak locations.

In summary, the PSLDA proves to be effective and robust in detecting either leaks or nonleaks in real-world scenarios, exhibiting potentials in monitoring sewer/water leakage and providing instruction for timely maintenance.

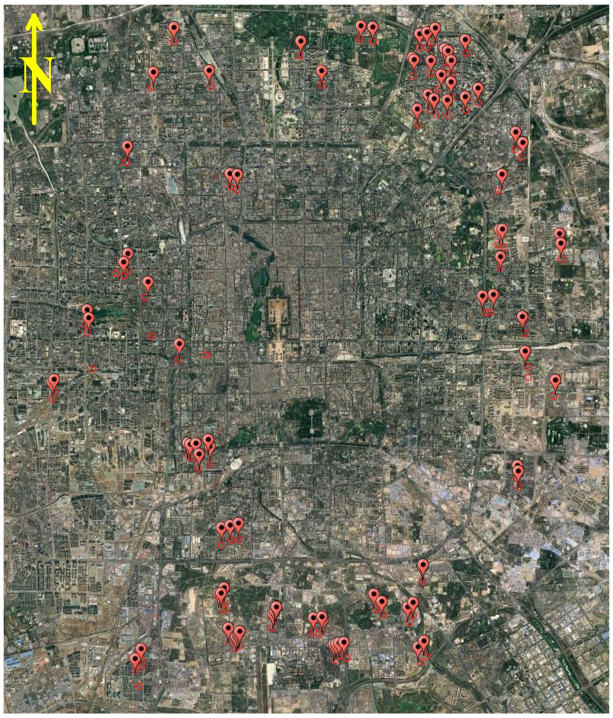


Fig. 9. Distribution of identified leak points in the water pipeline within the fifth ring road of Beijing. The red circles represent the designated detection areas, while the red pins indicate the detected leak points.

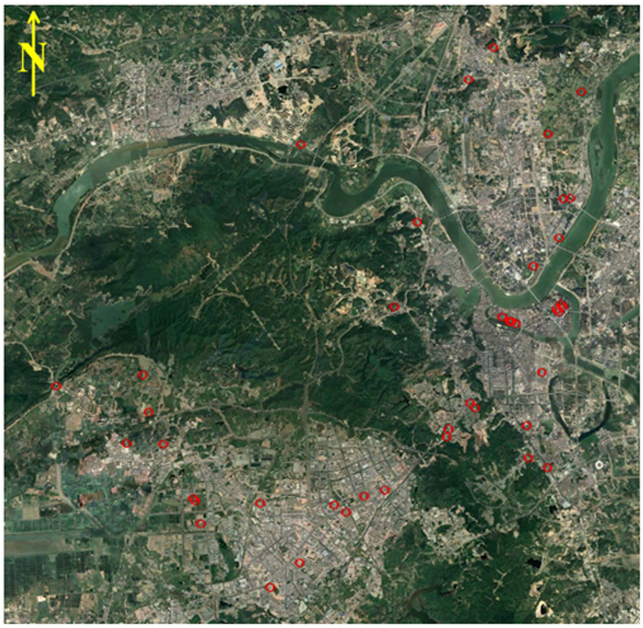


Fig. 10. Nonleak ROIs (red circle) in the water pipeline system within Huizhou city, Guangdong Province, China.

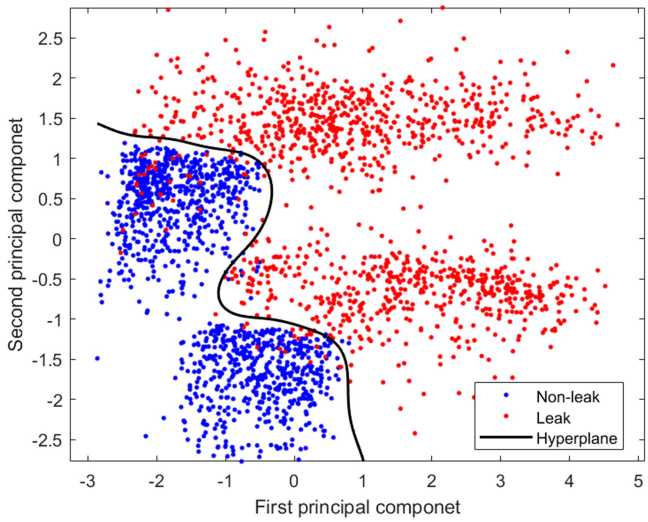


Fig. 11. Scatterplot of the nonleak and leak points in the training dataset with Gaussian kernel used in the SVM. The SVM hyperplane line is black.

V. Discussion

Theoretically, the learning ability of nonlinear function is stronger than that of linear function. Therefore, the performance of nonlinear kernel-based SVM is explored in detecting the leaks/nonleaks. The Gaussian kernel was adopted here. With identical training area and training dataset in Section III, a total of 2428 leak and nonleak samples were first processed with the feature extraction module of PSLDA. Subsequently, the reserved PCs were fed into the Gaussian kernel-based SVM, generating a classification model in Fig. 11. Notably, two types of points are distinguishable. Leak points (red) primarily occupy the right and upper positions, while nonleak points (blue) are situated in the left and lower positions. The nonlinear SVM hyperplane is represented by the black line. Compared with the linear SVM model in Fig. 6, more leak points (red) in Fig. 11 are correctly categorized. Quantitatively, there are 13 false alarm locations, and 56 missed detection locations. Considering that there are 1222 leak training points and 1206 nonleak training points, the FA is calculated to be 1.06% (13/1222) and MR 4.64% (56/1206). As a result, the MR of PSLDA is 2.85%, which is smaller than that of linear SVM. Therefore, it is implied that PSLDA incorporating nonlinear kernel-based SVM is capable of detecting water/sewer leakage more accurately. As indicated in Sections IV-A and IV-B, the accuracy verification of PSLDA through actual cases is complicated. One needs to first detect potential leak/nonleak points, based on which the listening sticks will be utilized to affirm the actual leaks/nonleaks. Therefore, it is low of possibility that the detection accuracy of PSLDA incorporating nonlinear kernel-based SVM could be validated through real-world scenarios. Under this condition, the well-trained SVM model in Fig. 11 was not furtherly applied to predict leaks in Beijing or nonleaks in Huizhou City, China.

VI. Conclusion

In order to detect sewer/water leakage effectively and efficiently, we presented an interpretable PSLDA. The network takes the moisture contents and InSAR closure phase as inputs, then conduct feature extraction and dimensionality reduction with the mathematically derived PCA instead of the convolutional and pooling layer in DL, finally implement binary

classification with the mathematically interpretable SVM instead of the loss-function layer in DL, identifying the leaks or nonleaks. The PSLDA was trained over 1222 in situ leak points and 1206 nonleak points. The well-trained model was applied to detect the leak/nonleak locations for two real-world scenarios, and achieved an accuracy of as high as 91.86%. The experimental results verify that the PSLDA is an effective and robust sewer/water leak detection algorithm.

Acknowledgment

Sentinel-1 data were downloaded from the http://vertex.daac.asf.alaska.edu/ of Alaska Satellite Facility, USA. The DEM used in this study is the NASADEM and was downloaded from the EARTHDATA center https://www.earthdata.nasa.gov/. The PODPOE data used in this study were downloaded from the website https://step.esa.int/auxdata/orbits/.

References

[1] D. Ma et al., "A multi-defect detection system for sewer pipelines based on styleGAN-SDM and fusion CNN," *Construction Build. Mater.*, vol. 312, Dec. 2021, Art. no. 125385, doi: 10.1016/j.conbuildmat.2021.125385.

[2] C. David and C. Richard, *Nature-Based Solutions for Water*. Place de Fontenoy, France: United Nations Educational, Scientific and Cultural Organization, 2018.

[3] S. Tariq, B. Bakhtawar, and T. Zayed, "Data-driven application of MEMS-based accelerometers for leak detection in water distribution networks," *Sci. Total Environ.*, vol. 809, Feb. 2022, Art. no. 151110, doi: 10.1016/j.scitotenv.2021.151110.

[4] A. Boretti and L. Rosa, "Reassessing the projections of the world water development report," *NPJ Clean Water*, vol. 2, Jul. 2019, Art. no. 15, doi: 10.1038/s41545-019-0039-9.

[5] M. J. Anbari, M. Tabesh, and A. Roozbahani, "Risk assessment model to prioritize sewer pipes inspection in wastewater collection networks," *J. Environ. Manage.*, vol. 190, pp. 91–101, Apr. 2017, doi: 10.1016/j.jenvman.2016.12.052.

[6] E. L. Kuniansky, D. J. Weary, and J. E. Kaufmann, "The current status of mapping karst areas and availability of public sinkhole-risk resources in karst terrains of the United States," *Hydrogeology J.*, vol. 24, no. 3, pp. 613–624, May 2016, doi: 10.1007/s10040-015-1333-3.

[7] M. Gamal, Q. Di, J. Zhang, C. Fu, S. Ebrahim, and A. A. El-Raouf, "Utilizing ground-penetrating radar for water leak detection and pipe material characterization in environmental studies: A case study," *Remote Sens.*, vol. 15, no. 20, Oct. 2023, Art. no. 4924, doi: 10.3390/rs15204924.

[8] P. M. Bach and J. K. Kodikara, "Reliability of infrared thermography in detecting leaks in buried water reticulation pipes," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4210–4224, Sep. 2017, doi: 10.1109/JSTARS.2017.2708817.

[9] Y. Tong et al., "Water leakage detection technology and practice of water supply pipe network based on long band radar satellite," *Water Wastewater Eng.*, vol. 49, no. 2, pp. 131–137, 2023, doi: 10.13789/j.cnki.wwe1964.2022.02.18.0002.

[10] D. Massonnet and K. L. Feigl, "Radar interferometry and its application to changes in the Earth's surface," *Rev. Geophys.*, vol. 36, pp. 441–500, Nov. 1998, doi: 10.1029/97RG03139.

[11] Z. Lu and D. Dzurisin, *InSAR Imaging of Aleutian Volcanoes: Monitoring a Volcanic Arc From Space*. Berlin, Germany: Springer-Verlag, 2014.

[12] B. D. Stanford, S. Pochiraju, T. Yokoyama, P. Maari, L. Grijalva, and Z. Bukhari, "Evaluating satellite and in situ monitoring technologies for leak detection and response," *AWWA Water Sci.*, vol. 4, no. 3, May 2022, Art. no. e1288.

[13] X. Le, H. Yu, and Y. Wang, "An interpretable neural network algorithm for leaking detection in the urban water and sewer pipeline network Tianjin, China," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Pasadena, CA, USA, 2023, pp. 2053–2056.

[14] L. He, Q. Qin, R. Panciera, M. Tanase, J. P. Walker, and Y. Hong, "An extension of the alpha approximation method for soil moisture estimation using time-series SAR data over bare soil surfaces," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1328–1332, Aug. 2017, doi: 10.1109/LGRS.2017.2711006.

[15] A. K. Holtgrave, M. Förster, F. Greifeneder, C. Notarnicola, and B. Kleinschmit, "Estimation of soil moisture in vegetation-covered floodplains with Sentinel-1 SAR data using support vector regression," *PFG-J. Photogrammetry Remote Sens. Geoinf. Sci.*, vol. 86, no. 2, pp. 85–101, Apr. 2018, doi: 10.1007/s41064-018-0045-4.

[16] J. Guo, J. Liu, J. Ning, and W. Han, "Construction and validation of soil moisture retrieval model in farmland based on sentinel multi-source data," *Trans. Chin. Soc. Agricultural Eng.*, vol. 35, no. 14, pp. 71–78, Jul. 2019, doi: 10.11975/j.issn.1002-6819.2019.14.009.

[17] F. De Zan, M. Zonno, and P. López-Dekker, "Phase inconsistencies and multiple scattering in SAR interferometry," *IEEE Trans. Geosci. Remote Sens..*, vol. 53, no. 12, pp. 6608–6616, Dec. 2015, doi: 10.1109/TGRS.2015.2444431.

[18] C. Anitescu, E. Atroshchenko, N. Alajlan, and T. Rabczuk, "Artificial neural network methods for the solution of second order boundary value problems," *CMC-Comput. Mater. Continua*, vol. 59, no. 1, pp. 345–359, doi: 10.32604/cmc.2019.06641.

[19] D. S. Li, A. R. Cong, and S. Guo, "Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification," *Autom. Construction*, vol. 101, pp. 199–208, May 2019, doi: 10.1016/j.autcon.2019.01.017.

[20] S. K. Sinha and P. W. Fieguth, "Neuro-fuzzy network for the classification of buried pipe defects," *Autom. Construction*, vol. 15, no. 1, pp. 73–83, Jan. 2006, doi: 10.1016/j.autcon.2005.02.005.

[21] J. Kang, Y.-J. Park, J. Lee, S.-H. Wang, and D.-S. Eom, "Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution system," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4279–4289, May 2018, doi: 10.1109/TIE.2017.2764861.

[22] X. Hu, H. Zhang, D. Ma, R. Wang, and J. Zheng, "Minor class-based status detection for pipeline network using enhanced generative adversarial networks," *Neurocomputing*, vol. 424, pp. 71–83, Jan. 2021, doi: 10.1016/j.neucom.2020.11.009.

[23] L. I. Smith, *A Tutorial on Principal Component Analysis*, Cornell University, Ithaca, NY, USA, 2002.

[24] L. Li, S. Liu, Y. Peng, and Z. Sun, "Overview of principal component analysis algorithm," *Optik*, vol. 127, no. 9, pp. 3935–3944, 2016, doi: 10.1016/j.ijleo.2016.01.033.

[25] R. Gholami and N. Fakhari, "Support vector machine: Principles, parameters, and applications," in *Handbook of Neural Computation*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 515–535.

[26] G. T. Kaya, "A hybrid model for classification of remote sensing images with linear SVM and support vector selection and adaptation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1988–1997, Aug. 2013, doi: 10.1109/JSTARS.2012.2233463.

[27] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[28] Y. Yan, H. Yu, and Y. Wang, "Alarming a tailings dam failure with a joint analysis of InSAR-derived surface deformation and SAR-derived moisture content," *Remote Sens. Environ.*, vol. 300, Nov. 2023, Art. no. 113910, doi: 10.1016/j.rse.2023.113910.

[29] S. Zwieback, S. Hensley, and I. Hajnsek, "A polarimetric first-order model of soil moisture effects on the DInSAR coherence," *Remote Sens.*, vol. 7, pp. 7571–7596, 2015, doi: 10.3390/rs70607571.

[30] S. Zwieback et al., "A statistical test of phase closure to detect influences on DInSAR deformation estimates besides displacements and decorrelation noise: Two case studies in high-latitude regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5588–5601, Sep. 2016, doi: 10.1109/TGRS.2016.2569435.

[31] H. Yu, Y. Lan, Z. Yuan, J. Xu, and H. Lee, "Phase unwrapping in InSAR: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 40–58, Mar. 2019, doi: 10.1109/MGRS.2018.2873644.

[32] B. Zhu, Y. Wang, and H. Yu, "An algorithm measuring urban building heights by combining the PS-InSAR technique and two-stage programming approach framework," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7624–7635, Sep. 2023, doi: 10.1109/JSTARS.2023.3305890.

[33] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS-J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011, doi: 10.1016/j.isprsjprs.2010.11.001.

[34] W. Li et al., "Development of systems for detection, early warning, and control of pipeline leakage in drinking water distribution: A case study," *J. Environ. Sci.*, vol. 23, no. 11, pp. 1816–1822, 2011, doi: 10.1016/S1001-0742(10)60577-3.

[35] Y. Tong et al., "Study on the applicability of satellite leak detection of water supply networks," *Water Wastewater Eng.*, vol. 49, no. 6, pp. 128–133, 2023, doi: 10.13789/j.cnki.wwe1964.2023.05.04.0007.

Yan Yan (Member, IEEE) received the B.S. degree in environmental engineering and the Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016 and 2024, respectively.

She is currently a Postdoctoral Researcher Fellow with the School of Resources and Environment, UESTC. Her research interests include the fields of phase unwrapping and synthetic aperture radar interferometry application.

Xujie Le (Member, IEEE) received the B.S. and M.S. degrees in electronic and information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2021 and 2024, respectively.

Taoli Yang (Senior Member, IEEE) received the Ph.D. degree in signal processing from the National Key Lab of Radar Signal Processing, Xidian University, Xi'an, China, in 2014.

From 2015 to 2016, she was a Postdoctoral Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. She is currently an Associate Professor with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include SAR/ISAR imaging, interferometric SAR, GRACE, and ground moving target indication

Hanwen Yu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Xidian University, Xian, China, in 2007 and 2012, respectively.

He was a Postdoctoral Research Fellow with the Department of Civil and Environmental Engineering, National Center for Airborne Laser Mapping, University of Houston, Houston, TX, USA. He is currently a Full Professor with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, China, and an adjunct Full Professor with the Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an, China, and the Department of Engineering, University of Naples, Naples, Italy. He has authored more than 70 scientific articles and given scientific presentation about "Advanced Techniques in InSAR Phase Unwrapping" invited by the IEEE Geoscience and Remote Sensing Society (GRSS) Webinar in 2021. His research interests include InSAR, and this work has led to new insights into the worldwide deformation monitoring and topographic mapping.

Dr. Yu reviewed more than 300 manuscripts for more than 20 different journals. He has been involved in IEEE (in general) and IEEE GRSS in particular. He was elected as a Best Reviewer of IEEE TRANSACTIONS GEOSCIENCE AND REMOTE SENSING in 2019. He was a recipient of several awards and honors from IEEE GRSS, including the 2022 Transactions Prize Paper Award, the Technical Program Committee Member and the session chair of 2022 IGARSS, and the Principal investigator of two IEEE GRSS 50