

Article

# Gas Pipeline Leakage Detection Method Based on IUPLCD and GS-TBSVM

Haiou Shan and Yongqiang Zhu \* 

School of Information and Control Engineering, Liaoning Petrochemical University, Fushun 113001, China

\* Correspondence: lnpuzyq@163.com

**Abstract:** To improve the identification accuracy of gas pipeline leakage and reduce the false alarm rate, a pipeline leakage detection method based on improved uniform-phase local characteristic-scale decomposition (IUPLCD) and grid search algorithm-optimized twin-bounded support vector machine (GS-TBSVM) was proposed. First, the signal was decomposed into several intrinsic scale components (ISC) by the UPLCD algorithm. Then, the signal reconstruction process of UPLCD was optimized and improved according to the energy and standard deviation of the amplitude of each ISC, the ISC components dominated by the signal were selected for signal reconstruction, and the denoised signal was obtained. Finally, the TBSVM was optimized using a grid search algorithm, and a GS-TBSVM model for pipeline leakage identification was constructed. The input of the GS-TBSVM model was the data processed by the IUPLCD algorithm, and the output was the real-time working conditions of the gas pipeline. The experimental results show that IUPLCD can effectively filter the noise in the signal and GS-TBSVM can accurately judge the working conditions of the gas pipeline, with a maximum identification accuracy of 98.4%.

**Keywords:** leak detection; grid search method (GS); twin-bounded support vector machine (TBSVM); gas pipeline

## 1. Introduction

Oil/gas pipelines are the arteries of industrial development, shouldering the task of transporting oil, gas, and other energy to refineries [1]. Oil and gas pipelines are complicated along the line, and most of them are laid in wild or remote places, and the laying environments are rather harsh. Therefore, the pipelines are prone to leakage. Natural gas is flammable and explosive [2,3], so once it leaks, it may cause accidents such as explosions. Therefore, leak detection of pipelines is important in pipeline operation management [4,5].

At present, there are many methods for pipeline leakage detection. According to the current application situation, it can be divided into non-acoustic detection methods and acoustic detection methods [6,7]. The acoustic method has the advantages of high detection accuracy and low false alarm rate [8]. However, the acoustic wave is easily disturbed by the environment during its propagation in the pipeline, which makes the acoustic signal contain considerable noise [9,10], thus reducing the accuracy of leakage identification. Therefore, it is particularly important to denoise acoustic signals. Meng proposed a denoising method combining ensemble empirical mode decomposition (EEMD) with cross-spectrum analysis, which improved the reconstruction process of the EEMD algorithm, so that the reconstructed signal contained more leakage information and effectively reduced the noise content [11]. Shi et al. improved the local mean decomposition (LMD) denoising method. First, the effective component reconstruction signal is obtained by the peak value obtained by signal correlation analysis. Then, the wavelet analysis is used to further filter out the noise. This method can filter out most of the noise in the signal, but there is still residual noise in the denoised signal [12]. Li et al. adopted a combined noise



**Citation:** Shan, H.; Zhu, Y. Gas Pipeline Leakage Detection Method Based on IUPLCD and GS-TBSVM. *Processes* **2023**, *11*, 278. <https://doi.org/10.3390/pr11010278>

Academic Editor: Wei Sun

Received: 25 December 2022

Revised: 12 January 2023

Accepted: 13 January 2023

Published: 15 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

reduction method based on cross-spectral analysis and independent component analysis (ICA) to optimize complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). Intrinsic mode function (IMF) components were filtered by the obtained characteristic frequency band and effective amplitude ratio [13].

The twin support vector machine (TWSVM) is one of the most classic and popular classification models, but it has some problems, such as slow training speed and low classification accuracy, when dealing with multiclassification problems [14,15]. TBSVM is improved on the basis of TWSVM. Compared with TWSVM, TBSVM has a faster training speed and higher accuracy. Lang input the processed signal into a least square twin support vector machine (LSTSVM) to identify the pipeline leakage, and the identification accuracy of the model was obviously improved, which could identify the pipeline leakage more accurately [16]. Kang used a convolution neural network and support vector machine (SVM) to identify pipeline leakage, which could accurately judge pipeline leakage. However, the calculation of this method is complicated and the running time is long [17].

To improve the accuracy of pipeline leakage identification, this paper proposes a gas pipeline leakage detection method based on the IUPLCD denoising algorithm and GS-TBSVM model. Acoustic waves often contain noise, which affects their recognition accuracy. In this paper, the IUPLCD denoising algorithm is used to denoise acoustic signals. The noise content in the denoised signal is extremely low, and the original information is retained to a great extent. In addition, according to the grid search algorithm and TBSVM, the GS-TBSVM model is established to identify the working conditions of gas pipelines. By inputting the data after noise reduction into the model, the leakage of the gas pipeline can be accurately identified, and the false alarm of the system can be effectively reduced.

The rest of the paper is arranged as follows: In Section 2, a signal denoising algorithm based on IUPLCD is proposed. In Section 3, the GS-TBSVM model is constructed to identify the leakage of gas pipelines. In Section 4, the experimental analysis is carried out. Finally, the conclusion of this paper is provided in Section 5.

## 2. Improved Uniform-Phase Local Characteristic-Scale Decomposition

### 2.1. Uniform-Phase Local Characteristic-Scale Decomposition

Compared with empirical mode decomposition (EMD), local characteristic-scale decomposition (LCD) has better performance, and it can effectively reduce the envelope fitting error and endpoint effect [18]. However, LCD can not completely eliminate modal aliasing. UPLCD adds a narrow-band cosine signal with a uniform phase change to the signal to be decomposed, which uniformizes the distribution of each extreme point, thus achieving the effect of eliminating modal aliasing [19]. UPLCD is decomposed, as follows:

- (1) calculate the cycle period  $T_c$  and the times  $n_{isc}$  of the decomposition algorithm.

$$n_{isc} = \log_2(l) \quad (1)$$

$$T_c = 2^m, m = 1 : n_{isc} \quad (2)$$

where  $l$  is the length of the data, and the frequency  $f$  can be obtained according to the period,  $f = 1/T_c$ . According to the actual situation of the signal, the value of amplitude  $\varepsilon$  and the number of phases  $n_p$  are set.

- (2) Let  $x(t) = r_0(t)$ ,  $w(t; \varepsilon_c; f_c; \theta_k)$  be a narrow-wave cosine signal and there are:

$$w(t; \varepsilon_c; f_c; \theta_k) = \varepsilon_c \cdot \cos(2\pi \cdot f_c \cdot t + \theta_k) \quad (3)$$

$$\varepsilon_c = \varepsilon \cdot r_{m-1}(t) \quad (4)$$

$$\theta_k = 2\pi(k-1)/n_p \quad (5)$$

where,  $r_{m-1}(t)$  is the standard deviation and  $\theta_k$  is the phase of the signal.

- (3) The  $i$ -th ISC component obtained by LCD decomposition is represented by  $L_i(\cdot)$ , and the first component obtained by LCD decomposition is:

$$c_{m,k}(t) = L_1(x(t) + w(t; \varepsilon_c; f_c; \theta_k)) \quad (6)$$

$$k = 1, 2, \dots, n_p; m = 1 : n_{isc}$$

$$ISC_1 = \left( \sum_{k=1}^{n_p} c_{m,k}(t) - w(t; \varepsilon_c; f_c; \theta_k) \right) / n_p \quad (7)$$

where,  $c_{m,k}(t)$  is the signal to be decomposed after adding narrow-wave cosine signal, and  $ISC_1$  is the first ISC component.

- (4) Subtract the obtained  $ISC_1$  component from the source signal  $x(t)$ ; use the remaining signal as the new source signal  $r_i(t)$ ; and repeat the above steps to obtain all ISC components until all ISC components are decomposed; stop the cycle and  $r_{n_{isc}}(t)$  is the residual term.

## 2.2. Improved Uniform-Phase Local Characteristic-Scale Decomposition

Although the performance of UPLCD is much better than that of LCD, there is no change in signal reconstruction rules. Similar to LCD, UPLCD does not have a signal reconstruction principle, but all ISC components are used for signal reconstruction, which will lead to some ISC components with more noise being used for reconstruction, resulting in some noise in the reconstructed signal. In this paper, the energy of ISC components and the standard deviation of amplitude after Fourier spectrum analysis are used as the selection criteria of ISC, and more components containing useful information are screened out from all ISC components.

The ISC component contains both useful information and useless information (such as noise signals). The ISC component with a large proportion of information is called the noise-dominant component, and the ISC component with a large proportion of useful signals is called the signal-dominant component. In the energy curve of the ISC component, the first extreme point is critical for distinguishing whether the ISC component is dominated by noise or is a useful signal [20]. That is, all components before (including) the critical point of the energy curve are noise-dominated components, and those after the critical point are signal-dominated components. The energy of the ISC component can be obtained by the following equation:

$$e_i = \sum_{j=1}^M ISC_j^2(t) \quad (8)$$

$$i = 1, 2, \dots, n + 1$$

where,  $e_i$  is the energy of the  $i$ -th ISC component, and  $M$  is the length of the sequence.

Standard deviation (STD) is the arithmetic square root of variance, which can reflect the dispersion degree of a data set, and compared with variance, standard deviation can intuitively reflect the deviation degree of the data set [21]. Fourier transform is carried out on several ISC components obtained by signal UPLCD decomposition. Among the obtained spectrograms, the spectrogram dominated by useless signals is more complicated. At this point, the standard deviation can be used to measure the dispersion of the data set. The standard deviation of data with more useful information is smaller, and that of a data set with more useless information is significantly higher than that of a data set with more useful information. Therefore, the standard deviation can be used to screen out ISC components with more leakage information. The calculation equation of standard deviation is as follows:

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}} \quad (9)$$

where  $X_i$  is the sample value of the data set,  $\bar{X}$  is the average of all sample values, and  $N$  is the number of sample values.

Because UPLCD doesn't give a reasonable screening principle for ISC components, this paper makes a screening criterion to optimize and improve it, and screens out the effective ISC components. The steps of IUPEMD are as follows:

- (1) First, the original signal is decomposed by UPLCD algorithm to obtain several ISC components.
- (2) The energy of each ISC component is calculated, the energy curve of each component is obtained, and the first extreme point of the energy curve is determined.
- (3) Fourier transform is carried out on each ISC component, and then the amplitude of each ISC component is normalized, and then the standard deviation is obtained using Equation (9).
- (4) According to the energy value and the standard deviation of the amplitude of the ISC components, the ISC components dominated by the signal are screened out, and the ISC components satisfying the above two conditions are used for signal reconstruction. The reconstructed signal is the denoised signal. The flow chart is shown in Figure 1.

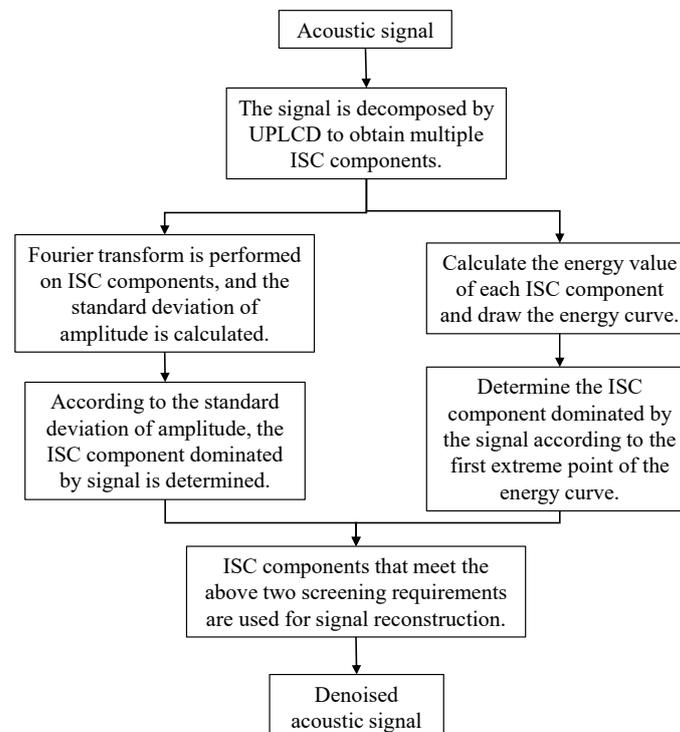


Figure 1. Denoising process of the signal.

### 3. Grid Search Algorithm-Optimized Twin-Bounded Support Vector Machine

#### 3.1. Twin-Bounded Support Vector Machine

TBSVM is an improvement of TWSVM, which uses successive over-relaxation technology to increase the training speed. In addition, TBSVM adopts the principle of structural risk minimization, which makes it superior to TWSVM in classification accuracy and calculation time. Similar to TWSVM, TBSVM finds two non-parallel decision hyperplanes to classify and judge the input data. The optimization problems of the linear TBSVM are as follows [22]:

$$\begin{aligned}
 \min_{w_1, b_1, \zeta, \zeta^*} \quad & \frac{1}{2}c_3(\|w_1\|^2 + b_1^2) + \frac{1}{2}\zeta^{*T}\zeta^* + c_1e_2^T\zeta \\
 \text{s.t.} \quad & Aw_1 + e_1b_1 = \zeta^* \\
 & -(Bw_1 + e_2b_1) + \zeta > e_2, \zeta \geq 0
 \end{aligned} \tag{10}$$

$$\begin{aligned} \min_{w_2, b_1, \eta, \eta^*} \quad & \frac{1}{2}c_4(\|w_2\|^2 + b_2^2) + \frac{1}{2}\eta^{*T}\eta^* + c_2e_1^T\eta \\ \text{s.t.} \quad & Bw_2 + e_2b_2 = \eta^* \\ & (Aw_2 + e_1b_2) + \eta > e_1, \eta \geq 0 \end{aligned} \quad (11)$$

where  $c_1, c_2, c_3, c_4$  is the penalty parameter.  $e_1$  and  $e_2$  are column vectors with two constituent elements of 1,  $\xi$  and  $\eta$  are relaxation variables that control the tolerance-to-noise ratio.

It can be obtained from Equations (10) and (11) that the constraint of one objective function is determined by the mode of the other objective function, and the two hyperplanes are mutually restricted. In the linear state, the Lagrange dual formula of TBSVM is:

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) = & \frac{1}{2}c_3(\|w_1\|^2 + b_1^2) + \frac{1}{2}\|Aw_1 + e_1b_1\|^2 \\ & + c_1e_2^T\xi + \alpha^T(Bw_1 + e_2b_1 - \xi + e_2) - \beta^T\xi \end{aligned} \quad (12)$$

From the above, it can be concluded that under the linear condition, in order to have the optimal solution to the linear programming problem, the dual problem can be obtained from the KKT condition, as follows:

$$\begin{aligned} \min_{\alpha} \quad & e_2^T\alpha - \frac{1}{2}\alpha^TG(H^TH + c_3I)^{-1}G^T\alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 \end{aligned} \quad (13)$$

$$\begin{aligned} \min_{\gamma} \quad & e_1^T\gamma - \frac{1}{2}\gamma^TH(G^TG + c_4I)^{-1}H^T\gamma \\ \text{s.t.} \quad & 0 \leq \gamma \leq c_2 \end{aligned} \quad (14)$$

When a new sample input is to be determined, the linear TBSVM can determine the label to which the sample belongs, according to the following formula:

$$\text{Class } i = \arg \min_{k=1,2} \frac{|(w_k^T \cdot x) + b_k|}{\|w_k\|} \quad (15)$$

When the data are linearly indivisible in low-dimensional space, SVM maps it to high-dimensional space by a kernel function, and then classifies the data. Like SVM, TBSVM uses a kernel function to map data, and then uses a linear method to classify data. With the introduction of the kernel function, the two hyperplanes of TBSVM can be expressed as:

$$\begin{aligned} K(x^T, C^T)u_1 + b_1 &= 0 \\ K(x^T, C^T)u_2 + b_2 &= 0 \end{aligned} \quad (16)$$

The optimization problems of the nonlinear TBSVM are as follows:

$$\begin{aligned} \min_{u_1, b_1, \xi, \xi^*} \quad & \frac{1}{2}c_3(\|u_1\|^2 + b_1^2) + \frac{1}{2}\xi^{*T}\xi^* + c_1e_2^T\xi \\ \text{s.t.} \quad & K(A, C^T)u_1 + e_1b_1 = \xi^* \\ & -(K(B, C^T)u_1 + e_2b_1) + \xi > e_2, \xi \geq 0 \end{aligned} \quad (17)$$

$$\begin{aligned} \min_{u_2, b_2, \eta, \eta^*} \quad & \frac{1}{2}c_4(\|u_2\|^2 + b_2^2) + \frac{1}{2}\eta^{*T}\eta^* + c_2e_1^T\eta \\ \text{s.t.} \quad & K(B, C^T)u_2 + e_2b_2 = \eta^* \\ & -(K(A, C^T)u_2 + e_1b_2) + \eta > e_1, \eta \geq 0 \end{aligned} \quad (18)$$

At this point, define:

$$\begin{aligned} S &= [K(A, C^T)e_1] \\ R &= [K(B, C^T)e_2] \end{aligned} \quad (19)$$

The Lagrange dual problem corresponding to TBSVM can be obtained as follows:

$$\begin{aligned} \min_{\alpha} \quad & e_2^T \alpha - \frac{1}{2} \alpha^T R(SS^T + c_3 I)^{-1} R^T \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 \end{aligned} \quad (20)$$

$$\begin{aligned} \min_{\gamma} \quad & e_1^T \gamma - \frac{1}{2} \gamma^T S(RR^T + c_4 I)^{-1} S^T \gamma \\ \text{s.t.} \quad & 0 \leq \gamma \leq c_2 \end{aligned} \quad (21)$$

At this time, when a new sample is to be input, the nonlinear TBSVM can judge the label to which the sample belongs according to the following formula:

$$\text{Class } i = \arg \min_{k=1,2} \frac{|(K(x^T, C^T)u_k + b_k)|}{\sqrt{u_k^T K(x^T, C^T)u_k}} \quad (22)$$

### 3.2. Grid Search Algorithm

The grid search method is an exhaustive search algorithm for specifying parameter values, and the optimal learning algorithm is obtained by optimizing the parameters of the estimation function through cross-validation [23].

First, a series of parameters are selected from the candidate parameter set and combined to obtain the candidate parameter list [24]. Then, the parameter list is traversed, the candidate parameters are entered into the model, and the score of the parameter combination is calculated. Finally, from the candidate parameter list, the parameter with the highest score is selected as the optimal parameter of the model.

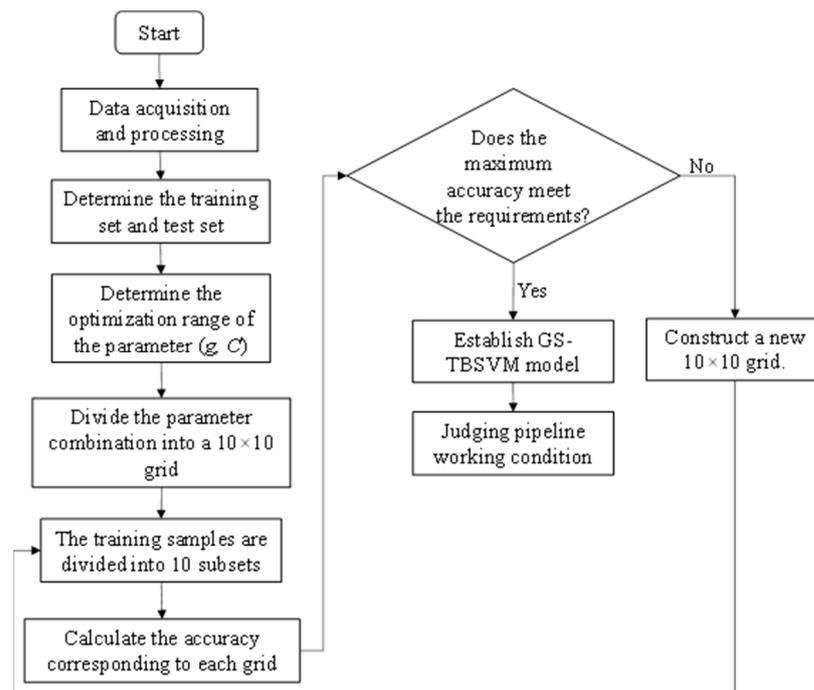
For grid search, if a point in the parameter space points to the real minimum value of the loss function, the minimum value and the corresponding parameters can be captured. The larger and denser the parameter space, the greater the possibility that the combination in the parameter space just covers the minimum point of the loss function. That is, in extreme cases, when all possible values in the parameter space are exhausted, the grid search will certainly be able to find the optimal parameter combination corresponding to the minimum value of the loss function, and the generalization ability of this parameter combination must be stronger than that obtained manually.

### 3.3. Grid Search Algorithm-Optimized Twin-Bounded Support Vector Machine (GS-TBSVM)

Because the parameters of the traditional TBSVM model are randomly obtained, the accuracy of the model is not high. The kernel function of the TBSVM model determines the accuracy of the model to a great extent, so optimizing the parameters of the kernel function can improve the performance of the model. In this paper, a GS-TBSVM model is constructed according to the grid search method and TBSVM. The model adopts the grid search method to optimize the parameters of kernel function in TBSVM and improve recognition accuracy.

The RBF kernel function is one of the commonly used kernel functions. It can realize nonlinear mapping and has few parameters, which will not affect the complexity of the model. Therefore, in the nonlinear case, the RBF kernel function is adopted in this paper,  $K(x_i, x_j) = \exp(-\lambda \|x_i - x_j\|^2)$ .

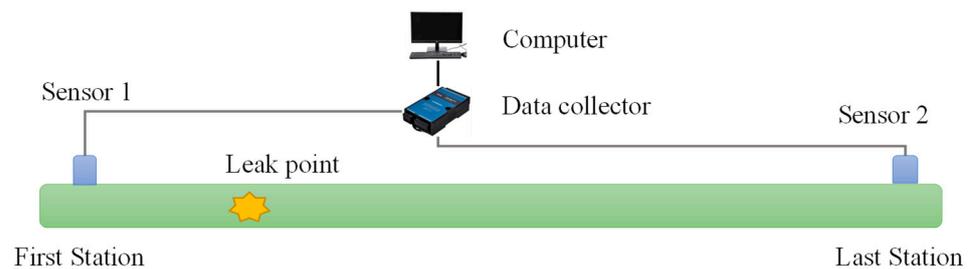
The parameters of RBF include kernel parameter  $g$  and penalty parameter  $C$ , and the grid search algorithm will first take these two parameters as a combination to form a parameter space. Then, the space is divided into equal grids according to the proposed coordinate system, and each group of vectors in the coordinate system represents a group of  $(C, g)$  values. By bringing each group of data in the space into the TBSVM, the performance of the model is verified until the optimal values of kernel parameter  $g$  and penalty parameter  $C$  are found. The specific process is shown in Figure 2.



**Figure 2.** Optimization process of the TBSVM model.

#### 4. Experiment and Discussion

In this paper, a gas pipeline of a petrochemical company was used for leakage experiments. The length of the pipeline was 12 km, the diameter was 200 mm, and the pressure at the first end of the pipeline was 2.3 MPa and 0.6 MPa, respectively. At 1 km, 3 km, 5 km, 7 km, and 11 km of the pipeline there was a ball valve of 8-mm and 5-mm bore. A PCB 106B acoustic sensor with a range of 0–57 kPa, a sensitivity of 43.5 Mv/kPa, and a sampling frequency of 1000 Hz was installed at each end of the pipeline to collect the acoustic signals in the pipeline. The NI-DAQ9181 data acquisition device was used to collect the data. The computer and software used were a PC with an Intel Pentium processor (2.90 GHz) and 6 GB RAM and MATLAB R2014a. The experimental schematic is shown in Figure 3. The experimental site is shown in Figure 4. Figure 5 shows the PCB 106b acoustic sensor.



**Figure 3.** Experimental schematic.



Figure 4. Experimental site.



Figure 5. PCB 106b acoustic sensor.

The acoustic signal collected by sensor 1 was used as an example for the denoising experiment. The 8-mm valve at 3 km of the pipeline was opened to simulate a leak in the pipeline, and the original signal collected by sensor 1 at the first station of the pipeline is shown in Figure 6.

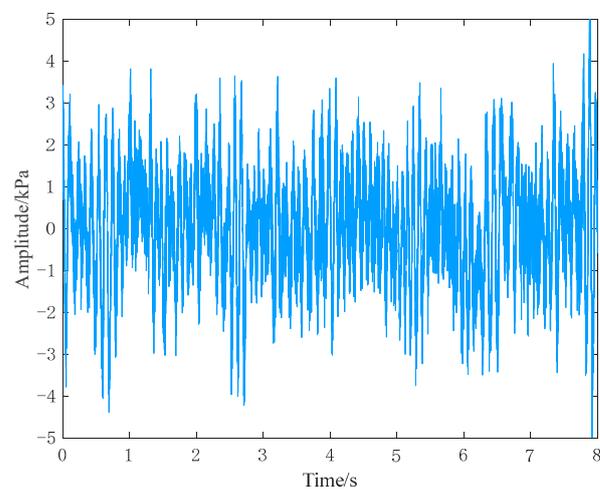
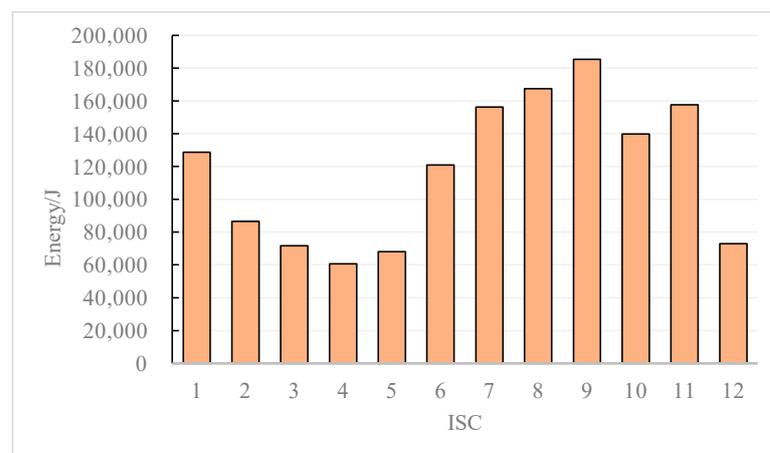


Figure 6. Original signal.

Figure 6 shows that the original signal contained more noise and the signal fluctuated more, and the trend of the signal curve could not be judged. Then, the IUPLCD algorithm was used to reduce the noise of the original signal. First, the original signal was decomposed into 12 ISC components of different scales and a residual term by using the UPLCD algorithm. Then, the energy value of each ISC component and the standard deviation of the amplitude after FFT transform of the components were calculated. Finally, the ISC components were filtered according to the energy values and standard deviations and the signal was reconstructed.

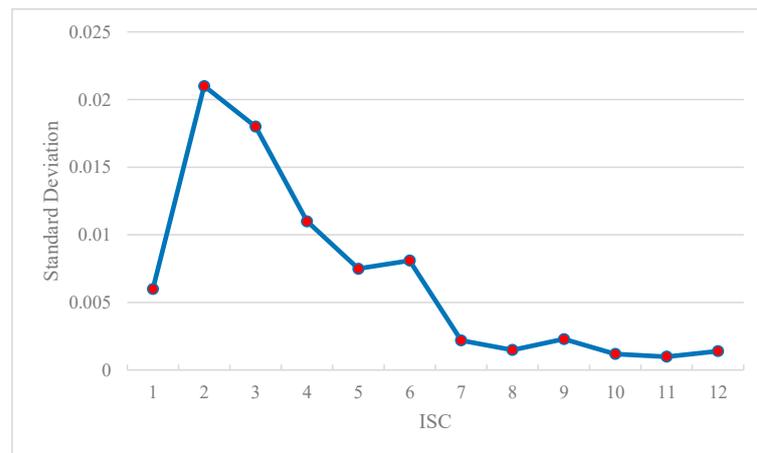
In the energy curve of ISC components, the first extreme point is critical for distinguishing whether the ISC components are dominated by noise or useful signals, that is, all components before the critical point of the energy curve (including at the critical point) are noise-dominated, and the components after the critical point are signal-dominated. Figure 7 shows the energy values of each ISC component.



**Figure 7.** Energy value of each ISC component.

As shown in Figure 7, the first extreme point of the energy curve was the point corresponding to ISC4. Therefore, ISC1-ISC4 are the noise-dominated components and ISC5-ISC12 are the signal-dominated components.

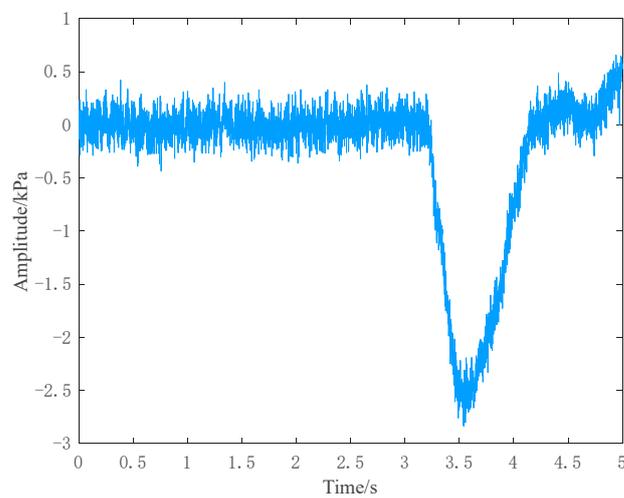
The Fourier transform of each ISC component obtained from the UPLCD decomposition yields spectrograms that vary widely and with different degrees of dispersion. The spectrograms of the noise-dominated ISC components are more complex. The standard deviation reflects the degree of dispersion of a data set, and it can visually reflect the degree of deviation of a data set compared to the variance. Therefore, the standard deviation can be used to determine which components are signal-dominated. To prevent the ISC components of different orders after the Fourier transform from differing in order of magnitude, this paper first normalizes the values and then calculates the standard deviations between different ISC components. Figure 8 shows the standard deviation of the magnitude of each ISC component after Fourier transform.



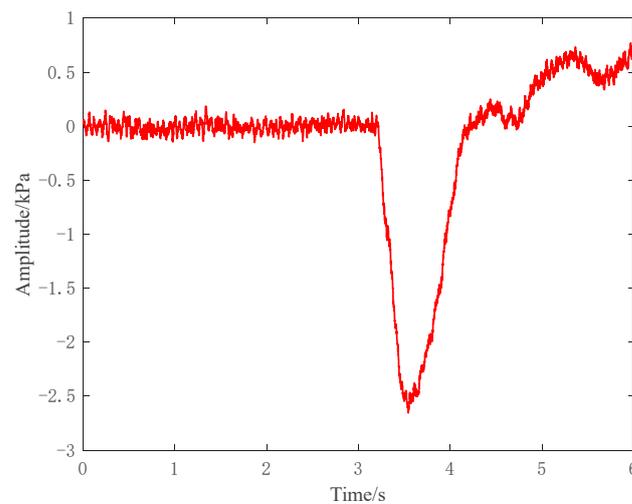
**Figure 8.** Standard deviation of amplitude.

From Figure 8, the standard deviation of ISC1-ISC6 had a large degree of variation and a high degree of dispersion, so it was the noise-dominated component. The standard deviation of ISC7-ISC12 had a smaller degree of variation and a low degree of dispersion, so ISC7-ISC12 was the signal-dominated component.

To preserve the original information to the greatest extent, the ISC components that met both the energy and standard deviation requirements were used for signal reconstruction. The energy and standard deviation curves of the ISC components were combined, and ISC7-ISC12 could meet both requirements, so ISC7-ISC12 was selected for signal reconstruction. Under the same conditions, UPLCD was selected to denoise the signal, and the obtained signal curve was compared with the signal curve after denoising by IUPLCD, as shown in Figures 9 and 10.



**Figure 9.** Signal curve after denoising by the UPLCD method.



**Figure 10.** Signal curve after denoising by the IUPLCD method.

From Figures 9 and 10, it is obvious that the curve obtained by the method used in this study is smoother, contains less noise, retains the original leakage information to the greatest extent, and can clearly illustrate the trend of the signal and the inflection point.

Leakage was simulated by opening 8-mm and 5-mm valves at 1 km, 3 km, 5 km, 7 km, and 11 km of the pipeline, and 3000 sets of leakage acoustic signals were collected for each. The same 3000 sets of data were collected under normal pipeline conditions. From the data, 2000 sets of data were selected as the training set of the model, and the remaining 1000 sets were used as the test set.

For the working conditions of the pipeline, the normal working condition category was set as 1 and the occurrence of leakage category as 2, and the GS-TBSVM model was applied to identify the working condition of the pipeline. The values of the kernel parameter  $g$  and the penalty parameter  $C$  were set in the range  $[0,100]$ , and the optimal parameters were obtained by using the grid search algorithm:  $g = 0.13$  and  $C = 51.6$ .

To verify the accuracy of the GS-TBSVM model, the GS-SVM model and GS-TWSVM model were selected for comparison with the model established in this paper. According to the grid search algorithm, the optimal parameters of the GS-SVM model are  $g = 0.43$  and  $C = 19.8$ . The optimal parameters of the GS-TWSVM model are  $g = 0.22$  and  $C = 31.1$ . Each of the above optimal parameters was brought to the model to judge the working conditions of the pipeline. The optimal model was selected by considering the training effect and testing effect of the model.

Leakage data at 3 km of the pipeline with an 8-mm leak aperture were used for training and testing, and the performance comparison of the three models is shown in Table 1.

**Table 1.** Performance comparison of different methods.

Denoising Method	Model	Time/s	Accuracy of Training Set %	Accuracy of Test Set %
UPLCD	GS-SVM	18.72	92.78	94.51
	GS-TWSVM	11.38	93.21	95.23
	GS-TBSVM	6.65	93.84	95.66
IUPLCD	GS-SVM	16.33	93.46	95.16
	GS-TWSVM	9.69	94.67	95.53
	GS-TBSVM	5.13	94.29	96.28

As shown in Table 1, the recognition accuracies of all three optimized models were above 94%. Among them, the recognition accuracy of the model built based on the IUPLCD denoising algorithm was significantly higher than that of the model built based on the

UPLCD denoising algorithm. This shows that the IUPLCD algorithm can effectively improve the recognition accuracy after processing the signal. The GS-TBSVM model based on the IUPLCD algorithm had the highest recognition accuracy, which could reach 96.28%, and the running time was 5.13 s. In a comprehensive view, the GS-TBSVM model had the best performance, the GS-TWSVM had the second-best performance, and the GS-SVM had the worst performance, relatively speaking.

The collected data were then all denoised using the IUPLCD algorithm, and then the GS-TBSVM model, GS-SVM model, and GS-TWSVM model were trained and tested. The comparison of the recognition accuracy of the three models under different working conditions is shown in Table 2.

**Table 2.** The FAR and FDR of IRPCA-based and PCA-based leak detection methods.

Work Conditions	Leakage Caliber	GS-SVM	GS-TWSVM	GS-TBSVM
Leakage at 1 km	8 mm	93.9%	95.6%	97.3%
	5 mm	94.3%	94.8%	96.8%
Leakage at 3 km	8 mm	95.6%	95.5%	96.3%
	5 mm	95.1%	95.5%	96.1%
Leakage at 5 km	8 mm	96.6%	96.1%	97.5%
	5 mm	94.1%	96.5%	97.1%
Leakage at 7 km	8 mm	95.4%	95.3%	98.4%
	5 mm	94.6%	94.8%	97.5%
Leakage at 11 km	8 mm	96.3%	96.7%	97.1%
	5 mm	95.8%	96.5%	96.9%

According to Table 2, the recognition accuracy of GS-TBSVM for pipeline leaks was significantly higher than that of GS-SVM and GS-TWSVM models, regardless of leaks at different locations or leaks of different sizes at the same location, and the recognition accuracy of the GS-TBSVM model for pipeline leaks was up to 98.4%.

## 5. Conclusions

To identify pipeline leaks accurately and in a timely manner, improve pipeline management, and prevent safety accidents, a pipeline leak detection method based on IUPLCD and GS-TBSVM is proposed in this paper. In response to the drawback that the UPLCD algorithm cannot select the effective ISC components, the algorithm was optimized and improved. The energy value and standard deviation of the amplitude of ISC components were used as the screening criteria for effective ISC components, and the signal-dominated ISC components were screened out and used in the signal reconstruction. The reconstructed signal had extremely low noise content, which better restored the original leakage signal. Finally, to identify pipeline leakage accurately and in a timely manner, a new pipeline leakage identification model GS-TBSVM was constructed based on the advantages of the grid search algorithm and TBSVM. This field experiment verifies that the noise reduction signal was input into the GS-TBSVM model, and the model can accurately determine the working condition of the pipeline and has a high recognition accuracy.

Compared with UPLCD and other denoising algorithms, the IUPLCD algorithm proposed in this paper is optimized in the screening of ISC components, which can accurately screen out the effective ISC components for signal reconstruction, and the noise content of the reconstructed signal is lower, and the original information of leakage is retained to a greater extent. In the identification of pipeline working conditions, compared with GS-SVM and GS-TWSVM models, the GS-TBSVM model constructed in this study has faster parameter seeking and higher identification accuracy, which effectively improves the performance of the leak detection system.

The leak detection method proposed in this paper was applied to gas pipelines with good results and was found to accurately identify leaks in gas pipelines and effectively improve the performance of leak detection. For a complete leak detection system, in addition to the leak detection function, the system should also have a leak locating function so that it can find the leak location in time to deal with the leak when it is detected. Therefore, in our future work, we plan to investigate the leak localization method for gas pipelines to achieve precise localization of leaks.

**Author Contributions:** Conceptualization, H.S. and Y.Z.; methodology, H.S. and Y.Z.; software, Y.Z. and H.S.; validation, H.S. and Y.Z.; formal analysis, H.S. and Y.Z.; investigation, Y.Z. and H.S.; resources, H.S. and Y.Z.; data curation, H.S. and Y.Z.; writing—original draft preparation, H.S. and Y.Z.; writing—review and editing, H.S. and Y.Z.; visualization, H.S. and Y.Z.; supervision, H.S. and Y.Z.; project administration, H.S. and Y.Z.; funding acquisition, H.S. and Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant no. 61673199 and no. 62073158), the Natural Science Foundation of Liaoning Province (grant no. 2019-BS-158), the Scientific Research Funds of Liaoning Provincial Department of Education (grant no. L2020017), the China Postdoctoral Science Foundation (grant no. 2020M670796), the Talent Scientific Research Fund of Liaoning Petrochemical University (grant no. 2019XJL-008).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Wang, W.; Mao, X.; Liang, H.; Yang, D.; Zhang, J.; Liu, S. Experimental research on in-pipe leaks detection of acoustic signature in gas pipelines based on the artificial neural network. *Measurement* **2021**, *183*, 109875. [[CrossRef](#)]
2. Holmes, C.; Yoshikawa, N.; Minns, R. Monitoring water contamination in jet fuel using silica-based Bragg gratings. *IEEE Sens. J.* **2019**, *19*, 2984–2990. [[CrossRef](#)]
3. Mujtaba, S.M.; Lemma, T.A.; Taqvi, S.A.A.; Ofei, T.N.; Vandrangi, S.K. Leak Detection in Gas Mixture Pipelines under Transient Conditions Using Hammerstein Model and Adaptive Thresholds. *Processes* **2020**, *8*, 474. [[CrossRef](#)]
4. Zhou, M.; Zhang, Q.; Liu, Y.; Sun, X.; Cai, Y.; Pan, H. An Integration Method Using Kernel Principal Component Analysis and Cascade Support Vector Data Description for Pipeline Leak Detection with Multiple Operating Modes. *Processes* **2019**, *7*, 648. [[CrossRef](#)]
5. Lang, X.; Zhu, Y. An analysis of detectable leakage rate for oil pipelines based on acoustic wave method. *Meas. Sci. Technol.* **2022**, *33*, 125108. [[CrossRef](#)]
6. Kousiopoulos, G.-P.; Kampelopoulos, D.; Karagiorgos, N.; Papastavrou, G.-N.; Konstantakos, V.; Nikolaidis, S. Acoustic Leak Localization Method for Pipelines in High-Noise Environment Using Time-Frequency Signal Segmentation. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [[CrossRef](#)]
7. Xiao, R.; Li, J. Evaluation of acoustic techniques for leak detection in a complex low-pressure gas pipeline network. *Eng. Fail. Anal.* **2023**, *143*, 106897. [[CrossRef](#)]
8. Lang, X.; Yuan, W. Localization Method of Multiple Leaks Based on Time-Frequency Analysis and Improved Differential Evolution. *IEEE Sens. J.* **2020**, *20*, 14383–14390. [[CrossRef](#)]
9. Zhang, Z.; Zhang, L.; Fu, M.; Ozevin, D.; Yuan, H. Study on leak localization for buried gas pipelines based on an acoustic method. *Tunn. Undergr. Space Technol.* **2022**, *120*, 104247. [[CrossRef](#)]
10. Liu, C.; Li, Y.; Fang, L. New leak-localization approaches for gas pipelines using acoustic waves. *Measurement* **2019**, *134*, 54–65. [[CrossRef](#)]
11. Meng, Q.; Lang, X.; Lin, M. Leak Localization of Gas Pipeline Based on the Combination of EEMD and Cross-spectrum Analysis. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 1–9. [[CrossRef](#)]
12. Shi, C.; Lang, X.; Zhang, L. Leak Detection and Localization of Pipelines Based on LMD and LSTSVM. *J. Liaoning Shihua Univ.* **2019**, *39*, 84–90.
13. Li, S.; Cai, M.; Han, M.; Dai, Z. Noise Reduction Based on CEEMDAN-ICA and Cross-Spectral Analysis for Leak Location in Water-Supply Pipelines. *IEEE Sens. J.* **2022**, *22*, 13030–13042. [[CrossRef](#)]
14. Li, J.; Ke, L.; Du, Q. Heart sound signal classification algorithm: A combination of wavelet scattering transform and twin support vector machine. *IEEE Access* **2019**, *7*, 179339–179348. [[CrossRef](#)]

15. Huang, L.; Shao, Y.; Zhang, J.; Zhao, Y.; Teng, J. Robust Rescaled Hinge Loss Twin Support Vector Machine for Imbalanced Noisy Classification. *IEEE Access* **2019**, *7*, 65390–65404. [[CrossRef](#)]
16. Lang, X.; Li, P.; Hu, Z.; Ren, H.; Li, Y. Leak Detection and Location of Pipelines Based on LMD and Least Squares Twin Support Vector Machine. *IEEE Access* **2017**, *5*, 8659–8668. [[CrossRef](#)]
17. Kang, J.; Park, Y.; Lee, J.; Wang, S.; Eom, D. Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4279–4289. [[CrossRef](#)]
18. Cheng, J.; Zheng, J.; Yang, Y. A nonstationary signal analysis approach—The local characteristic-scale decomposition method. *J. Vib. Eng.* **2012**, *25*, 215–220.
19. Zheng, J.; Su, M.; Pan, H.; Tong, J.; Pan, Z. Uniform phase local characteristic-scale decomposition and its applications in mechanical fault diagnosis. *J. Electron. Meas. Instrum.* **2021**, *35*, 50–58.
20. Zhu, S.; Liu, L.; Yao, Z. The application of threshold empirical mode decomposition de-noising algorithm for battlefield ambient noise. *Int. J. Model. Simul. Sci.* **2019**, *9*, 1850027. [[CrossRef](#)]
21. Fina, M.; Lauff, C.; Faes, M.; Valdebenito, M.; Wagner, W.; Freitag, S. Bounding imprecise failure probabilities in structural mechanics based on maximum standard deviation. *Struct. Saf.* **2023**, *101*, 102293. [[CrossRef](#)]
22. Shao, Y.; Zhang, C.; Wang, X.; Deng, N. Improvements on Twin Support Vector Machines. *IEEE Trans. Neural. Netw.* **2011**, *22*, 962–968. [[CrossRef](#)] [[PubMed](#)]
23. Huang, Q.; Mao, J.; Liu, Y. An improved grid search algorithm of SVR parameters optimization. In Proceedings of the 2012 IEEE 14th International Conference on Communication Technology, Chengdu, China, 9–11 November 2012; pp. 1022–1026.
24. Chang, Z.; Yuan, W.; Huang, K. Remaining useful life prediction for rolling bearings using multi-layer grid search and LSTM. *Comput. Electr. Eng.* **2022**, *101*, 108083.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.