An Integration Method Using Kernel Principal Component Analysis and Cascade Support Vector Data Description for Pipeline Leak Detection with Multiple Operating Modes

Mengfei Zhou [ID]**, Qiang Zhang, Yunwen Liu, Xiaofang Sun *, Yijun Cai and Haitian Pan**

Department of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310027, China;
mfzhou@zjut.edu.cn (M.Z.); aizjut@163.com (Q.Z.); threight@163.com (Y.L.); hgybcyj@zjut.edu.cn (Y.C.);
htpan@zjut.edu.cn (H.P.)

* Correspondence: zgdsxf@zjut.edu.cn

Abstract: Pipelines are one of the most efficient and economical methods of transporting fluids, such as oil, natural gas, and water. However, pipelines are often subject to leakage due to pipe corrosion, pipe aging, pipe weld defects, or damage by a third-party, resulting in huge economic losses and environmental degradation. Therefore, effective pipeline leak detection methods are important research issues to ensure pipeline integrity management and accident prevention. The conventional methods for pipeline leak detection generally need to extract the features of leak signal to establish a leak detection model. However, it is difficult to obtain actual leakage signal data samples in most applications. In addition, the operating modes of pipeline fluid transportation process often have frequent changes, such as regulating valves and pump operation. Aiming at these issues, this paper proposes a hybrid intelligent method that integrates kernel principal component analysis (KPCA) and cascade support vector data description (Cas-SVDD) for pipeline leak detection with multiple operating modes, using data samples that are leak-free during pipeline operation. Firstly, the local mean decomposition method is used to denoise and reconstruct the measured signal to obtain the feature variables. Then, the feature dimension is reduced and the nonlinear principal component is extracted by the KPCA algorithm. Secondly, the K-means clustering algorithm is used to identify multiple operating modes and then obtain multiple support vector data description models to obtain the decision boundaries of the corresponding hyperspheres. Finally, pipeline leak is detected based on the Cas-SVDD method. The experimental results show that the proposed method can effectively detect small leaks and improve leak detection accuracy.

Keywords: multiple operating modes; cascade support vector data description; leak detection; K-means; kernel principal component analysis

1. Introduction

Pipeline fluid transport is one of the more safe, efficient, and energy-saving methods, which plays an increasingly important role in the development of national economy. However, pipelines are often subject to leakage due to pipe corrosion, pipe aging, pipe weld defects, or damage by a third-party [1]. In addition, pipeline fluids are also often flammable, toxic, corrosive media. Therefore, pipeline leakage will not only lead to economic loss of products and resources, but also seriously pollute the environment [2,3]. Pipeline safety management and accident prevention are increasingly required in countries around the world. Establishing an efficient pipeline leak detection system can report the occurrence and localization of leak accidents in a timely manner, and minimize or even avoid economic

losses and environmental pollution. So far, some comprehensive review papers about pipeline leak and localization have been published [4–8]. From the perspective of measurement signal acquisition, pipeline leak detection and positioning systems can be broadly classified into external detection methods and internal detection methods [2]. Externally-based methods monitor external pipeline parameters, such as acoustic signals [9–11] and fiber-optic cables [12–14], while internally-based methods typically collect pressure, flow, and temperature signals, such as real time transient modeling [15–18], negative pressure wave method [19,20], pressure point analysis, and the flow balance method. Recently, some scholars have also pointed out that integrating multi-source signals, including internal sensors and external sensors, is also an effective way to improve the performance of pipeline leak detection and localization [21]. Saqib et al. [22] used pressure and vibration signals to detect and locate leaks in the water pipeline network. Lang et al. [23] combined the ultrasonic wave velocity and flow signals to detect small leaks in the pipelines of the experimental set-up. The experimental results showed that multi-sensor information fusion for leak detection has a greater improvement than single-type sensors.

From the perspective of the model used for pipeline leak detection, pipeline leak detection and localization methods can be divided into two categories, based on the mechanism model and the data-driven model. The mechanism model is usually based on the conservation of fluid mass, momentum conservation, as well as energy conservation in the pipeline. The mechanism model can be used to detect leaks in pipelines under steady-state and transient conditions [5]. Typically, the mechanism model-based methods compare the estimated value obtained from the model with the actual measured value, and if the residual between the estimated value and the actual measured value is greater than a pre-specified threshold, it indicates that a leak is detected [24]. The accuracy of these methods depends to a large extent on model parameters and sensor accuracy. Furthermore, the method requires extensive simulation and calibration work [8]. In addition, high computational loads are required to solve these complex nonlinear models [25,26]. In recent years, data-driven methods for pipeline leak detection and localization have been developed rapidly, which rely on measuring data and performing signal processing and statistical analysis for leak detection. The advantage of these methods is that they do not require any specific insight into the hydraulics mechanism, and only through machine learning algorithms or artificial intelligence algorithms, plus some statistical or pattern recognition tools, to obtain pipeline leak characteristics and knowledge from the collected data. Recently, Wu and Liu [8] presented a detailed review on data-driven approaches for leak detection (specifically, burst detection) in the water distribution. Typically, artificial neural network-based methods [11,27,28] and support vector machine-based methods [13,28–30] are the most widely used data-driven approaches for pipeline detection and localization. Besides these, there are other data-driven methods for pipeline leak detection, such as genetic algorithm [31], principal component analysis (PCA) [31], particle swarm optimization [12], support vector data description (SVDD) [32], and Bayesian reasoning [33,34]. In addition to being independently applied to pipeline leak detection, some of these methods are often integrated as a hybrid method for leak detection. Ahn et al. [31] used the genetic algorithm and PCA to extract acoustic emission signal features, and used support vector machine to detect leaks. The particle swarm optimization method, integrating with support vector machine, was proposed in the literature [12,35,36], in which the particle swarm optimization algorithm was used to optimize the parameters of support vector machine to improve its classification accuracy. The experimental results showed that the particle swarm optimization has strong global search ability when optimizing support vector machine parameters, which further improves the accuracy of leak detection. Similarly, the particle swarm optimization algorithm was used to optimize the parameters of the kernel functions of support vector machine and support vector regression in Jia et al. [13], in which the support vector machine was used for pipeline leak detection and support vector regression for pipeline leak localization. Mandal et al. [37] proposed a leak detection approach based on the rough set theory and support vector machine to improve leak detection accuracy. In this method, the rough set theory was used to reduce the length of experimental data as well as generated rules. Meanwhile, the artificial bee colony algorithm was used for the computational training for the support vector

machine. Li et al. [11] specifically studied the leak detection of a water distribution system subject to failure of the socket joint. The acoustic characteristics of leak signals in the socket and spigot pipe segments were extracted and selected. An artificial neural network was established as the classifier. More recently, an adaptive design was proposed that combined one-dimensional convolutional neural networks and support vector machine [28]. This method enabled fast and accurate leak detection. Moreover, a graph-based localization algorithm was proposed to determine the leak location within a real water distributed system.

To this end, most of the published data-driven methods require extracting features from pipeline leak signal data samples to develop a classification or prediction model for leak detection. To ensure the accuracy and versatility of an algorithm, it is necessary to introduce leaks with different leak levels at different leak locations before the leak detection system operates. This is impossible and unrealistic in most applications [38]. An alternative suggestion is to use simulation techniques to generate leak samples for training for data-driven methods. However, due to the uncertainty and complexity of actual leaks, it is difficult to fully simulate the actual leakage signal with complete characteristics. Therefore, some researchers have introduced some leak detection methods that require only leak-free sample signals. For example, Wang et al. [32] extracted the time-domain statistical characteristics of the acoustic sensor from the normal (no leak) sample signal, and constructed the SVDD model, which was implemented in a field leak detection system.

In addition, frequent changes in various operating modes, such as normal running, operating conditions adjustment, and pump operation, are often encountered during the pipeline fluid transportation process. The training samples in the feature space of each operating mode are different in nature and unevenly distributed. Moreover, the characteristics of working condition adjustment and pump operation have certain similarities with the one of pipeline leakage, resulting in a high false alarm rate for the pipeline leak system. As a result, the effective classification of pipeline operating modes is of great significance for improving the accuracy of the leak detection system.

In this paper, we present a novel integration method using kernel principal component analysis (KPCA) and cascade support vector data description (Cas-SVDD), namely, KPCA-Cas-SVDD, for pipeline leak detection with multiple operating modes. On the one hand, only the leak-free data samples are required from the actual running process, and the features are extracted and reconstructed by local mean decomposition (LMD). After that, KPCA is used to reduce the feature dimensions. On the other hand, based on the K-means clustering method, the various normal operating modes of the pipeline are classified, based on which the SVDD model is established for each operating mode. Subsequently, the Cas-SVDD method is used for pipeline leak detection, which greatly improves the leak detection accuracy. The rest of this paper is structured as follows. Section 2 presents the novel methodology for pipeline leak detection, including the basic algorithms of LMD, KPCA, K-means, and SVDD. Section 3 presents the background of the case study, data processing, and feature extraction. Section 4 consists of experimental results and discussion. Section 5 addresses the conclusion of this research and directions for future work.

2. Methodology

2.1. LMD Based Signal Processing and KPCA for Feature Extraction

The local mean decomposition (LMD) was recently developed to analyze time series signals with nonlinear and non-stationary features, that can adaptively denoise and extract features from the original signals [39]. Meanwhile, the LMD adaptively decomposes the original signal into pure frequency modulated signals and envelope components of different magnitude, and then obtains a set of product functions (PF), each of which is the product of a pure frequency-modulated signal and an envelope signal [40]. Combining the instantaneous amplitude and instantaneous frequency of all PF components, the complete time–frequency distribution of the original signal can be obtained, so that the feature information in the original signal can be extracted more effectively.

Given any non-stationary measurement signal $x(t)$, its decomposed PF components via the LMD algorithm can be written as follows:

$$\text{PF}_i(t) = a_i(t)s_i(t), \tag{1}$$

where $a_i(t)$ is the instantaneous amplitude of the PF component, and $s_i(t)$ is a pure frequency modulation signal. The instantaneous frequency of the PF component is obtained by processing the pure frequency modulation signal by the following formula:

$$f_i(t) = \frac{1}{2\pi}\frac{d[arccos(s_i(t))]}{dt}. \tag{2}$$

All the PF components are separated from the signal $x(t)$, and finally a residual component $e_k(t)$ is obtained, which is constant or monotonic. In this way, the original signal $x(t)$ can be represented into the sum of k PF components and $e_k(t)$:

$$x(t) = \sum_{p=1}^{k} \text{PF}_p(t) + e_k(t). \tag{3}$$

The noise in $x(t)$ can be removed by carefully selecting PF and $e_k(t)$ according to frequency. After the original signal is denoised and reconstructed by the LMD, 12 feature variables are extracted, including time-domain features (mean, variance, effective value, square root amplitude, and energy) and waveform-domain features (kurtosis, skewness parameter, kurtosis factor, pulse factor, shape parameter, peak coefficient, and valley factor), as shown in Table 1.

Table 1. The equations of features in the time domain and waveform domain.

No.	Parameters	Equations	No.	Parameters	Expression
T1	mean	$x_{am} = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert$	T7	pulse factor	$x_{imf} = \frac{max\{\lvert x_i\rvert\}}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$
T2	variance	$x^2 = \frac{1}{n}\left\{\sum_{i=1}^{n}(x(t_i)-\overline{x}(t))^2\right\}$	T8	shape factor	$S = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}}{\left(\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert\right)}$
T3	effective value	$E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$	T9	peak coefficient	$F = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}}{max(x_i)-min(x_i)}$
T4	kurtosis	$x_k = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^4\right)^{\frac{1}{4}}$	T10	square root amplitude	$x_r = \left(\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^{\frac{1}{2}}\right)^2$
T5	skewness parameter	$SF = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}x_i^3\right)}{\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)^3}$	T11	valley factor	$L = \frac{max\{\lvert x_i\rvert\}}{\left(\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^{\frac{1}{2}}\right)^2}$
T6	kurtosis factor	$x_{kf} = \frac{x_k}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}}$	T12	energy	$E = \sum_{i=1}^{n}\lvert x(t_i)\rvert^2$

However, too many features that may include invalid or redundant features can result in reduced accuracy of leak detection due to overfitting and greatly increase computational complexity. Principal components analysis (PCA) is one of the classical techniques for multivariate analysis, based on which the dimensionality is reduced by preserving the most variance of the original data. Usually, PCA decorrelates the variables by obtaining a linear relationship that handles the two-order correlation between the variables. However, it is less effective in extracting nonlinear features of the pipeline measurement variables. In this case, kernel PCA (KPCA), using the idea of kernel function, is proposed to handle nonlinear feature extraction for the original signal by finding a suitable nonlinear mapping function $\Phi(X)$, by which the low-dimensional data sample set $X = [x_1, x_2, \cdots, x_n]^T$, where $x_i \in R^m$ $(i = 1, 2, \cdots, n)$, n is the number of samples, and m is the number of variables, is mapped to high-dimensional space F to construct the feature space, and thereby each variable is subjected to principal component analysis from the high-dimensional feature space [41,42]. We shall briefly review the KPCA method here.

The covariance matrix C^F on the F space is expressed by:

$$C^F = \frac{1}{n} \sum_{j=1}^{n} \Phi(x_i) \Phi^T(x_j). \tag{4}$$

The eigenvalue decomposition equation of the Equation (4) can be obtained by:

$$\lambda_i v_i = C^F v_i, \tag{5}$$

where λ_i and v_i are, respectively, the eigenvalues and the eigenvectors of the covariance matrix. The eigenvectors v_i of C^F can be expressed as:

$$v_i = \sum_{j=1}^{n} a_{ij} \Phi(x_i), \tag{6}$$

where a_{ij} is the feature space expansion coefficient. Taking Equations (4) and (6) into Equation (5), the following equation can be obtained:

$$\lambda_i \sum_{j=1}^{n} a_{ij} \Phi(x_i) \Phi^T(x_j) = \frac{1}{n} \sum_{j=1}^{n} a_{ij} \left[\Phi(x_i), \sum_{j=1}^{n} \Phi(x_j) \right] \left[\Phi(x_j), \Phi(x_i) \right]. \tag{7}$$

Define an $n \times n$ kernel matrix K, where $K_{ij} = \left[\Phi(x_i), \Phi(x_j) \right]$, then Equation (7) can be expressed as:

$$n\lambda_i a_i = K a_i, \tag{8}$$

where $a_i = [a_{i1}, a_{i2}, \cdots, a_{in}]^T$, the eigenvalue of the matrix K satisfies $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. One can choose to preserve the pre-p ($p \leq n$) eigenvalues and eigenvectors according to the cumulative variance contribution rate criteria (e.g., up to 0.85). Through this method, KPCA achieves dimensionality reduction and nonlinear feature extraction for original signal. Commonly kernel functions used of KPCA include linear kernel, multi-layer perceptual kernel function, Gaussian kernel, and polynomial kernel. Of these, the Gaussian kernel was used in KPCA in our study, which provides better performance, regardless of the total sample size and feature dimension.

2.2. Pipeline Leak Detection Method Based on K-Means and Cas-SVDD

The SVDD algorithm only requires data samples under normal operating conditions, the main idea of which is to create a closed compact hypersphere that contains as many data samples as possible [43,44]. SVDD is an excellent method of one-class classifier with the advantages of robustness, good generalization, and high computation efficiency. Additionally, SVDD inherits small-sample characteristics. However, due to the uncertainties caused by natural changes (such as corrosion and sediments), and human factors (such as valve operation and demand changes), there are frequent changes in operating modes in the pipeline during the fluid transportation process. Moreover, the signal characteristics generated by some pipeline operations, such as valve adjustment and pump opening or closing, are similar to those generated by pipeline leakage. If the different normal operating mode data samples collected are used to create a single SVDD hypersphere, the volume of the hypersphere may be too large, which would lead to an increase in false alarm rate. Therefore, it is necessary that the various operating modes data during normal operation should be first classified, and then multiple SVDDs established.

In our work, the K-means clustering algorithm was used to identify each operating mode of the data samples collected from normal operation of the pipeline. The K-means clustering, as an unsupervised clustering algorithm, is a mature and widely used clustering method. It has the advantages of its simplicity, favorable execution time, and good clustering effect [45]. Assuming that there are P operating modes in the pipeline transportation process, the algorithm will divide the original data set W into P clusters, each of which has high data similarity, with low similarity between

clusters. First, the *P* data samples are randomly selected in the data set *W*, each as the center of the initial clusters, and then the distance between the remaining data samples and the initial cluster center is calculated. After that, the data samples of the minimum distance from the certain center of the cluster are assigned to the nearest cluster. Subsequently, the average of all the data in each cluster is calculated. Thereby, the new *P* cluster centers are obtained. The iterative calculation is performed until each cluster center value does not change. At that time the update process stops. The K-means algorithm chooses squared Euclidean distance as the dissimilarity measure, so the optimization problem can be formulated as:

$$\min_{Q_1,\cdots,Q_i} H = \frac{1}{W} \sum_{i=1}^{P} \sum_{q \in Q_i} q - C_i^2, \tag{9}$$

where *H* is the function value of the sum of mean squared deviations, *q* is the given data samples in cluster Q_i, and C_i is the mean of the cluster Q_i (i.e., cluster center).

Through the K-means method, the original data set was classified into several sub-data sets with corresponding operating modes. However, as the K-means clustering method does not consider the anomaly data samples in each cluster and remove them, it was necessary to further describe the cluster-intensive data samples using SVDD for each cluster after K-means clustering. In this way, it the method not only overcomes the influence of abnormal sample noise, but also overcomes the shortcoming of the SVDD algorithm when the sample population density is not high.

As mentioned earlier, the data samples for each type of operating mode only participate in the training of constructing an SVDD hypersphere, so that each SVDD hypersphere contains all or as many data samples of the corresponding operating mode. In this way, the multiple SVDD models can be established. Figure 1 gives the schematic diagram.
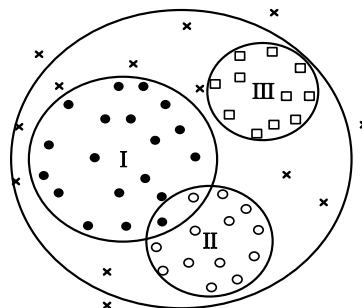

Figure 1. Schematic diagram of hyperspheres obtained by multiple SVDDs (i.e., each SVDD corresponds to one operating mode), where the symbols (○, ●, and □) denote the corresponding samples from different modes, and the symbol (×) denotes outlier samples.

Even for single class data samples obtained by the K-means algorithm, there are often some data samples of a non-target class, which lead to a larger volume of SVDD hypersphere obtained by training, which will increase the false alarm rate. Additionally, if these data samples near the boundary of the hypersphere are removed, the volume of the obtained hypersphere becomes smaller, resulting in an increase in the false negative rate [46]. Therefore, in order to enhance the robustness of single SVDD, a relaxation factor and a penalty parameter were introduced. The problem of determining the minimum SVDD hypersphere is formulated as the following optimization problem:

$$\min_{R_p, a_p, \xi_{p,i}} J = R_p^2 + C_p \sum_{i=1}^{N} \xi_{p,i}$$
$$s.t.$$
$$\|x_{p,i} - a_p\|^2 \le R_p^2 + \xi_{p,i}, \ i = 1, 2, ..., N \tag{10}$$
$$\xi_{p,i} \ge 0, \ i = 1, 2, ..., N; \ p = 1, 2, ..., P,$$

where ξ is the relaxation factor, which indicates the probability that some training data samples are allowed to be erroneously classified; C_p is the penalty parameter, which is used to control the degree

of punishment for misclassified data samples, and thus plays a role in hypersphere volume and misclassification; N is the number of data samples; P is the number of operating modes; a_p and R_p are the center and radius of the p-th hypersphere, respectively.

The kernel function used in SVDD maps raw training data from a low-dimensional space to a high-dimensional feature space, thereby constructing a compact hypersphere in a high-dimensional feature space that contains all or all of the target training data samples. The Gaussian kernel was used for SVDD here, which is commonly used for a one-class classifier, such as SVDD, support vector machine, and Parzen density [47]. After introducing the kernel function, the dual problem of the optimization problem of Equation (9) can be obtained as follows:

$$
\begin{aligned}
&\min_{a_{p,i}} J = \sum_{i=1}^{n} a_{p,i} K\left(x_{p,i},\ x_{p,j}\right) - \sum_{i=1}^{n} \sum_{j=1}^{n} a_{p,i} a_{p,j} K\left(x_{p,i},\ x_{p,j}\right) \\
&s.t. \\
&0 \leq a_{p,i} \leq C_p,\ \sum_{i=1}^{n} a_{p,i} = 1,
\end{aligned}
\tag{11}
$$

where a a is a Lagrange multiplier; $K\left(x_{p,i} \cdot x_{p,j}\right) = \left\langle \Phi\left(x_{p,i}\right), \Phi\left(x_{p,j}\right)\right\rangle$ is the kernel function used to calculate the inner product of the feature space. By solving the above quadratic programming problem, the radius of the p-th hypersphere is calculated by Equation (12):

$$
R_p = \sqrt{K\left(x_{p,i},\ x_{p,j}\right) - 2\sum_{i=1}^{n} a_{p,i} K\left(x_{p,i},\ x_{p,k}\right) + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{p,i} a_{p,j} K\left(x_{p,i},\ x_{p,j}\right)},
\tag{12}
$$

where $x_{p,k}$ is the support vector. With the same method, the spherical center and radius of other SVDD hyperspheres with different operating modes can be obtained.

Assuming that x_{new} is a new sample to be tested, then the distance d_p between the test sample and the p-th hypersphere center is obtained by:

$$
d_p = \sqrt{K\left(x_{new,i},\ x_{new,j}\right) - 2\sum_{i=1}^{n} a_{p,i} K\left(x_{new,i},\ x_{p,i}\right) + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{p,i} a_{p,j} K\left(x_{p,i},\ x_{p,j}\right)}.
\tag{13}
$$

Therefore, if d_p is greater than R_p, then the test sample does not belong to the p-th class.

For multi-SVDD classification issues, it should be noted that due to the intersection of multiple hyperspheres some data samples may have difficulty determining which operating mode the data sample belongs to. Therefore, there are uncertain regions. However, the training data set collected from leak-free historical data, regardless of which hypersphere the data sample in the overlap region belongs to, indicates that the data sample is leak-free one. It is not necessary to determine which SVDD hypersphere the data sample of the overlap region belongs to. For pipeline leak detection, we are concerned with whether the data sample is a leak or a leak-free one, and do not care which type of operating mode the data sample belongs to. Since all SVDD hyperspheres are obtained under different normal operating modes, as long as the test data samples fall in any of the SVDD hyperspheres, this indicates that no leak has occurred. This is the basis for our proposed Cas-SVDD for pipeline leak detection. In our work, the Cas-SVDD is defined that the SVDDs obtained under different operating modes are cascaded one after the other.

After obtaining the SVDD model for each operating mode, Cas-SVDD can be applied with the use of the following strategies:

(1) Before the Cas-SVDD is used, each SVDD should be reasonably ordered. The SVDD corresponding to the operating mode with high probability occurrence should be placed in front of other SVDDs, which can effectively reduce the online detection time;

(2) The data samples collected online are sequentially passed through the Cas-SVDD model. If the sample is included in a certain SVDD hypersphere, it indicates that there is no leak and it is no longer necessary to enter the remaining SVDD hyperspheres. Conversely, if the sample is not included in any SVDD hypersphere, it indicates that the pipeline is leaking.

2.3. Procedure

Based on the KPCA and Cas-SVDD methods, the procedure of the proposed method for pipeline leak detection with multiple operating modes can be divided into an offline part and online part, as shown in Figure 2.
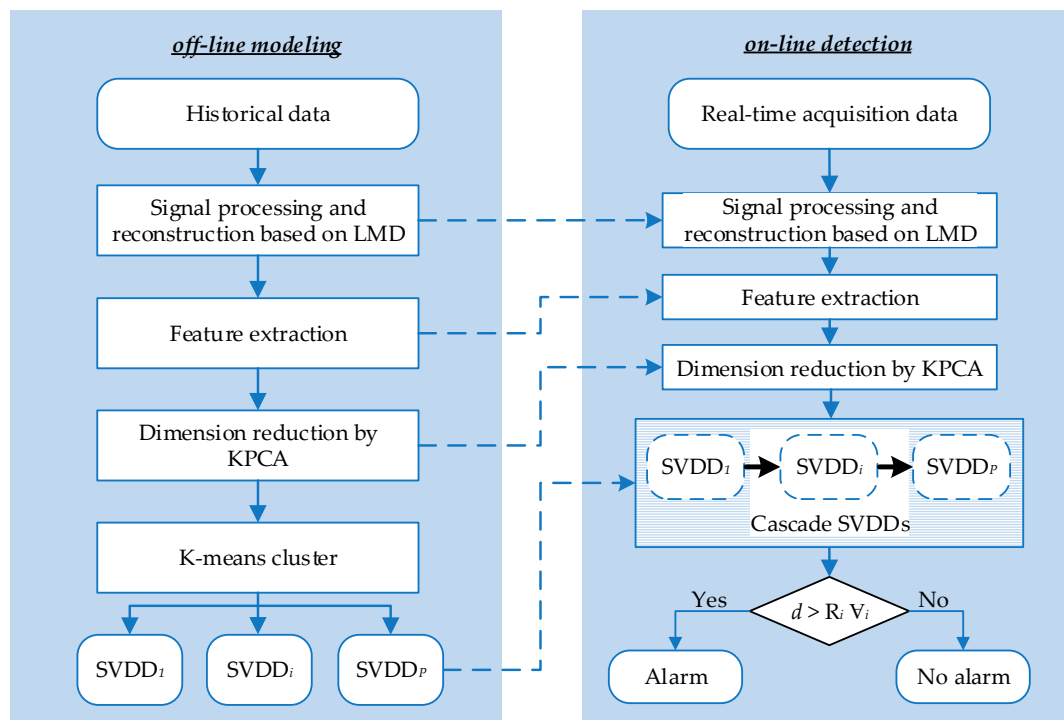

Figure 2. Flowchart of the proposed method.

The detailed procedure was as follows:

I: Offline training model module

(1) Collect historical data of pipeline normal operation, and set initial parameters of LMD, KPCA, K-means algorithm, and SVDD;

(2) Denoise and reconstruct data samples by LMD, and extract the feature variables, each of which can be calculated according to the formulas provided in Table 1. Then, the value of each feature is normalized to the same range (between 0 and 1);

(3) Reduce the Dimension of feature variables by KPCA;

(4) Identify various operating modes by using the K-means algorithm;

(5) Establish the SVDD model for each operating mode, and obtain the center and radius of each hypersphere.

II: Online detection module

(1) Obtain real-time operational data sample;

(2) and (3) are the same as the steps as for the offline module;

(4) Leak alarm, or not, by using Cas-SVDD according to the strategy mentioned above.

It should be noted that the issue of pipeline leak localization is not considered in the paper. For the integrity of the methodology for pipeline leak problem, the most widely used method for pipeline leak location, namely the negative pressure wave method, is briefly introduced here. Once the pipeline leak is determined, the generalized correlation analysis can be used to obtain the time delay estimation of the negative pressure wave generated by the leak signal reaching the sensors on both sides of the leak point. The leak localization can be calculated according to the formula $L_x = (L + v \times \Delta t)/2$, where L_x is the leak point from the upstream reservoir, L is pipeline length, v is the propagation speed of negative

pressure wave, and Δt is the time delay. A detailed description of the leak location method can be found in the literature [48].

3. Case Study

3.1. Data Generation by Flowmaster Software

The pipeline model and pipeline leak scenarios were established by using Flowmaster software, as shown in Figure 3. The length of the pipeline was 2000 m, between an upstream node and a downstream node. The pipeline model parameters were as follows: the inner diameter was 70 mm, the inner wall relative roughness 0.015 mm, the reservoir height of constant head upstream and downstream were 130 m and 0 m, and the negative pressure wave velocity was 1000 m/s, the temperature was 20 degrees Celsius. The position of the leak simulated was 500 m away from the upstream reservoir. Leak ball valves were selected to simulate the pipeline leak. The simulation time was 40 s, the sampling time was 0.01 s. The leak ball valve was opened within 2 s, and the leak occurred at 20 s. For the purpose of better demonstration, only two normal operating modes were considered in our study, one was normal running, where the pipeline was running without any operation adjustment, the other was pipeline valve adjustment. To verify the validity of the proposed method, three leak scenarios were simulated, namely small leak, medium leak, and large leak. Here, small leak was defined as smaller than 1% of the total instantaneous flow within the pipeline, medium leak ranged from 1% of the flow up to 5%, and large leak was larger than 5% of the total flow within the pipeline. Each type of scenario simulated 80 sets of data samples. The pressure signals were collected at node 1 and node 2. To simulate the real signals coming from the pressure sensors, the normally distributed random number was added to the pressure data collected at the nodes.
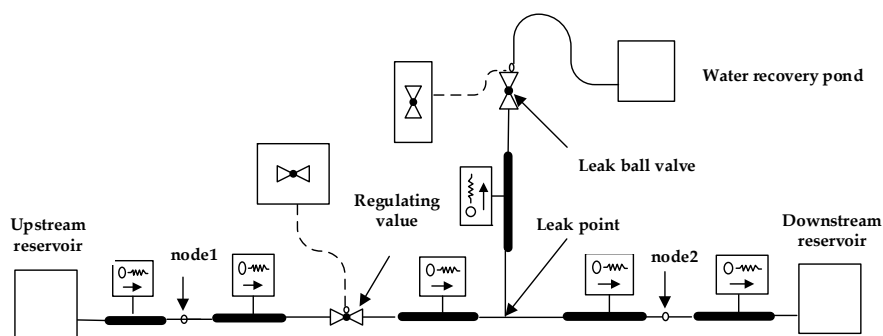

Figure 3. Pipeline modeling by Flowmaster.

3.2. Data Processing and Feature Extraction

The collected pressure signal was used for noise reduction and signal reconstruction by using LMD. The pressure signal of 500 m from the upstream reservoir was used as an example to illustrate the noise reduction effect, as shown in Figure 4. The denoised and reconstructed pressure signal better showed the characteristics of time domain and waveform domain, which provided a basis for the subsequent feature extraction of the pressure variables.

A total of 400 sets of data samples were generated, of which 80 sets were normal running mode and 80 sets were pipeline valve adjustment mode. The data samples for small leak, medium leak, and large leaks were each 80 sets. Table 2 gives an example of a set of extracted feature variables for each mode and leak scenario. It should be noted the value of each feature was normalized to the same range (between 0 and 1) in order to eliminate cross-modal amplitude differences caused by different feature extraction mechanisms.
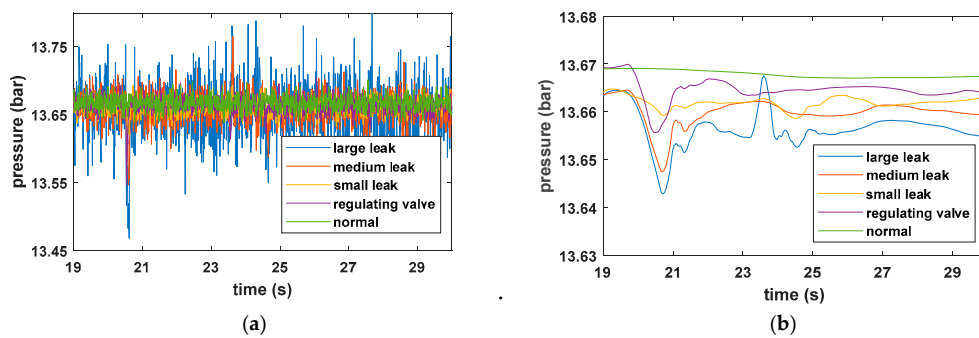
Figure 4. Pressure signals before and after noise reduction. (**a**) Original data samples. (**b**) Reconstructed signals.

Table 2. An example of time-domain features and waveform-domain features under different operating modes and leak scenarios.

| Features	Normal Running	Valve Adjustment	Small Leak	Medium Leak	Large Leak
T1	0.8630	0.9715	0.4743	0.2972	0.0646
T2	0.0201	0.0119	0.0393	0.1705	0.3626
T3	0.8613	0.9709	0.4684	0.2924	0.0619
T4	0.8588	0.9697	0.4605	0.2876	0.0629
T5	0.1169	0.0102	0.5081	0.6859	0.9171
T6	0.9663	0.9869	0.9147	0.7176	0.7101
T7	0.0524	0.0253	0.073	0.2617	0.5905
T8	0.0178	0.0107	0.0384	0.1610	0.3588
T9	0.6369	0.8704	0.4115	0.1295	0.0846
T10	0.8639	0.9718	0.4776	0.300	0.0667
T11	0.0523	0.0253	0.0730	0.2619	0.5911
T12	0.8584	0.9702	0.4622	0.2873	0.0604

It can be seen from Table 2 that the time-domain feature variables and waveform-domain feature variables reflect the characteristic changes of the pressure signal to some extent. Moreover, the changes in certain feature variables are obvious under various operation modes and leak scenarios. However, too many feature variables will greatly increase the computational complexity, which in turn affects the real-time performance of pipeline leak detection. Additionally, some feature variables may have redundancy. Here, the KPCA was used to reduce the feature dimensions, through which the redundant information was removed from the feature variables and nonlinear elements were extracted. The KPCA adopted the Gaussian radial basis function, which has the advantage of fewer parameters and satisfies the Mercer condition, and its kernel width is 20. The first three kernel principal components (KPC) were selected as the new comprehensive feature variables, i.e., KPC 1, KPC 2, and KPC 3. The variance contribution rate was 85% or more, so that the selected kernel principal components could reflect the comprehensive characteristics of the original feature variables. Table 3 gives an example of four sets of kernel principal components data samples in each operating mode and leak scenario. Table 3 shows that, compared with Table 2, the kernel principal components in different operating modes and leak scenarios have a better difference.

Table 3. Kernel principal components (KPC) data samples in different operating modes and leak scenarios.

Mode	KPC 1	KPC 2	KPC 3
normal running	0.11587706	−0.10176848	0.00542158
	0.11507034	−0.10450996	0.00801958
	0.11585059	−0.10521875	0.00449056
	0.11120097	−0.10419620	0.00683419
valve adjustment	0.13148375	−0.09658969	0.00207181
	0.12462393	−0.10080797	0.00200395
	0.12116591	−0.09986308	0.00152137
	0.11844755	−0.09243056	0.00294898
small leak	0.05001575	−0.15987215	−0.00856215
	0.03419228	−0.14363320	0.00159545
	0.04376509	−0.15863410	−0.00645285
	0.04046785	−0.14860605	0.00441316
medium leak	−0.01347718	−0.14295158	0.00504162
	−0.01062501	−0.13432576	0.00606295
	−0.01134602	−0.14564260	−0.02915774
	−0.01123048	−0.15335873	−0.01754892
large leak	−0.10735391	−0.10583311	0.04107728
	−0.12025648	−0.08014315	−0.06043058
	−0.12778299	−0.07435172	−0.01165562
	−0.10255444	−0.11753984	0.02651577

4. Results and Discussions

As mentioned above, during the pipeline fluid transport process, the operating modes, such as valve adjustment and normal running, change frequently, and the collected leak-free data samples contain multiple operating conditions. The single SVDD hypersphere is not compact enough in a variety of operating modes, which in turn leads to low classification accuracy. The K-means algorithm was used to cluster data samples that were processed by KPCA, and then the data samples of each operating mode could be obtained. The clustering result using K-means is shown in Figure 5. The K-means clustering obtained a good clustering result. Next, the SVDD model was established for each operating mode in order to obtain the Cas-SVDD model, consisting of multiple compact SVDD hyperspheres. It can also be seen from Figure 5 that establishing a unique SVDD hypersphere for all the data samples in all normal operating modes will result in a data sample set that is not compact enough. In some cases, a leak data sample may be located between the SVDD hyperspheres, and the unique SVDD may consider it as a normal operating data sample, while the Cas-SVDD makes it easy to identify that the data sample is a pipeline leak sample.
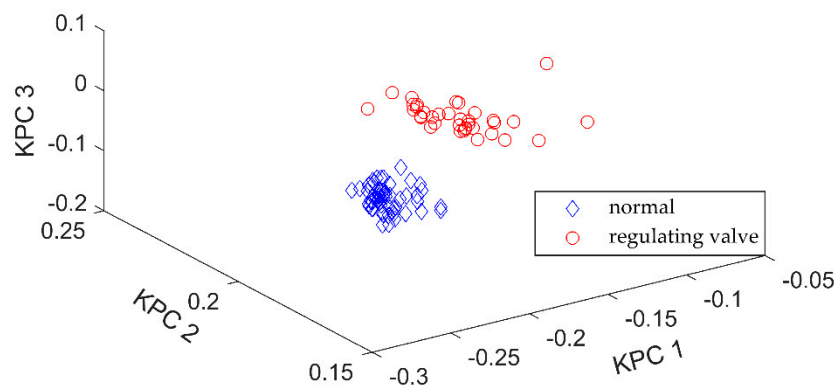
Figure 5. The clustering result using K-means.

To verify the proposed KPCA-Cas-SVDD method, the performance of the pipeline leak detection will be compared to the corresponding data from the single SVDD (S-SVDD) and Cas-SVDD methods. Here, the S-SVDD method means that only one hypersphere model was established using all data samples with different operating modes under normal operation; the difference between the Cas-SVDD and the KPCA-Cas-SVDD is that the former does not use the KPCA for the dimension reduction of feature variables, while the latter uses the KPCA. The ratio of training data and test data was 5:5. The SVDD adopted a Gaussian kernel function. After 5-fold cross-validation and grid search, the optimal parameter kernel width σ and target class error rate w of the SVDD in the three methods were obtained, as shown in Table 4.

Table 4. Optimal parameters of SVDD.

	S-SVDD	Cas-SVDD		KPCA-Cas-SVDD	
		SVDD1	**SVDD2**	**SVDD1**	**SVDD2**
σ	3.8075	5.1248	2.8289	2.8289	7.7426
w	0.5412	0.2154	0.2154	0.0464	0.2154

For the convenience of comparing the performance of the three methods, three cases were considered and the three methods were applied to each of these three cases. Case 1: including 120 sets of data samples, of which 40 were from normal running, 40 from pipeline valve adjustment, and 40 from small leak; Case 2: including 120 sets of data samples, of which 40 were from normal running, 40 from pipeline valve adjustment, and 40 from medium leak; Case 3: including 120 sets of data samples, of which 40 were from normal running, 40 from pipeline valve adjustment, and 40 from large leak.

For the S-SVDD method, the pipeline leak detection results are shown in Figure 6, where Figure 6a–c represent Case 1, Case 2, and Case 3, respectively. For the convenience of graphic demonstration, the first 80 data samples are leak-free data samples, and the leak data sets are set from the 81st to the 120th samples. It can be seen from Figure 6a that 15 small leak samples were not detected. Figure 6b shows that 12 medium leak samples were undetected. Figure 6c shows that seven large leak samples were undetected. Therefore, these results show that the performance of the S-SVDD method for pipeline leak detection is poor, with low leak detection accuracy.
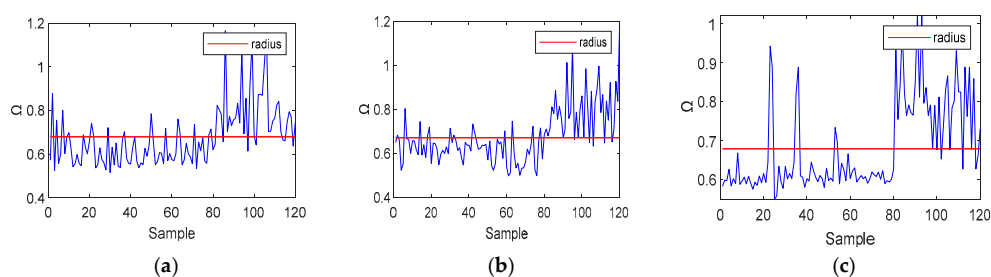

Figure 6. The results of the S-SVDD method for pipeline leak detection. (**a**) small leak; (**b**) medium leak; (**c**) large leak.

For the Cas-SVDD method, the pipeline leak detection results are shown in Figure 7. The 120 sets of data samples were first classified by using the first SVDD of the Cas-SVDD. Then, the remaining data samples that were not in the first SVDD hypersphere were further classified by using the second SVDD of the Cas-SVDD. Figure 7(a1) shows that there were eight samples with false-positive results, and Figure 7(a2) shows 10 samples with false-negative results. In summary, these results from Figure 7 show that the performance of Cas-SVDD for pipeline leak detection was much better than that of S-SVDD.
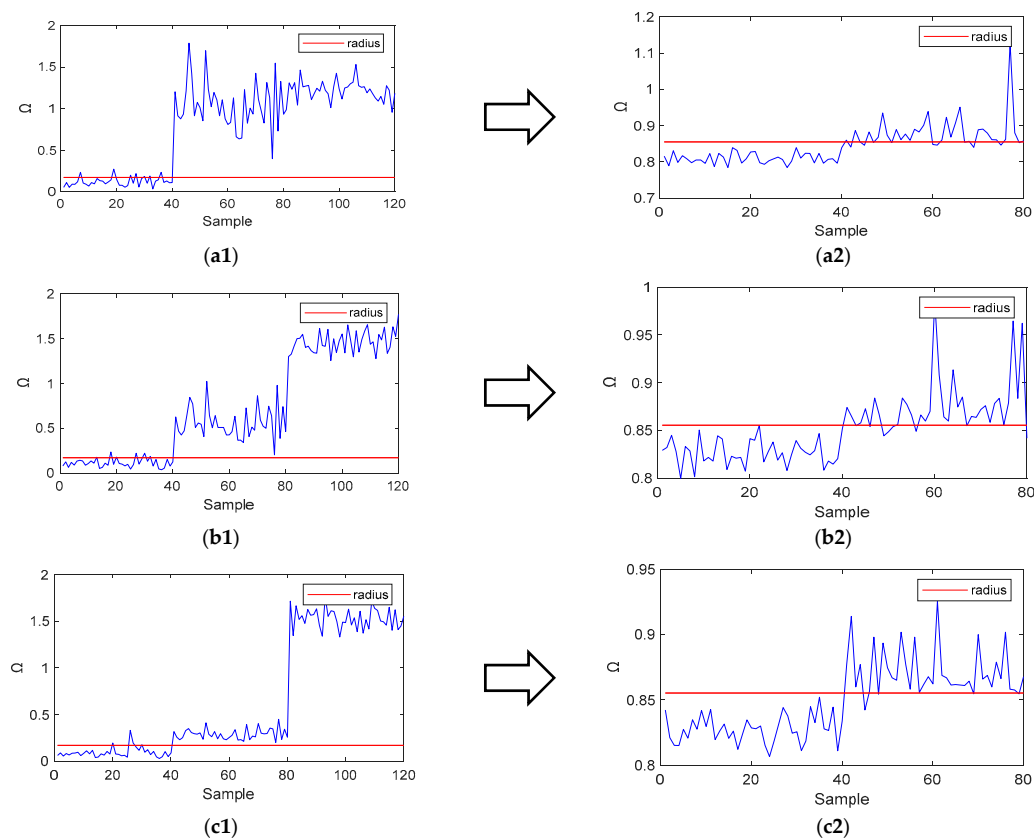
Figure 7. The results of the Cas-SVDD method for pipeline leak detection. (**a1**) Cas-SVDD$_1$ (small leak); (**a2**) Cas-SVDD$_2$ (small leak); (**b1**) Cas-SVDD$_1$ (medium leak); (**b2**) Cas-SVDD$_2$ (medium leak); (**c1**) Cas-SVDD$_1$ (large leak); (**c2**) Cas-SVDD$_2$ (large leak).

As shown in Figure 8, the KPCA-Cas-SVDD method had very few false-positive results and false-positive results. This also shows that the feature dimension reduction using KPCA has a great influence on the leak detection performance for the Cas-SVDD. This is because the KPCA algorithm has the ability to process high-dimensional projections for nonlinear data, and can obtain new comprehensive features that contain most of the information of the original feature variables. Table 5 shows the results of the comparison of the three methods for pipeline leak detection. It also shows that the pressure at the location of the leak has a great influence on the accuracy of the pipeline leak detection. The high pressure at the location of the leak indicates that a large leak has occurred. Large leaks are easily detected compared to small leaks, is because a large leak produces a pressure change of higher magnitude that travels faster to the upstream and downstream pressure sensors, causing obvious changes in the feature variables. However, for small leaks, the pressure change at the location of the leak is less obvious and the variation in the feature variables is also small. In this case, the proposed method showed a high detection accuracy compared to the other two methods.
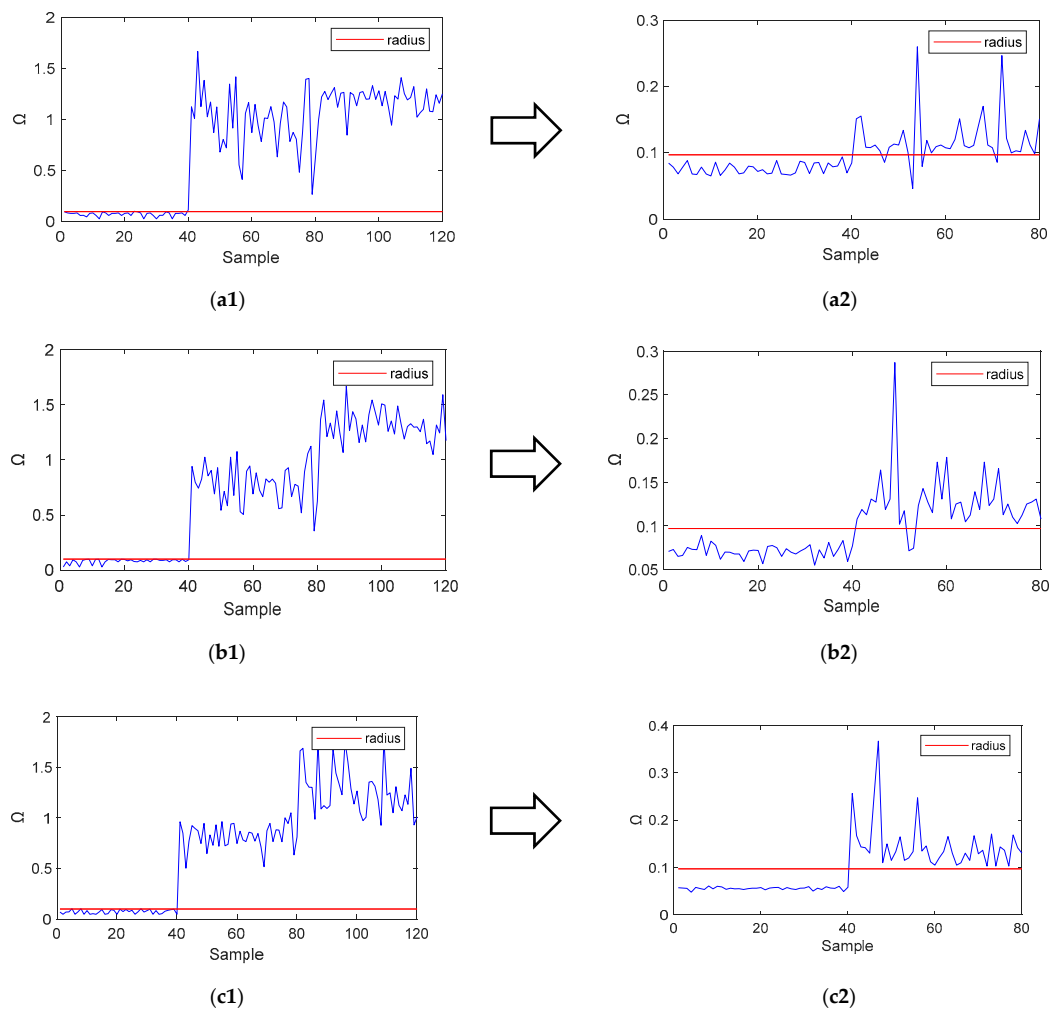
Figure 8. The results of the KPCA-Cas-SVDD method for pipeline leak detection. (**a1**) KPCA-Cas-SVDD$_1$ (small leak); (**a2**) KPCA-Cas-SVDD$_2$ (small leak); (**b1**) KPCA-Cas-SVDD$_1$ (medium leak); (**b2**) KPCA-Cas-SVDD$_2$ (medium leak); (**c1**) KPCA-Cas-SVDD$_1$ (large leak); (**c2**) KPCA-Cas-SVDD$_2$ (large leak).

Table 5. A comparison of the results of detection accuracy.

	Small Leak	**Medium Leak**	**Large Leak**
S-SVDD	62.5%	70%	82.5%
Cas-SVDD	75%	80%	90%
KPCA-Cas-SVDD	90%	95%	97.5%

5. Conclusions

In this paper, a hybrid intelligent method for pipeline leak detection was proposed. The method firstly performed signal denoising and signal reconstruction based on LMD. After that, the KPCA was used for feature dimension reduction. Subsequently, the K-means algorithm was used for the clustering of various operating modes. Finally, the Cas-SVDD was used for pipeline leak detection. The characteristics of this paper are mainly two aspects: (1) the proposed method is only necessary to collect the pressure signal of pipeline normal operation, without the need to collect the leak data samples that are difficult to obtain in the actual pipeline operation; (2) based on the integration of KPCA and Cas-SVDD, a novel method was proposed for pipeline leak detection with multiple operating modes. Compared with the S-SVDD method and Cas-SVDD method, the proposed method in this

paper comprehensively considered the various operating modes in the pipeline transportation process, and can effectively reduce the false alarm rate.

However, only two normal operating modes were considered in our work. There may be more operating modes during the actual pipeline operation. In this case, more SVDD models would need to be established, furthermore, the order of each SVDD in the cascade structure should be optimized, which can reduce the time of online detection and improve the accuracy of leak detection. Therefore, how to optimize the order of Cas-SVDD to minimize the time of online detection while ensuring the accuracy of leak detection is an issue worthy of further study. In addition, the better clustering algorithm also contributes to the construction of the Cas-SVDD and improves the accuracy of pipeline leak detection. Finally, some practical issues of the proposed method should be considered for leak detection in a real case. Future work will focus on these issues.

Author Contributions: Conceptualization, M.Z., X.S. and H.P.; methodology, M.Z. and Q.Z.; software, Y.L.; formal analysis, X.S.; data curation, Q.Z.; writing—original draft preparation, M.Z. and Q.Z.; writing—review and editing, H.P.; supervision, X.S.; resources, Y.C.; funding acquisition, H.P.

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

References

1. Yalçin, B.C.; Demir, C.; Gökçe, M.; Koyun, A. Water leakage detection for complex pipe systems using hybrid learning algorithm based on ANFIS Method. *J. Comput. Inf. Sci. Eng.* **2018**, *18*, 041004. [CrossRef]
2. Arifin, B.M.S.; Li, Z.; Shah, S.L.; Meyer, G.A.; Colin, A. A novel data-driven leak detection and localization algorithm using the Kantorovich distance. *Comput. Chem. Eng.* **2018**, *108*, 300–313. [CrossRef]
3. Fayaz, M.; Ahmad, S.; Hang, L.; Kim, D. Water supply pipeline risk index assessment based on cohesive hierarchical fuzzy inference system. *Processes* **2019**, *7*, 182. [CrossRef]
4. Datta, S.; Sarkar, S. A review on different pipeline fault detection methods. *J. Loss Prev. Proc.* **2016**, *41*, 97–106. [CrossRef]
5. Adedeji, K.B.; Hamam, Y.; Abe, B.T.; Abu-Mahfouz, A.M. Towards achieving a reliable leakage detection and localization algorithm for application in water piping networks: An overview. *IEEE Access* **2017**, *5*, 20272–20285. [CrossRef]
6. Darsana PVarija, K. Leakage detection studies for water supply systems—A review. In *Water Resources Management*; Singh, V.P., Yadav, S., Yadava, R.N., Eds.; Water Science and Technology Library: Singapore, 2018; Volume 78, pp. 141–150.
7. Adegboye, M.A.; Fung, W.K.; Karnik, A. Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches. *Sensors* **2019**, *19*, 2548. [CrossRef] [PubMed]
8. Wu, Y.; Liu, S. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* **2017**, *14*, 972–983. [CrossRef]
9. Yazdekhasti, S.; Piratla, K.R.; Matthews, J.C.; Khan, A.; Atamturktur, S. Optimal selection of acoustic leak detection techniques for water pipelines using multi-criteria decision analysis. *Manag. Environ. Qual.* **2018**, *29*, 255–277. [CrossRef]
10. Wang, W.; Li, Z.; Jing, L.; Lee, P.; Murch, R. A straightforward method for estimating the size of leaks in water pipelines using acoustic transients. *J. Acoust. Soc. Am.* **2018**, *144*, EL404–EL409. [CrossRef]
11. Li, S.Z.; Song, Y.J.; Zhou, G.Q. Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition. *Measurement* **2018**, *115*, 39–44. [CrossRef]
12. Huang, Y.; Wang, Q.; Shi, L.L.; Yang, Q.H. Underwater gas pipeline leakage source localization by distributed fiber-optic sensing based on particle swarm optimization tuning of the support vector machine. *Appl. Opt.* **2016**, *55*, 242–247. [CrossRef] [PubMed]
13. Jia, Z.G.; Ren, L.; Li, H.N.; Jiang, T.; Wu, W.L. Pipeline leakage identification and localization based on the fiber Bragg grating hoop strain measurements and particle swarm optimization and support vector machine. *Struct. Control Health Monit.* **2019**, *26*, e2290. [CrossRef]

14. Zhang, S.; Liu, B.; He, J. Pipeline deformation monitoring using distributed fiber optical sensor. *Measurement* **2019**, *133*, 208–213. [CrossRef]

15. Zhang, H.; Liang, Y.; Zhang, W.; Xu, N.; Guo, Z.; Wu, G. Improved PSO-based method for leak detection and localization in liquid pipelines. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3143–3154. [CrossRef]

16. Smith, J.; Chae, J.; Hugo, R.; Learn, S.; Park, S. Pipeline rupture detection using real-time transient modelling and convolutional neural networks. In Proceedings of the 2018 12th International Pipeline Conference (IPC 2018), Calgary, Canada, 24–28 September 2018; American Society of Mechanical Engineers (ASME): New York, NY, USA, 2018; Volume 3.

17. Alawadhi, A.; Boso, F.; Tartakovsky, D.M. Method of distributions for water hammer equations with uncertain parameters. *Water Resour. Res.* **2018**, *54*, 9398–9411. [CrossRef]

18. Zhang, C.; Zecchin, A.C.; Lambert, M.F.; Gong, J.Z.; Simpson, A.R. Multi-stage parameter-constraining inverse transient analysis for pipeline condition assessment. *J. Hydroinform.* **2018**, *20*, 281–300. [CrossRef]

19. Li, J.; Zheng, Q.; Qian, Z.; Yang, X. A novel location algorithm for pipeline leakage based on the attenuation of negative pressure wave. *Process Saf. Environ. Prot.* **2019**, *123*, 309–316. [CrossRef]

20. Ge, C.; Wang, G.; Ye, H. Analysis of the smallest detectable leakage flow rate of negative pressure wave-based leak detection systems for liquid pipelines. *Chem. Eng. Sci.* **2008**, *32*, 1669–1680. [CrossRef]

21. Murvay, P.-S.; Silea, I. A survey on gas leak detection and localization techniques. *J. Loss Prev. Proc.* **2012**, *25*, 966–973. [CrossRef]

22. Saqib, N.U.; Mysorewala, M.F.; Cheded, L. A multiscale approach to leak detection and localization in water pipeline network. *Water Resour. Manag.* **2017**, *31*, 3829–3842. [CrossRef]

23. Lang, X.; Li, P.; Cao, J.; Li, Y.; Ren, H. A small leak localization method for oil pipelines based on information fusion. *IEEE Sens. J.* **2018**, *18*, 6115–6122. [CrossRef]

24. Zhang, T.; Tan, Y.; Zhang, X.; Zhao, J. A novel hybrid technique for leak detection and location in straight pipelines. *J. Loss Prev. Proc.* **2015**, *35*, 157–168. [CrossRef]

25. Chan, T.K.; Chin, C.S.; Zhong, X.H. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access* **2018**, *6*, 78846–78867. [CrossRef]

26. Lu, Z.; She, Y.T.; Loewen, M. A sensitivity analysis of a computer model-based leak detection system for oil pipelines. *Energies* **2017**, *10*, 1226. [CrossRef]

27. Waleed, D.; Mustafa, S.H.; Mukhopadhyay, S.; Abdel-Hafez, M.F.; Jaradat, M.A.K.; Dias, K.R.; Arif, F.; Ahmed, J.I. An in-pipe leak detection robot with a neural-network based leak verification system. *IEEE Sens. J.* **2019**, *19*, 1153–1165. [CrossRef]

28. Kang, J.; Park, Y.J.; Lee, J.; Wang, S.H.; Eom, D.S. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4279–4289. [CrossRef]

29. Pan, S.; Xu, Z.; Li, D.; Lu, D. Research on detection and location of fluid-filled pipeline leakage based on acoustic emission technology. *Sensors* **2018**, *18*, 3628. [CrossRef]

30. Li, Z.L.; Zhang, H.F.; Tan, D.J.; Chen, X.; Lei, H.X. A novel acoustic emission detection module for leakage recognition in a gas pipeline valve. *Process Saf. Environ. Prot.* **2017**, *105*, 32–40. [CrossRef]

31. Ahn, B.; Kim, J.; Choi, B. Artificial intelligence-based machine learning considering flow and temperature of the pipeline for leak early detection using acoustic emission. *Eng. Fract. Mech.* **2019**, *210*, 381–392. [CrossRef]

32. Wang, F.; Lin, W.G.; Liu, Z.; Wu, S.C.; Qiu, X.B. Pipeline leak detection by using time-domain statistical features. *IEEE Sens. J.* **2017**, *17*, 6431–6442. [CrossRef]

33. Zhou, Z.J.; Hu, C.H.; Xu, D.L.; Yang, J.B.; Zhou, D.H. Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection. *Expert Syst. Appl.* **2011**, *38*, 3937–3943. [CrossRef]

34. El-Zahab, S.; Abdelkader, E.M.; Zayed, T. An accelerometer-based leak detection system. *Mech. Syst. Signal Process.* **2018**, *108*, 58–72. [CrossRef]

35. Ni, L.; Jiang, J.; Pan, Y. Leak location of pipelines based on transient model and PSO-SVM. *J. Loss Prev. Proc.* **2013**, *26*, 1085–1093. [CrossRef]

36. Wang, X.Y.; Chen, Z.G.; Zhong, X.R.; Inc, D.E.P. Research on leak detection of water pipeline base on PSO-SVM. In Proceedings of the 2016 International Conference on Applied Mechanics, Electronics and Mechatronics Engineering, Beijing, China, 28–29 May 2016; Destech Publications, Inc.: Lancaster, UK, 2016; pp. 227–233.

37. Mandal, S.K.; Chan, F.T.S.; Tiwari, M.K. Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM. *Expert Syst. Appl.* **2012**, *39*, 3071–3080. [CrossRef]

38. Geiger, G.; Werner, T.; Matko, D. Knowledge-based leak monitoring for pipelines. *IFAC Proc. Vol.* **2001**, *34*, 249–254. [CrossRef]

39. Park, C.; Looney, D.; van Hulle, M.M.; Mandic, D.P. The complex local mean decomposition. *Neurocomputing* **2011**, *74*, 867–875. [CrossRef]

40. Smith, J.S. The local mean decomposition and its application to EEG perception data. *J. R. Soc. Interface* **2005**, *2*, 443–454. [CrossRef]

41. Lee, J.M.; Yoo, C.K.; Choi, S.W.; Vanrolleghem, P.A.; Lee, I.B. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2004**, *59*, 223–234. [CrossRef]

42. Zeng, L.; Long, W.; Li, Y.Y. A novel method for gas turbine condition monitoring based on KPCA and analysis of statistics T^2 and SPE. *Processes* **2019**, *7*, 124. [CrossRef]

43. Tax, D.M.J.; Duin, R.P.W. Support vector domain description. *Pattern Recognit. Lett.* **1999**, *20*, 1191–1199. [CrossRef]

44. Tax, D.M.J.; Duin, R.P.W. Support vector data description. *Mach. Learn.* **2004**, *54*, 45–66. [CrossRef]

45. Zhu, X.J.; Jin, X.N.; Jia, D.D.; Sun, N.W.; Wang, P. Application of data mining in an intelligent early warning system for rock bursts. *Processes* **2019**, *7*, 55. [CrossRef]

46. Qi, Y.; Lin, W.; Wu, H. A leak detection method for natural gas pipelines based on time-domain statistical features. *Acta Pet. Sin.* **2013**, *34*, 1195–1199.

47. Hoffmann, H. Kernel PCA for novelty detection. *Pattern Recognit.* **2007**, *40*, 863–874. [CrossRef]

48. 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]