Article

A Reliable Acoustic EMISSION Based Technique for the Detection of a Small Leak in a Pipeline System

Thang Bui Quy , Sohaib Muhammad and Jong-Myon Kim *

School of Computer Engineering, University of Ulsan, Ulsan 44610, Korea; bqthangcndt@gmail.com (T.B.Q.); md.sohaibdurrani@gmail.com (S.M.)

* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Received: 10 March 2019; Accepted: 15 April 2019; Published: 18 April 2019

Abstract: This paper proposes a reliable leak detection method for water pipelines under different operating conditions. This approach segments acoustic emission (AE) signals into short frames based on the Hanning window, with an overlap of 50%. After segmentation from each frame, an intermediate quantity, which contains the symptoms of a leak and keeps its characteristic adequately stable even when the environmental conditions change, is calculated. Finally, a k-nearest neighbor (KNN) classifier is trained using features extracted from the transformed signals to identify leaks in the pipeline. Experiments are conducted under different conditions to confirm the effectiveness of the proposed method. The results of the study indicate that this method offers better quality and more reliability than using features extracted directly from the AE signals to train the KNN classifier. Moreover, the proposed method requires less training data than existing techniques. The transformation method is highly accurate and works well even when only a small amount of data is used to train the classifier, whereas the direct AE-based method returns misclassifications in some cases. In addition, robustness is also tested by adding Gaussian noise to the AE signals. The proposed method is more resistant to noise than the direct AE-based method.

Keywords: acoustic emissions; k-nearest neighbor; leak detection; pipeline diagnostics; reliability

1. Introduction

Water pipeline systems are required to accomplish commercial and domestic activities. A leak in such a system can cause financial loss, waste of resources, and even human deaths due to the collapse of the system. Thus, numerous approaches have been introduced to diagnose pipeline leaks. The approaches include passive and active systems [1], hardware-based and software-based methods [2]. Previous works [3,4] also provided an overview of research on leak detection. Among those approaches, acoustic emission (AE)-based techniques, which are passive and hardware-based, are promising because AE sensors can detect small leaks quickly, offering high sensitivity regarding fault growth in a pipeline system. Thus, many researchers have exploited AE-based techniques for pipeline diagnostics [5–11].

The first stage of the diagnosis process is to spot an existing fault. Because of the complexity of AE activity, it is difficult to establish a leak detection model directly from mathematical equations; therefore, different methods have been proposed to train the model using recorded datasets [6,7]. Although these techniques offer high accuracy, they could be inefficient in diverse circumstances because the classifiers are trained using features extracted directly from AE signals. Thus, it makes the classifiers dependent on the absolute signal levels that are influenced by the flow rate and pressure in a pipeline system [12,13]. Moreover, AE signals also vary with temperature, which is uncertain in industries. As a result, AE measurements recorded under such uncertain working conditions have different signal levels, and thus, a fault diagnosis model based on the absolute amplitude values of the

signals is unreliable. In [14], the authors presented a study that investigates the inherent properties of the vibro-acoustic signals instead of relative amplitude values. It delivers useful insights and results for water leak detection, but the study focused on pipelines buried under soil less influenced by external factors while AE signals from pipelines above soil in factories are prone to noise. This issue can be addressed by acquiring many training datasets under different working conditions to provide enough information for the classifier. However, it is not an optimal solution because the design of experiments can become immensely expensive or even prohibitively complicated. Therefore, this study extracts features from transformed signals independent of an absolute level instead of directly from AE signals. The obtained classifier can detect small leaks in a pipeline with high accuracy and reliability.

For the experiments, this study uses a water pipeline system deployed in a laboratory. Three AE sensors are mounted on different parts of the system with the known distances among the sensors. A pinhole is drilled through the pipe wall to simulate leakage of the pipeline and a valve is attached to the wall to control the flow of water through the leak. When the valve is closed, and system is working in a normal state (no leakage) then there is no variation in the levels of the recorded signals. If a small leak appears between sensors 2 and 3, an imbalance is created between AE channels 1 and 2 by reporting dissimilar amounts of noise from the leak. Since sensor 2 is nearer the leak than sensor 1, the leak noise on channel 2 is greater than that on channel 1, as explained by the theory of wave attenuation [15–17]. Thus, this uneven behavior of the signals can be used to detect small leaks in a pipeline system. The leak-signal-to-normal-noise ratio and the attenuation law of wave propagation can be used to detect the leak; these characteristics are demonstrated in Section 3.

To demonstrate the advantages of the balance/imbalance-based approach, this paper uses the theory of Kullback-Leibler distance [18] to measure the separability between the two classes (i.e., NORMAL and ABNORMAL) in different experiments. Furthermore, the paper applies a k-nearest neighbor (KNN) model to classify between normal and leak. The KNN classifier is trained by two approaches, one using features extracted directly from the AE signals and the other using features extracted from the transformed signals. The KNN classifier in both the approaches are trained on the data of one working condition and tested using data of other conditions. In addition, Gaussian noise is added to the AE measurements to evaluate the robustness of the classifiers trained by the two approaches.

2. Experimental Setup

2.1. Data Acquisition

Figures 1 and 2 show a setup of the AE signals acquisition from a water pipeline system. The testbed installation and the sensors configuration parameters are listed in Table 1.

Table 1. Experimental parameters.

No	Quantity	Detail
1	Distance from Sensor 2 to Sensor 3 (D)	2000 (mm)
2	Distance from Sensor 2 to Sensor 1 (d)	500 (mm)
3	Distance from Sensor 2 to Leak (d2)	900 (mm)
4	Distance from Sensor 3 to Leak (d3)	1100 (mm)
5	Thickness of pipeline	6.02 (mm)
6	Outer diameter of pipeline	114.3 (mm)
7	Material of pipeline	Stainless steel 304

The experiments are based on four leaks with different diameters, i.e., 2.0, 1.0, 0.5, and 0.3 mm (mm), abbreviated as L1, L2, L3, and L4, respectively. The water flow is controlled using pressures of 7, 13, and 18 bar, given as P1, P2, and P3, respectively. The experiments are conducted under a stable temperature of approximately 29 °C.

Figure 1. Data acquisition setup.

Figure 2. The testbed for pipeline leak detection.

2.2. AE Sensors

To acquire AE signals, R15I-AST sensors are used to provide high sensitivity during the data recording and the recorded signals tend to be free of low-frequency components. The sensors' characteristics are summarized in Table 2.

Table 2. Specifications of R15I-AST.

No	Parameter	Value
1	Peak sensitivity, ref (V/(m/s))	109 (dB)
2	Peak sensitivity, ref (V/ μ bar)	−22 (dB)
3	Operating frequency range	50–400 (kHz)
4	Resonant frequency, ref (V/(m/s))	75 (kHz)
6	Resonant frequency, ref (V/ μ bar)	150 (kHz)
7	Directionality	+/−1.5 (dB)
8	Temperature range	−35 to 75 ($^{\circ}$ C)

2.3. Data Record

The normal and abnormal states of the system refer to the closed and open positions of the valve installed on the leak, respectively. The data for every pair of (P, L) combinations are recorded for 2 min with a sampling frequency of 1 megahertz (MHz) after the water flow are stable. Thus, there are a total of 72 datasets for the three signal channels in the experimental conditions (3 channels $\times$ 3 pressures $\times$ 4 leaks $\times$ 2 classes). Figure 3 presents mode setting of water flow. In the first stage, the pump is turned on, the leak is deactivated (closed) and the pressure is controlled around 7 Bar. Then, AE signals are recorded in 2 min. In the second stage, the leak is activated (open). The acquisition device waits for the flow stabilization and records AE signals in 2 min. The process continues, as shown in Figure 3. Figure 4 shows the AE signals from channel 2 for leak L3 under three different flow pressures.

Figure 3. Input flow rate experimented with leak L3.

Figure 4. Acoustic emission (AE) signals from channel 2 for leak L3.

3. Small Leak Detection Methodology

The mathematical modeling of the leaks and their symptoms is based on the testbed, as shown in Figure 1.

3.1. Symptoms of Leak Presence

There exists AE activity in pipelines, even when they are operating in a healthy state. This activity could be due to the mechanical sources, such as pumps or particles in the flow, hitting a pipeline wall, or hydraulic sources caused by pressure pulses at vortices in the fluid inside the pipeline [12]. Let assume that n_i ($i = 1, 2$) are AE signals acquired by sensors 1, 2 in the normal state, and their mean and variance are 0 and $N_1 \approx N_2 \approx N > 0$, respectively. When a small leak occurs in the testbed pipeline, it makes a small disturbance in the flow around the leak, which introduces a new AE source into the system. The previous source n_i can be deemed as background noise; suppose that this noise is uncorrelated with the source of the signal obtained from the leak. Therefore, the model for AE measurements from the sensors for this scenario can be given as: $z_i = s_i + n_i$ ($i = 1, 2$), where s_i is the leak AE signal received by sensor i . If the variance of z_i and s_i is Z_i and S_i , respectively, then the uncorrelation between n_i and s_i , $Z_i = S_i + N$ can be explored by setting $g = Z_2/Z_1$ and it is transformed as follows:

$$g = \frac{Z_2}{Z_1} = \frac{S_2 + N}{S_1 + N} = \frac{S_2/N + 1}{S_1/N + 1}. \quad (1)$$

If the background noise n_i ($i = 1, 2$) is bandlimited white noise, then its variance is always N over its entire frequency range. Next, consider the measurement model when the measurement consists of the background noise n_i and the leak signal s_i ($i = 1, 2$) at a frequency ω . In fact, the spectrum of the

leak noise is a broadband range of frequencies; however, all the components demonstrate an identical behavior toward the leak phenomenon.

The AE attenuation characteristic from a power law [17], S_1 and S_2 are related by:

$$S_1 = S_2 e^{-\alpha d}. \quad (2)$$

In (2), $\alpha = \alpha(\omega)$ is the attenuation coefficient in wave propagation which is dependent on frequency, d is the distance between sensors 1 and 2, and S_i ($i = 1, 2$) is the leak signal with frequency ω .

By substituting (2) into (1), and abbreviating $r = S_2/N$, which is the signal to noise ratio measured by sensor 2 at frequency ω , and symbolizing $\beta = e^{-\alpha d}$, (1) transforms as follow:

$$g = \frac{r+1}{\beta r+1}. \quad (3)$$

Next, take the partial derivative of g according to r :

$$\frac{\partial g}{\partial r} = \frac{1-\beta}{(\beta r+1)^2}. \quad (4)$$

Since $0 < \beta < 1$ for $0 < d < \infty$, then $\partial g / \partial r > 0$. As a result, $g(r)$ is a monotonically increasing function according to r . If $r_1 \neq r_2$, then $g(r_1) \neq g(r_2)$. Naturally, a normal state has $r = 0$ at every frequency ω , and an abnormal state always has $r \neq 0$; thus, the function $g(r)$ is applicable for leak detection. Figure 5 shows the dependence of $g(r)$ on r with different β values at a particular frequency. It can be easily observed that all the curves $g(r)$ increase from the normal state when r increases.

Figure 5. The dependence of $g(r)$ on r with different β at frequency ω .

3.2. Robustness of $g(r)$ in Leak Manifestation

This section investigates the reliability of leak manifestation using the function $g(r)$ when the level of background noise increases. Gaussian noise is added to the signals to emulate the presence of noise. Suppose that an amount of noise Δn with the mean 0 and variance ΔN is added to the noise background while the leak signal remains the same. At this moment, the background noise n is replaced by $n' = n + \Delta n$; its variance is $N' = N + \Delta N$, and r is replaced by $r' = S_2/N'$.

Setting $\gamma = \Delta N/N$, produces $r' = r/(1 + \gamma)$, which is a function of two variables (r, γ). Consider the partial derivative of r' with respect to γ :

$$\frac{\partial r'}{\partial \gamma} = -\frac{r}{(1 + \gamma)^2}. \quad (5)$$

Now, the function g is replaced by:

$$g' = g(r') = \frac{r + 1 + \gamma}{\beta r + 1 + \gamma} \quad (6)$$

and its partial derivative according to γ is given by:

$$\frac{\partial g'}{\partial \gamma} = \frac{(\beta - 1)r}{(\beta r + 1 + \gamma)^2}. \quad (7)$$

If $0 < \beta < 1$ and $\partial g' / \partial \gamma < 0$, then g' decreases when γ increases. In other words, the discrimination quality of g' becomes poorer as the intensity of the background noise is higher. This characteristic is similar to r' ; however, g' is more reliable than r' because the decline in g' is smaller than the decline in r' .

Shortening $\left| \frac{\partial r'}{\partial \gamma} \right| = \Delta_1$, $\left| \frac{\partial g'}{\partial \gamma} \right| = \Delta_2$, and taking the proportion $\frac{\Delta_2}{\Delta_1}$ produces:

$$\frac{\Delta_2}{\Delta_1} = (1 - \beta) \frac{(1 + \gamma)^2}{(\beta r + 1 + \gamma)^2} \quad (8)$$

Obviously, if the parameter β in (10) is selected suitably, then $\frac{\Delta_2}{\Delta_1} \ll 1 \quad \forall r, \gamma$. As a result, $g(r')$ varies more slowly than r' if γ is increasing. It turns out that if the noise background increases to some extent, the variable r' exceeds the limitation of leak discrimination, whereas the function $g(r')$ still provides enough differentiation.

In (8), if β approaches 0, the ratio Δ_2/Δ_1 converges 1 and the variation of the function g is similar to r if the background noise changes, the characteristic of g is no longer robust. In contrast, (3) reveals that if β approaches 1, g converges 1 for any r . This means that the function g does not manifest any abnormal state of the system. Hence, the parameter β should be chosen optimally to trade off between the two above cases so that both high sensitivity and reliability can be achieved.

3.3. Detection Procedures

This paper proposes an algorithm to detect a small leak in a pipeline based on $g(r)$ function because it can indicate the presence of leak as described in the previous section. Figure 6 shows a generic framework for leak detection using two approaches: direct AE-based and $g(r)$ -based.

Figure 6. Leak detection procedures.

The AE signals from sensors 1, 2, and 3 are the inputs to the leak detection framework. In the direct AE-based method, there is no $g(r)$ -construction block. After dividing the AE signals into frames, they are provided to the feature extraction block. The feature extraction process is carried out after the completion of $g(r)$ -construction process.

3.3.1. Frame Division

The recorded AE signals are segmented into a series of frames by the frame division block. The AE waves propagate through different distances from the leak to the sensors. Thus, their arrivals are lagged. It indicates that the frame indexes associated with different channels are not exactly correlation.

Hence, at the detection stage, the position of the leak is obscure, and the time of arrival of the signals at the sensors is unknown. Thus, one way to deal with this problem is to select a reasonable frame size. Figure 7 uses an example of two signals s_1 and s_2 to explain the method. In this figure, Δt_{lag} and t_{frame} are the lag time and frame size (in time) of the two signals, respectively. Due to the existence of the lag time Δt_{lag} , a lag part of signal s_1 cannot be correlated with any part of signal s_2 in the same frame index i because it has already propagated in frame $(i - 1)$ of s_2 . Thus, a formula for the frame size can be defined as:

$$t_{frame} = \Delta t_{lag} + t_{ext} \quad (9)$$

where t_{ext} is an amount of time to extend the frame size. Obviously, the bigger t_{ext} is, the smaller the lag compared with the remains, which reduces the effect of the lag on the correlation.

Figure 7. Frame division dependent on lag time.

Next, the parameter Δt_{lag} is calculated. The location of the leak in the pipeline is unknown; however, the leak lies somewhere within the tested pipeline. According to this condition, the following equations can calculate the maximum lag time and this result is used to calculate the reasonable frame size.

$$\Delta t_{lag} = \frac{(D + d)}{C} \quad (10)$$

In (10), C is the wave speed. According to [19], AE signals can be propagated in fluid in the frequencies range of 20 kHz to 80 kHz besides propagating through the pipe wall in high frequencies. Furthermore, AE signals of leaks are from the flow turbulence and interaction of particles at the leak point. Thus, AE signals might contain both kinds of propagation. In other words, the wave speed in water is smaller than in solid materials [20]. Hence, the value of C should be calculated with the propagation in water. The wave speed can be calculated as follows [21,22]:

$$C = \sqrt{\frac{K}{\rho \left(1 + \frac{D_p}{e} \frac{K}{E} \psi \right)}} \quad (11)$$

$$\psi = \left[1 / \left(1 + e / D_p \right) \right] \left[1 + 2 \left(e / D_p \right) (1 + u) \left(1 + e / D_p \right) \right] \quad (12)$$

where K and ρ are the volumetric compressibility modulus and the liquid density of the medium inside the pipeline, e and D_p are the thickness and inner diameter of the pipe, ψ is a factor related to the pipe supporting condition, and u is Poisson's ratio.

Thus, the frame size is counted as:

$$t_{frame} = (1 + \zeta) \Delta t_{lag} \quad (13)$$

$$\zeta = \frac{t_{ext}}{\Delta t_{lag}}. \quad (14)$$

In (14), ζ must keep the lagged part of the signals, which is not very large as compared to the rest of the signal.

3.3.2. $g(r)$ -Construction

The quantity $g(r)$ in Section 3.1 is formulated by dividing the variance of one frequency for the leak signal. In this section, $g(r)$ vector is constituted from the signal of sensors 1 and 2. The proportions of the amplitudes over all the frequencies are considered in the $g(r)$ vector (Figure 8).

Figure 8. $g(r)$ -construction. FFT, fast Fourier transform.

In Figure 8, time domain signals are converted to the frequency domain by fast Fourier transform (FFT), taking only their amplitudes as components of the divider for every frequency. After the transformation, we have a new signal in the form of $g(r)$ containing information about leakage symptoms.

3.3.3. Feature Extraction

In this study, the three features given in Table 3 are used to compare the performance of the direct AE-based method with that of the $g(r)$ -based one. These features are selected because of their effectiveness in both time and frequency domains.

Table 3. Typical features.

No	Feature	Definition	Abbreviation
1	Root mean square	$RMS = \sqrt{\sum_{i=1}^N x_i^2 / N}$	RMS
2	Short time energy	$STE = \sum_{i=1}^N x_i^2$	STE
3	Average amplitude	$\mu = \sum_{i=1}^N x_i / N$	AVA

3.3.4. Classification

KNN is a popular kernel function used to identify instances belonging to different classes during the diagnosis. The theory of KNN is presented clearly in [18,23]. This paper uses the KNN-based classifier to solve a binary classification problem, i.e., whether a sample belongs to the normal or abnormal conditions of the pipeline.

4. Experimental Results

For a comparison between the direct AE-based and $g(r)$ -based approaches, the necessary dataset parameters are listed in Table 4, where P is one of the pressures {P1, P2, P3} and L is one of the leaks {L1, L2, L3, L4} defined in Section 2.1.

Table 4. Dataset parameters.

Parameters	Training	Testing
Sample frequency (MHz)	1	1
Window type	Hanning	Hanning
Window size (points)	3000	3000
Overlapping of window	50%	50%
{P, L, 'NORMAL'} (frame)	6665	73,331
{P, L, 'ABNORMAL'} (frame)	6665	73,331

After the AE signals are divided into frames, they are transferred directly to the feature extraction block in the direct AE-based method, and to the $g(r)$ -construction block in the proposed method (Figure 6). Figure 9 illustrates the AE signals of the pair (P1, L1), and Figure 10 presents the signals in a frame index.

The R15I-AST AE sensors used in the experiment are manufactured by MISTRAS in accordance with the industrial standards, and they can capture signals with low level as shown in Figures 9 and 10. These signals demonstrate that noise levels of all the channels are nearly same in the normal condition, but differ from those of the abnormal condition. Channel 1 is mounted at the furthest place from the leak, and the signal level at this point is the lowest, which resembles the noise, while the two others get higher amplitudes for being closer to the leak. Moreover, such a low level of signals is prone to noise in industry. Therefore, a classification algorithm relying on absolute levels is unreliable if it is only trained by a limited number of datasets. In a real application, the leak position is obscure; AE signals can vary between low level like noise (when AE sensors are far from the leak) and higher levels (when they are closer to the leak). To address the issue, a classifier is trained by the quantity $g(r)$ formed by AE signals from critical conditions; for example, when sensors are mounted far from the leak, and pressure is low. The proposed model can detect a leak correctly even if the environment fluctuates widely.

Figure 9. Datasets of the pair (P1, L1).

The $g(r)$ -construction block in Figure 8 converts the signals of channels 1 and 2 from the time domain to the frequency domain and then transforms into the quantity $g(r)$. Figure 11 presents a frame of $g(r)$ in which the normal $g(r)$ fluctuates around 0, but the abnormal $g(r)$ has a different trend. This behavior is independent of distance and resistant to noise.

Figure 10. A frame of datasets of the pair (P1, L1).

Figure 11. A frame of $g(r)$ for datasets of the pair (P1, L1).

4.1. Effectiveness of the $g(r)$ -Based Approach Compared with the Direct AE-Based Method

The effectiveness of the $g(r)$ -based approach compared with the direct AE-based method is illustrated by state scattering in 3D space, as shown in Figures 12–15. The plot uses the features and a separability comparison that relies on Kullback-Leibler (KL) distances [18] between the two classes under various experimental conditions. The KL distance is given as follows:

$$d_{kl} = D_{12} + D_{21} \quad (15)$$

$$D_{12} = \sum p(x|w_1) \ln \frac{p(x|w_1)}{p(x|w_2)} \quad (16)$$

$$D_{21} = \sum p(x|w_2) \ln \frac{p(x|w_2)}{p(x|w_1)} \quad (17)$$

where w_1, w_2 are the two classes (i.e., NORMAL and ABNORMAL) and $x = [x_1, x_2, \dots, x_n]^T$ is the features for n data frames, and p is the probability density function.

Figures 12–15 show state scattering based on the features extracted from direct AE signals and $g(r)$.

Figure 12. State scattering based on features extracted from the AE signal on channel 1 (CH1).

Figure 13. State scattering based on features extracted from the AE signal of channel 2 (CH2).

Table 5 shows the *KL* distances for each pair of pressure, leak (P, L) values. Since the *KL* distances depend on features and their values can be relatively large, this study uses a logarithm calculation to convert the distances into dB for demonstration. Obviously, the *KL* distances based on features extracted directly from the AE signals vary in a wide range. In contrast, the *KL* distances calculated

using features extracted from $g(r)$ remain around a certain value for each type of feature under diverse conditions. Moreover, the average values also demonstrate that the class separability of the $g(r)$ -based approach is greater than that of the direct AE-based one. From the state scattering and the KL distances, it is clearly observed that the $g(r)$ -based approach is more effective than the direct AE-based one.

Figure 14. State scattering based on features extracted from the AE signal of channel 3 (CH3).

Figure 15. State scattering based on features extracted from $g(r)$.

Table 5. Kullback-Leibler distances in decibels (dB) (d_{kl} (dB) = $10 \cdot \log_{10}(d_{kl})$).

P	L	CH1			CH2			CH3			$g(r)$		
		RMS	STE	AVA	RMS	STE	AVA	RMS	STE	AVA	RMS	STE	AVA
P1	L1	11	1	16	1	22	−10	8	25	2	59	112	60
P1	L2	9	−1	14	11	42	5	16	44	11	63	118	63
P1	L3	−6	13	2	3	34	−6	8	35	4	70	109	70
P1	L4	−2	6	5	11	44	6	17	46	12	63	110	63
P2	L1	0	24	−11	12	42	8	16	42	12	70	102	71
P2	L2	2	36	−7	15	58	11	21	61	17	73	104	73
P2	L3	3	29	−4	24	65	20	29	67	25	60	116	60
P2	L4	4	33	−2	13	49	8	18	51	14	68	110	68
P3	L1	14	48	10	23	64	19	27	65	23	72	101	73
P3	L2	10	48	5	17	62	12	22	65	18	76	99	77
P3	L3	7	−3	11	5	36	0	9	35	4	73	107	74
P3	L4	6	37	1	12	48	8	19	52	15	69	108	70
MEAN		5	23	3	12	47	7	18	49	13	68	108	69

4.2. Classifier Training and Testing

4.2.1. Training

This paper uses datasets of pressure and leak {P, L} pairs to train the KNN classifier used in the two approaches. Table 6 presents the performance of the two approaches. The P4 column in the table represents the combination of datasets with all the pressures and leaks. Although Table 6 shows high accuracies with both methods, their performances become different when applying them under different conditions, as shown in the next section.

Table 6. Trained classifiers and their accuracy (%).

	Direct AE-Based				$g(r)$ -Based			
	P1	P2	P3	P4	P1	P2	P3	P4
L1	97.50	100.0	100.0	99.00	98.80	100.0	100.0	100.0
L2	100.0	100.0	100.0		100.0	100.0	100.0	
L3	100.0	100.0	100.0		100.0	100.0	100.0	
L4	100.0	100.0	100.0		100.0	100.0	100.0	

4.2.2. Cross Testing

The previous section trained 12 classifiers using datasets for conditions corresponding to each pair {P_i, L_j} for each method: direct AE-based and $g(r)$ -based. This section uses datasets in conditions {P_m, L_n} to test them, where $i, m \in \{1, 2, 3\}$ and $j, n \in \{1, 2, 3, 4\}$. The purpose of the test is to verify the reliability of each approach if the training is carried out using only some of the datasets. Table 7 shows the results of the experiment carried under this configuration. In each table, the trained classifiers are symbolized in the columns, and the test datasets are listed in the rows.

It is obvious from Tables 7–10 that when the testing datasets use the same conditions as the training datasets, the accuracies reach 100% or nearly 100%. However, if the testing conditions differ from the training conditions, the accuracies are not promising, and misclassification can occur when using the direct AE-based method. In Table 8, an accuracy of 00.00% indicates misclassification. In the same situation; however, the $g(r)$ -based approach identifies leaks without misclassification. Furthermore, the average accuracy of the $g(r)$ -based approach is higher than that of the direct AE-based method in every test.

Table 7. Accuracy (%) of classifiers trained with leak L1.

		Direct AE-Based			$g(r)$ -Based		
		P1	P2	P3	P1	P2	P3
L1	P1	98.55	69.64	62.66	99.78	99.82	99.82
	P2	99.54	99.98	65.85	99.82	99.88	99.88
	P3	96.71	99.94	100.0	99.88	99.82	99.80
L2	P1	99.88	100.0	64.45	99.98	99.76	99.76
	P2	98.23	99.86	100.0	100.0	99.88	99.86
	P3	99.76	100.0	100.0	99.98	99.96	99.96
L3	P1	100.0	100.0	62.91	99.96	99.90	99.88
	P2	99.42	99.96	67.25	100.0	99.82	99.82
	P3	99.32	99.98	82.05	99.76	99.82	99.84
L4	P1	99.96	64.27	62.94	99.98	99.90	99.88
	P2	99.80	99.94	63.18	99.94	99.96	99.94
	P3	99.70	99.94	63.83	99.19	99.84	99.84
MEAN		99.24	94.46 89.43	74.59	99.86	99.86 99.86	99.86

Table 8. Accuracy (%) of classifiers trained with leak L2.

		Direct AE-based			$g(r)$ -Based		
		P1	P2	P3	P1	P2	P3
LS1	P1	65.77	62.54	00.00	86.92	99.54	99.90
	P2	100.0	63.25	62.69	62.91	99.52	99.92
	P3	100.0	77.53	64.53	62.65	99.58	99.92
LS2	P1	100.0	62.65	62.51	100.0	99.98	99.84
	P2	99.94	100.0	99.36	70.40	100.0	99.92
	P3	100.0	100.0	100.0	63.00	99.96	99.98
LS3	P1	100.0	62.58	62.53	98.33	99.88	99.92
	P2	99.96	64.15	62.91	98.06	100.0	99.92
	P3	100.0	62.99	62.74	98.06	99.64	99.74
LS4	P1	63.97	62.58	00.00	64.73	99.94	99.98
	P2	99.98	62.74	62.58	98.56	99.84	99.96
	P3	99.96	63.04	62.68	95.72	98.95	99.52
MEAN		94.13	70.34 74.34	58.54	83.28	99.74 94.30	99.88

Table 9. Accuracy (%) of classifiers trained with leak L3.

		Direct AE-Based			$g(r)$ -Based		
		P1	P2	P3	P1	P2	P3
LS1	P1	70.51	73.06	63.19	97.96	95.10	97.83
	P2	99.98	99.96	76.28	81.45	66.99	79.70
	P3	99.92	99.88	100.0	76.79	64.07	74.84
LS2	P1	100.0	100.0	99.82	99.98	100.0	99.98
	P2	99.86	99.80	100.0	99.90	99.64	99.88
	P3	100.0	100.0	100.0	98.95	78.53	98.64
LS3	P1	100.0	100.0	63.87	99.32	99.07	99.30
	P2	99.96	99.96	72.89	99.98	99.70	99.96
	P3	99.98	99.96	100.0	99.21	98.85	99.13
LS4	P1	64.34	64.60	63.52	98.50	92.05	98.42
	P2	99.94	99.92	100.0	99.52	99.27	99.50
	P3	99.94	99.94	80.45	97.83	97.35	97.79
MEAN		94.54	94.76 91.43	85.00	95.78	90.89 94.03	95.41

Table 10. Accuracy (%) of classifiers trained with leak L4.

		Direct AE-Based			$g(r)$ -Based		
		P1	P2	P3	P1	P2	P3
LS1	P1	97.00	63.79	67.29	99.78	98.08	98.33
	P2	99.72	99.52	100.0	99.88	82.62	86.62
	P3	97.60	100.0	99.98	99.96	77.92	82.90
LS2	P1	99.94	100.0	100.0	99.98	99.98	99.98
	P2	98.64	100.0	99.94	100.0	99.90	99.94
	P3	99.86	100.0	100.0	99.98	99.03	99.36
LS3	P1	100.0	98.62	100.0	99.98	99.32	99.42
	P2	99.62	88.80	99.98	100.0	99.98	100.0
	P3	99.56	100.0	100.0	99.78	99.21	99.23
LS4	P1	99.96	63.72	64.06	100.0	98.52	98.93
	P2	99.80	100.0	99.94	99.96	99.54	99.60
	P3	99.80	99.44	99.94	99.30	97.87	97.96
MEAN		99.29	92.82 95.46	94.26	99.88	96.00 97.58	96.86

4.2.3. Testing Classifiers Trained by Combined Datasets

To detect a fault, a classifier that depends on data must be trained using many datasets representing various experimental conditions to provide high reliability. Table 11 shows accuracies of the two approaches when using the dataset of each pair {P, L}.

Table 11. Accuracy (%) of classifiers of two approaches by combined datasets.

		Direct AE-Based			$g(r)$ -Based		
		P1	P2	P3	P1	P2	P3
L1		96.49	99.72	97.75	99.94	100.0	100.0
L2		99.96	98.76	99.88	99.90	100.0	100.0
L3		100.0	99.66	99.64	100.0	99.92	99.96
L4		99.96	99.80	99.80	100.0	100.0	99.90
MEAN		99.10	99.48 99.28	99.27	99.96	99.98 99.97	99.97

The average accuracies are 99.28 and 99.97 for the direct AE-based and $g(r)$ -based approaches, respectively, as shown in Table 11. Thus, a model trained by many datasets can work in varied conditions. However, the $g(r)$ -based method is more efficient than the direct AE-based because it still achieves the expected accuracy with only a small number of training datasets.

4.2.4. Evaluating Two Approaches Using Combined Datasets with Added Noise

To evaluate robustness of the proposed method, we add Gaussian noise to the acquired signals to increase the intensity of background noise. The parameter γ is counted in (dB) for each channel of the three AE sensors. Figure 16 illustrates the signals from channel 2 before and after added noise with $\gamma = 0$ (dB). This evaluation is conducted using the KNN classifier trained by the combination of all datasets. The testing accuracies are shown in Table 12.

Figure 16. AE on channel 2 before/after adding noise.

Table 12. Accuracy with added noise.

γ (dB)	Model	
	Direct AE-Based	$g(r)$ -Based
−30	99.28	99.97
−20	99.28	99.97
−10	99.28	99.97
−5	99.25	99.96
0	99.08	99.83
5	00.00	98.46
10	00.00	94.38
15	00.00	88.83
20	00.00	76.90
25	00.00	70.74
30	00.00	66.96
35	00.00	00.00

Table 12 shows that if γ is small, the accuracies of the two methods are approximately the same as in Table 11. When γ increases, the accuracy of both approaches decreases, and at a certain value, they cannot guarantee classification. However, the direct AE-based method fails in classification when the added noise reaches 5 (dB), whereas, the proposed $g(r)$ -based method works until γ reaches 35 (dB). Thus, the $g(r)$ -based approach produces a classifier that can detect a small leak more robustly than the direct AE-based classifier.

5. Conclusions

AE-based techniques have the advantage of detecting the physical changes within a structure such as a water pipeline with high accuracy. Due to the complication of the AE phenomenon, model-based fault diagnosis of the water pipeline is difficult, whereas a data-driven fault diagnosis is relatively easy to implement. However, if the training relies on extracting features directly from AE signals, the resulting classifier is unreliable because such data cannot reflect all possible information under divers working conditions. This paper introduced an intermediate step in a water pipeline fault diagnosis framework in which the signals are preprocessed before the extraction of features.

The preprocessing step calculates an intermediate quantity $g(r)$ by applying an attenuation equation to the acquired AE signals. The quantity $g(r)$ is more stable than the original AE signals in leakage identification. The results showed that the $g(r)$ -based approach achieves higher accuracy and higher reliability than the direct AE-based method. This approach can be applied in a variety of systems because the $g(r)$ function always possesses the same behavior reflecting leaks in a system. This paper also proposed a way to select a reasonable window size when dividing signal frames. The proposed technique uses the speed of acoustic emission signal to calculate a possible maximum lag time between signals on the sensor channels. From that computation, the frame size can be adjusted appropriately.

Author Contributions: All authors contributed equally to the conception of the idea, as well as implementing and analyzing the experimental results, and writing the manuscript.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20172510102130). This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2016R1D1A3B03931927).

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

References

1. Chan, T.K.; Chin, C.S.; Zhong, X. Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection. *IEEE Access* **2018**, *6*, 78846–78867. [CrossRef]
2. Datta, S.; Sarkar, S. A review on different pipeline fault detection methods. *J. Loss Prev. Process Ind.* **2016**, *41*, 97–106. [CrossRef]
3. Puust, R.; Kapelan, Z.; Savic, D.A.; Koppel, T. A review of methods for leakage management in pipe networks. *Urban Water J.* **2010**, *7*, 25–45. [CrossRef]
4. Murvay, P.-S.; Silea, I. A survey on gas leak detection and localization techniques. *J. Loss Prev. Process Ind.* **2012**, *25*, 966–973. [CrossRef]
5. Martini, A.; Troncossi, M.; Rivola, A. Vibroacoustic Measurements for Detecting Water Leaks in Buried Small-Diameter Plastic Pipes. *J. Pipeline Syst. Eng. Pract.* **2017**, *8*, 04017022. [CrossRef]
6. Li, S.; Song, Y.; Zhou, G. 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]
7. 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]
8. Xiao, Q.; Li, J.; Bai, Z.; Sun, J.; Zhou, N.; Zeng, Z. A Small Leak Detection Method Based on VMD Adaptive De-Noiseing and Ambiguity Correlation Classification Intended for Natural Gas Pipelines. *Sensors* **2016**, *16*, 2116. [CrossRef] [PubMed]
9. Nicola, M.; Nicola, C.I.; Vintila, A.; Hurezeanu, I.; Duta, M. Cross-Correlation, Acoustic emission, Leakage detection, LabVIEW environment. *J. Mech. Eng. Autom.* **2018**, *8*, 59–67.
10. Sun, J.; Wen, J. Target location method for pipeline pre-warning system based on HHT and time difference of arrival. *Measurement* **2013**, *46*, 2716–2725. [CrossRef]
11. Martini, A.; Troncossi, M.; Rivola, A. Leak Detection in Water-Filled Small-Diameter Polyethylene Pipes by Means of Acoustic Emission Measurements. *Appl. Sci.* **2016**, *7*, 2. [CrossRef]
12. Antaki, G. *Piping and Pipeline Engineering: Design, Construction, Maintenance, Integrity, and Repair*; Dekker Mechanical Engineering; CRC Press: Boca Raton, FL, USA, 2003; Volume 159, ISBN 978-0-8247-0964-8.
13. Rienstra, S.W.; Hirschberg, A. An Introduction to Acoustics. 298. Available online: <https://www.win.tue.nl/~{sjoerdr/papers/boek.pdf> (accessed on 2 December 2018).
14. Martini, A.; Rivola, A.; Troncossi, M. Autocorrelation Analysis of Vibro-Acoustic Signals Measured in a Test Field for Water Leak Detection. *Appl. Sci.* **2018**, *8*, 2450. [CrossRef]
15. He, P. Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **1998**, *45*, 114–125. [PubMed]
16. Szabo, T.L.; Wu, J. A model for longitudinal and shear wave propagation in viscoelastic media. *J. Acoust. Soc. Am.* **2000**, *107*, 2437–2446. [CrossRef] [PubMed]

17. Chen, W.; Holm, S. Modified Szabo's wave equation models for lossy media obeying frequency power law. *J. Acoust. Soc. Am.* **2003**, *114*, 2570. [[CrossRef](#)] [[PubMed](#)]
18. Koutroumbas, K.; Theodoridis, S. *Pattern Recognition*, 4th ed.; Academic Press: New York, NY, USA, 2008.
19. BS EN 15856 Non-Destructive Testing—Acoustic Emission—General Principles of AE Testing for the 416 Detection of Corrosion with Metallic Surrounding Filled with Fluid 2010. Available online: <https://shop.bsigroup.com/ProductDetail/?pid=000000000030165681> (accessed on 2 December 2018).
20. Measurement & Control. Sound Speed & Pipe Size Data 2014. Available online: <https://www.rshydro.co.uk/files/SNSP-and-Pipe-Data.pdf> (accessed on 2 December 2018).
21. Twyman, J. Wave speed calculation for water hammer analysis. *Obras y Proyectos* **2016**, *20*, 86–92. [[CrossRef](#)]
22. Watters, G.Z. *Analysis and Control of Unsteady Flow in Pipelines*; Butterworths: Boston, MA, USA, 1984; ISBN 0-250-40492-3.
23. Devroye, L.; Györfi, L.; Lugosi, G. *A Probabilistic Theory of Pattern Recognition*; Springer: New York, NY, USA, 1996; ISBN 978-0-387-94618-4.

© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).