

Article

Real-Time Leak Detection for a Gas Pipeline Using a k -NN Classifier and Hybrid AE Features

Thang Bui Quy  and Jong-Myon Kim * 

School of Electrical, Electronics, and Computer Engineering, University of Ulsan, Ulsan 44610, Korea; bqthangcndt@gmail.com

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

Abstract: This paper introduces a technique using a k -nearest neighbor (k -NN) classifier and hybrid features extracted from acoustic emission (AE) signals for detecting leakages in a gas pipeline. The whole algorithm is embedded in a microcontroller unit (MCU) to detect leaks in real-time. The embedded system receives signals continuously from a sensor mounted on the surface of a gas pipeline to diagnose any leak. To construct the system, AE signals are first recorded from a gas pipeline testbed under various conditions and used to synthesize the leak detection algorithm via offline signal analysis. The current work explores different features of normal/leaking states from corresponding datasets and eliminates redundant and outlier features to improve the performance and guarantee the real-time characteristic of the leak detection program. To obtain the robustness of leak detection, the paper normalizes features and adapts the trained k -NN classifier to the specific environment where the system is installed. Aside from using a classifier for categorizing normal/leaking states of a pipeline, the system monitors accumulative leaking event occurrence rate (ALEOR) in conjunction with a defined threshold to conclude the state of the pipeline. The entire proposed system is implemented on the 32F746G-DISCOVERY board, and to verify this system, numerous real AE signals stored in a hard drive are transferred to the board. The experimental results show that the proposed system executes the leak detection algorithm in a period shorter than the total input data time, thus guaranteeing the real-time characteristic. Furthermore, the system always yields high average classification accuracy (ACA) despite adding a white noise to input signal, and false alarms do not occur with a reasonable ALEOR threshold.

Keywords: pipeline leak detection; acoustic emission analysis; signal classification; k -NN algorithm; hybrid AE features



Citation: Quy, T.B.; Kim, J.-M. Real-Time Leak Detection for a Gas Pipeline Using a k -NN Classifier and Hybrid AE Features. *Sensors* **2021**, *21*, 367. <https://doi.org/10.3390/s21020367>

Received: 5 December 2020

Accepted: 1 January 2021

Published: 7 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gas pipelines play a vital role in the fuel transportation field. Even though they are designed and assembled according to strict technical principles [1,2], a gas leak could still occur due to material aging and corrosion [3,4], leading to violent explosions causing injuries, human deaths, and pollution of the environment. Hence, a real-time gas pipeline leak detection system is extremely important to reduce catastrophic consequences.

In early times, acoustic emission (AE) was mainly used for detecting growing cracks and discontinuities in materials because it was defined as releasing elastic energies in a deformed material [5]. However, AE is currently referred to as a phenomenon where transient elastic waves are generated by the rapid release of energy from localized sources within a material, or the transient elastic waves so generated [5]. As a result, a leak is also the source of AEs and is detectable with AE equipment. This type of AE source is sometimes called a secondary source to distinguish it from the classic AEs which are caused by material deformation [5]. The AE-based leak detection is therefore feasible. Consequently, many researchers have applied this mechanism to detect a leak in a gas pipeline [6–13]. The leak detection using AE signals is extremely beneficial because it is a non-destructive technique,

thus it does not affect the working system [5,14–16]. Additionally, the symptom of small leakage is often extraordinarily subtle; for example, a small leak can be caused by an initial crack that does not create an obvious rupture. Hence, AE sensors can be applicable in this case because they offer high sensitivity regarding any early abnormality [17].

Researchers tend to adopt a data-driven approach that trains a classifier using AE features extracted from AE signals to separate pipeline health states to normal or leaking. This approach is appropriate because an AE signal acquired from a gas pipeline is non-stationary [18,19]. Moreover, AE waves attenuate along the pipeline from their emission source to AE sensors [20]; they vary with the environmental conditions of pressure, flow rate, and temperature [21]. Thus, it is challenging to draw an explicit model to identify a leak relying on AE signals exclusively. A classification model learns the leakage manifestation from the supplied training data; hence, it can identify the leak detection problem effectively. However, the computational complexity of existing leak detection methodologies restricts their exploitation in real-time applications, despite the fact that they show high classification accuracy. For example, the Wavelet transform and the signal decomposition algorithms are used to analyze AE signals, and machine learning-based models are used for state classification [10–12,22], which can improve accuracy, but their computation is highly complex.

A long gas transportation system usually comprises numerous pipeline segments with diversity in size, shape, and material. Many sensors are spread over that system to monitor the health of different pipeline segments. A wireless-based leak detection system with a server receiving and exploring signals dispatched from remote sensor nodes, as proposed by [23], would not be suitable for AE signal application due to the overload of communication and computation. Therefore, a sensor node should be a smart integrated system that can itself inspect a pipeline segment and report only the health state of pipeline to its server instead of sending a massive amount of AE signals to the server. The advantage of the integration is that it does not require a complex communication network topology between the sensor nodes and the server. Nonetheless, the integrated system must be low-power and compact, because if many devices are installed, they will result in high energy consumption and a bulky system. This is similar to the design presented in [24], which integrated a propane sensor with a low-power system-on-chip device. However, a propane sensor could only detect an obvious gas leak nearby, thus challenging the early gas leakage detection in a large pipeline network, where a tiny leak would occur at any place and any time.

Working from the demand for gas pipeline leak detection and the achievements and limitations of current studies, this work presents a microcontroller unit (MCU)-based system designed to diagnose leakage for a gas pipeline in real-time. The system analyzes AE signals locally to identify a leak and just issues a warning of state changing. Because an MCU-based system only supports a restricted resource in memory and execution speed for computing implementation, the paper exploits a k -nearest neighbor (k -NN) classifier trained by using hybrid AE features directly extracted from raw AE signals. The k -NN algorithm can execute on a limited-resource platform in real-time because it is made up of simple computations and neighbor-searching loops. To optimize the algorithm further, a filtering technique is exploited to remove the least useful elements from the feature pool relying on the three-sigma rule [25] and the Kullback–Leibler (KL) distance [26], which reduces the number of computation cycles and loops in the correspondingly implemented program, thus accelerating the proposed detection system. The selected features are normalized as well; hence, a trained k -NN model can be applied to various sensor nodes along a pipeline network. Moreover, the trained model can be updated in run-time to adapt to a sensor installation location or any change in the working conditions.

Prior to implementing the leak detection program on an MCU-based hardware platform, the proposed methodology is offline synthesized using the Matlab 2019a software and AE signal datasets recorded at a gas pipeline testbed under diverse experimental scenarios. Thus, the essential parameters of the k -NN classifier (training features and

number of nearest neighbors) are chosen to ensure not only the real-time characteristic, but also high accuracy of the leak detection program. Aside from ambient noise, any external factor that can cause the vibrations in the pipeline can trip AE signals. For instance, a random pipe collision triggers a mechanical vibration that generates plentiful elastic waves propagating through the pipeline. AE sensors with enough sensitivity can capture signals resulting from those elastic waves, thus interfering with measured target signals. Hence, a k -NN classifier based on AE signals is subjected to discrete events near the testing pipeline, generating false alarms. To address this problem, the current work proposes monitoring the accumulative leaking event occurrence rate (ALEOR) from the output of the state classifier. A final decision of pipeline health state is based on the comparison between the instant ALEOR and a defined threshold, hence avoiding a false alarm.

Finally, the work evaluates the gas pipeline leak detection system constructed from the proposed methodology on the 32F746G-DISCOVERY board (STMicroelectronics, Quakertown, PA, USA) using recorded AE signal datasets. Experimental results demonstrate that the system can identify a leak in real-time with high average classification accuracy under various pressure conditions, and its robustness is satisfactory, even with adding white noise to the input AE signal. Hence, the proposed MCU-based system is applicable for gas leak detection in real applications.

2. AE Signal Data Acquisition

A pipeline testbed is established to simulate the gas leakage as shown in Figure 1. The testbed is a part of a real gas pipeline system (see Figure 1c) made from stainless steel 304 pipelines with sizes of 114.3 millimeters (mm) and 6.02 mm in outer diameter and wall thickness, respectively. To create various leaks, we designed a leak tool as shown in Figure 1a, which is assembled to the testing pipeline. This tool is composed of a valve and an orifice of diameter 0.3 mm, 0.5 mm, or 1 mm (see Figure 1b). Hence, the normal/leaking states of the pipeline are connected to closed/open valve positions.

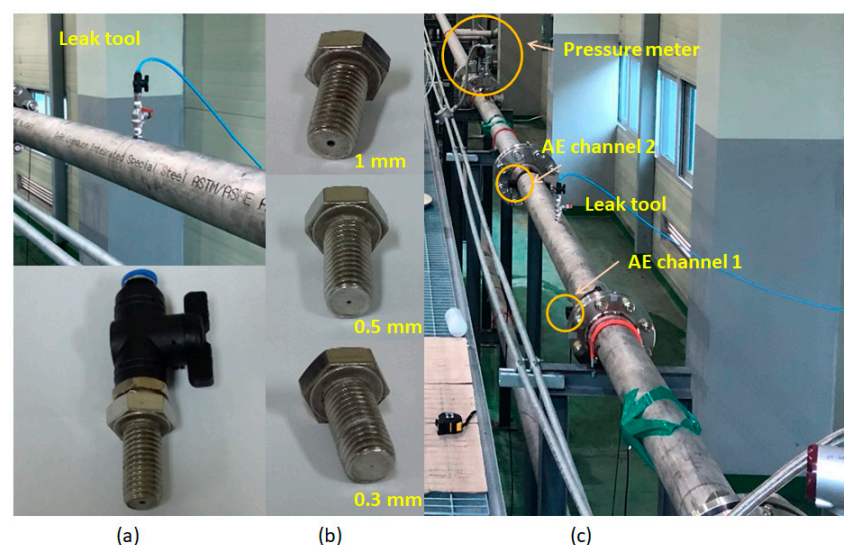


Figure 1. Pipeline testbed: (a) leak tool, (b) orifices, (c) test section.

The experimental configuration is shown in Figure 2. To capture AE signals, two R15i-AST sensors (AE channels), which were manufactured by MITRAS Group, Inc (Princeton Junction, NJ, USA), are mounted at downstream and upstream locations on the surface of the testing pipeline. These sensors can detect any elastic wave in a range of operating frequencies, which are 50 kilohertz (kHz) to 400 kHz [27]. Those elastic waves can be caused by diverse sources such as leak noise [10], negative pressure wave [4], ambient noise, and other vibrations of the pipe wall. Such R15i-AST sensors are selected because their operating frequency range covers the frequency ranges of AE waves propagating in

metal objects, which are from 100 kHz to 300 kHz, as stated in the BSI standard BS EN 15,856 [15]. AE signals are sampled at 1 megahertz (MHz) by the NI-9223 module. The sampling frequency of 1 MHz is more than double the maximum operating frequency of sensors, thus satisfying the Nyquist–Shannon sampling theorem [28] about converting analog signals into digital signals.

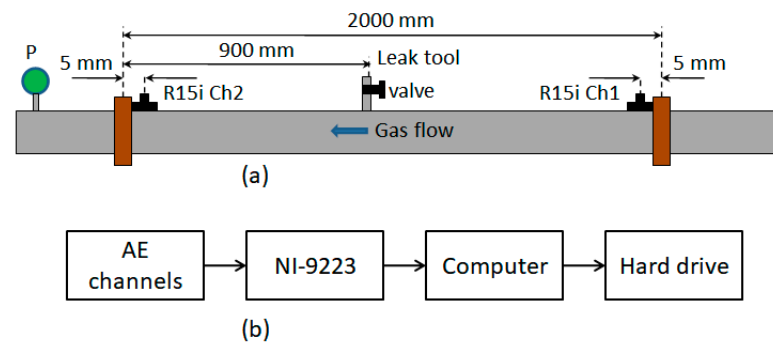


Figure 2. Experimental setup: (a) test section, (b) data acquisition system. (R15i Ch1 and R15i Ch2 are acoustic emission (AE) channels, P is a pressure meter).

After finishing the hardware setup, data recording software is installed on the computer to control the whole data acquisition. Additionally, we exploit the pencil lead break technique [29] to examine both sensitivity of sensors and the whole AE equipment. This ensures the reliability of AE signal datasets prior to storing them in the hard drive.

In the experiment, the three orifices are alternated to simulate different leakages at three inner relative pressures of 700 kPa, 1300 kPa, and 1800 kPa, resulting in three normal states of the testing pipeline (closed valve) and nine diverse leaking states (open valve). Specifically, data acquisition has been performed as follows. First, an orifice was installed, and the pipeline system was configured at a pressure level of 700 kPa, 1300 kPa, or 1800 kPa, and this condition was kept relatively stable before acquiring AE signals. At this time, the valve of the leak tool was closed to simulate the normal state of the pipeline. For this state, the signals were recorded for 2 min. Next, the valve was opened to simulate a leakage. Here, the data corresponding to a leaking state were collected after pressure stabilization. Figure 3 presents gas flow rates measured in front of the testing pipeline during the experimental stages.

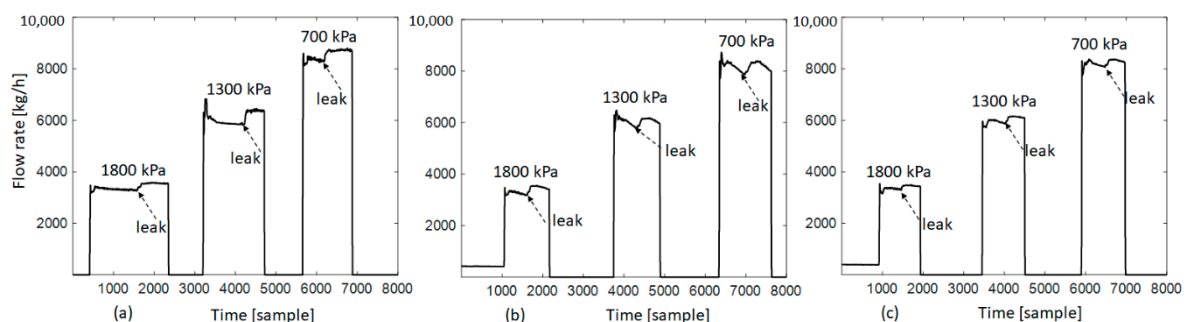


Figure 3. Gas flow rates corresponding to three orifices: (a) 0.3 mm, (b) 0.5 mm, (c) 1.0 mm.

3. Leak Detection Methodology

The overall gas pipeline leak detection diagram is shown in Figure 4. It is composed of two processes: one is offline, and the other is online. The offline analysis synthesizes and optimizes the leak detection algorithm, while the online process experiments and verifies the detection. We will describe the analysis blocks of the algorithm below.

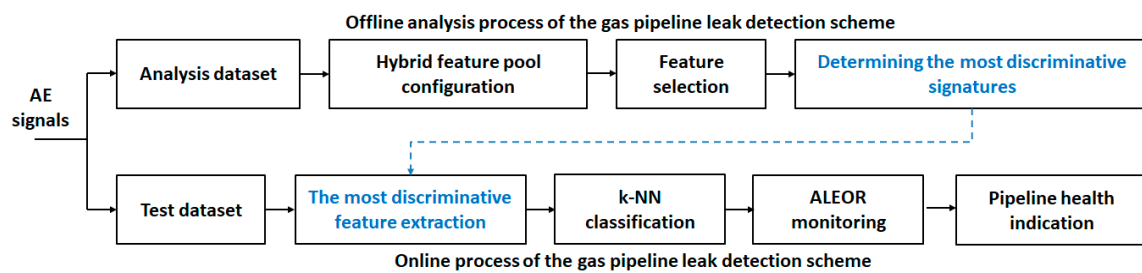


Figure 4. Entire flow diagram of the gas pipeline leak detection.

3.1. Hybrid Feature Pool and Feature Selection

To detect the leaking state of a gas pipeline, time and frequency domain statistical features are extracted, as shown in Table 1, from raw AE signals utilized as diagnosis leakage signatures. We therefore obtain a hybrid feature pool of size $R \times M$, where R is the number of feature types ($R = 12$, as shown in Table 1), and M is the number of analyzed signal frames. The value M should be large enough to reflect the statistical discrimination of the normal/leaking states precisely.

Table 1. Typical features for leak detection.

Features	Equations	Features	Equations	Features	Equations
Short time energy (STE)	$\sum_{n=0}^{N-1} x_n^2$	Standard deviation (STD, σ)	$\sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} (x_n - \mu)^2}$	Skewness (SKE)	$\frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{x_n - \mu}{\sigma} \right)^3$
Root mean square (RMS)	$\sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x_n^2}$	Zero crossing rate (ZCR)	$\frac{1}{N-1} \sum_{n=1}^{N-1} \text{sign}(x_n - \mu) - \text{sign}(x_{n-1} - \mu) $	Spectral peak (SPP)	$\text{argmax}_f (X(f))$
Average amplitude (AVA)	$\frac{1}{N} \sum_{n=0}^{N-1} x_n $	Entropy (ETY)	$-\sum_{n=0}^{N-1} q_n \log_2 q_n$	Spectral centroid (SPC, f_c)	$\frac{\sum_{m=0}^{M-1} f_m X_m^2}{\sum_{m=0}^{M-1} X_m^2}$
Mean (MEA, μ)	$\frac{1}{N} \sum_{n=0}^{N-1} x_n$	Kurtosis (KUS)	$\frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{x_n - \mu}{\sigma} \right)^4$	Spectral spread (SPS)	$\sqrt{\frac{\sum_{m=0}^{M-1} (f_m - f_c)^2 X_m^2}{\sum_{m=0}^{M-1} X_m^2}}$

Where x is an input signal, N is the total number of samples, X is the short-time spectral amplitude, f is the frequency, M is the total number of discrete frequencies, and $q_n = x_n^2 / \sum_{n=0}^{N-1} x_n^2$.

Next, the feature pool should be refined to enhance the pipeline health classification quality. Outliers, data points that differ significantly from the other aggregated data points in the same class can cause serious problems in statistical analyses. The existence of outliers in a feature extracted from an AE signal measured at a gas pipeline is inevitable, resulting from both exterior and interior factors. The exterior factor could be variability in the measurement. For example, power spikes can interfere with sensed signals, causing outliers in AE features. This problem can be fixed by perfect experimental configuration and the exploitation of high-quality equipment. Outliers may be created by interior factors of the pipeline system, such as burst emissions appearing in high amplitude and energy in AE signals. A gas pipeline itself generates such a signal due to the disturbance between inner gas flow and the gas flow–pipe wall interaction. Nevertheless, outliers should be eliminated from features used for training a classifier because they do not statistically characterize the normal/leaking state discrimination, thus leading to the deterioration of the classification performance. This paper assumes a normal distribution for the AE features; outliers can therefore be detected by the three-sigma rule [25]. This rule is expressed as follows:

$$\Pr(|Y_i - \mu_{y_i}| \leq 3\sigma_{y_i}) \approx 0.99. \quad (1)$$

where Y_i is an observation from a normally distributed feature y_i ; μ_{y_i} and σ_{y_i} are the mean and standard deviation of the distribution, respectively; $i = 1, 2, \dots, R$. According to (1), if $|Y_i - \mu_{y_i}| > 3\sigma_{y_i}$, the value Y_i is considered an outlier and it is removed from the set of y_i -feature observations. After unwanted observations are eliminated from the y_i -vector, the length of y_i -vector is shrunk as M_i^* ($M_i^* \leq M$). Because the feature types distribute dissimilarly, the outlier elimination might return different lengths M_i^* of the y_i -vectors ($i = 1, 2, \dots, R$). As a result, we compensate new satisfactory observations for the feature pool to gain $M_i^* = M$. The feature pool size is therefore intact ($R \times M$); however, its elements are refined, which satisfies (1).

Furthermore, all the extracted features may not be equally effective in highly accurate leak detection. Inferior signatures not only impair the classification accuracy but also increase the computational complexity. Thus, we need to filter out redundant features from the pool to enhance the detection performance while reducing the computational load. This paper scores features using the Kullback–Leibler distance [26] and eradicates low-ranked elements in the feature pool. The KL distance is calculated as follows:

$$d_{KL} = D_{12} + D_{21};$$

$$D_{12} = \sum p(y_i|w_1) \ln \frac{p(y_i|w_1)}{p(y_i|w_2)}; D_{21} = \sum p(y_i|w_2) \ln \frac{p(y_i|w_2)}{p(y_i|w_1)}. \quad (2)$$

where d_{KL} is the KL distance, w_1, w_2 are two classes indicating the normal and leaking states, respectively; $y_i = [y_{i1}, y_{i2}, \dots, y_{iM}]^T$ is a sort of y_i -feature in the refined feature pool, p is a conditional probability density function. Based on (2), we retain features with the dominant KL distance and remove the others in the feature pool, because the greater the KL distance is, the more discriminative the feature. Finally, we retrieve a purified feature pool with size $r \times M$, where r is the number of high-scored features ($r \leq R$).

3.2. Leak Detection Using a k -NN Classifier and Accumulative Leaking Event Occurrence Rate

With the purified feature pool, we utilize a k -NN classifier to distinguish the two normal/leaking states, in which an obscure new class is assigned to the most common class among its k nearest neighbors using the Manhattan distance given by:

$$\delta_j = \sum_{n=1}^r |z_n - y_{n,j}|. \quad (3)$$

where δ_j is the Manhattan distance between the input feature vector $z = \{z_1, z_2, \dots, z_r\}$ and the j th training feature vector $y_{*j} = \{y_{1j}, y_{2j}, \dots, y_{rj}\}$, and $j = 1, 2, \dots, M$. The k -NN classifier categorizes the input z into the major class in its k nearest neighbors corresponding to k minimum distances δ_j ($k < M$).

The detection approach aims at the extremely noisy industrial environment. A k -NN classifier is sensitive to noise involving ambient noise and discrete events and may subsequently yield a false alarm (the classified state is “leaking” but the true state is “normal”) or miss a true leaking event (the leakage is actually happening); thus, a normal/leaking state decision should depend on monitoring the ALEOR. The leak detection criterion is given by:

$$ALEOR = \frac{\Delta B}{\Delta t} \geq \gamma, \Delta t = t_2 - t_1. \quad (4)$$

where ΔB is the number of leaking events in a time period $\Delta t = t_2 - t_1$, which is from the moment t_1 to the moment t_2 , and γ is a threshold to issue a warning of pipeline health state. This threshold is flexibly adjusted by pipeline operators in their specific real environment.

4. Implementation of Proposed Gas Pipeline Leak Detection on an MCU-Based Architecture

4.1. Offline Analysis of AE Signal Datasets

Prior to developing the real-time gas leak detection program with the proposed methodology on an MCU-based architecture, we analyzed offline AE signal datasets to

search for a set of optimal parameters, thus enhancing the performance of the real-time leak detection program. The optimized parameters are the feature pool for training the k -NN classification model and the number k (the number of nearest neighbors used for the k -NN classifier). We perform the offline analysis process using a number of AE datasets, as shown in Table 2.

Table 2. Number of datasets used for the offline analysis and evaluation.

	P ₀		P ₁		P ₂	
	N _{FA}	N _{FE}	N _{FA}	N _{FE}	N _{FA}	N _{FE}
L ₀	600	30,000	600	30,000	600	30,000
L ₁	200	10,000	200	10,000	200	10,000
L ₂	200	10,000	200	10,000	200	10,000
L ₃	200	10,000	200	10,000	200	10,000

For feature selection, we should first normalize extracted features to place them on the same unit basis. The feature normalization is expressed by the following equation:

$$y_{new} = \frac{y_{old} - \mu_{yn}}{\sigma_{yn}}. \quad (5)$$

where y_{old} , y_{new} are original and rescaled features, respectively, and μ_{yn} , σ_{yn} are successively mean and standard deviation of the feature estimated from samples belonging to the normal pipeline state.

Table 3 exhibits feature scores using the KL distance method. The most highly ranked features are STE, RMS, AVA, and STD, and these are returned in every pressure condition. Hence, we only consider these kinds of features to build the real-time gas leak detection program. Figure 5 illustrates the 3-D visualization of three features with the highest scores under diverse pressure conditions, in which the normal/leaking states are obviously separated for all the cases. Moreover, we know that a large k may improve performance; however, too large a k destroys the locality. Therefore, to choose k appropriately, we employ the available k -NN fitting function “fitknn” supported by Matlab 2019a to trial different values of k using the analysis datasets and we obtain $k = 25$.

Table 3. Feature score based on KL distance.

	STE	RMS	AVA	MEA	STD	ZCR	ETY	KUS	SKE	SPP	SPC	SPS
P ₀	57.7	36.5	37.3	−30.2	36.2	10.8	4.0	6.9	−18.0	1.8	7.0	8.9
P ₁	71.9	44.2	44.4	−0.2	44.0	10.2	−1.7	3.3	−9.1	2.0	7.8	7.6
P ₂	77.7	47.2	47.4	5.1	47.1	13.3	−5.6	1.1	−6.5	3.3	11.1	9.0

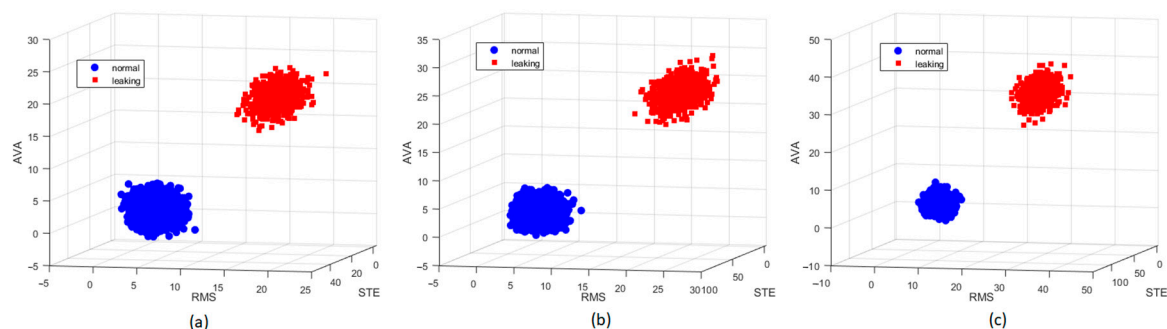


Figure 5. Three-dimensional visualization of the three most highly ranked features under various pressure conditions: (a) P₀, (b) P₁, (c) P₂.

The datasets belong to a signal channel (R15i Ch1 or R15i Ch2), corresponding to three pressure conditions: 700 kPa (P_0), 1300 kPa (P_1), and 1800 kPa (P_2), and pipeline health states: normal (L_0), leaking (0.3 mm (L_1), 0.5 mm (L_2), and 1 mm (L_3)), which were recorded in Section 2; N_{FA} and N_{FE} are the numbers of frames for the offline analysis and experiment respectively, and a frame consists of 8192 samples stored in the hard drive.

4.2. Gas Pipeline Leak Detection Implementation on an MCU-Based Hardware Architecture

4.2.1. Overview of the Experimental Hardware Design with an MCU Used for Real-Time Gas Pipeline Leak Detection

Figure 6 illustrates an MCU-based hardware architecture to implement the proposed method for real-time gas pipeline leak detection. A sensor channel is connected to a data acquisition (DAQ) module which converts analog AE signals to digital AE signals and directly writes them to a synchronous random-access memory (SRAM) through a communication module, along with a direct memory access (DMA) channel available in the MCU; hence, the leak detection program can investigate AE signals in real-time. We also design a portable memory (SDcard) to store some pre-defined parameters of the leak detection program and its runtime log files used for later analyses. Hence, the program can be adjusted and updated quickly. Additionally, a liquid crystal display (LCD) is installed to indicate the output of the diagnostic program. This entire experimental design is embedded in the 32F746G-DISCOVERY board, as shown in Figure 7.

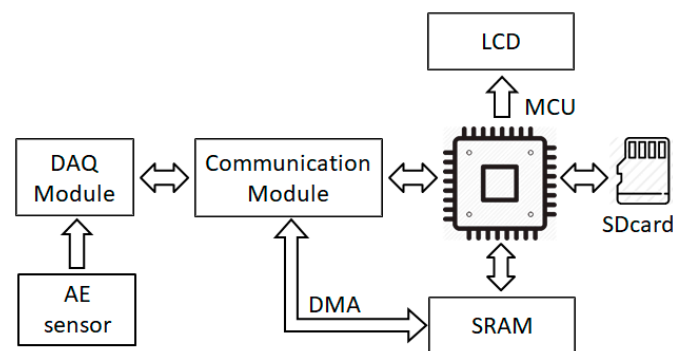


Figure 6. Experimental MCU-based hardware architecture for the gas pipeline leak detection.

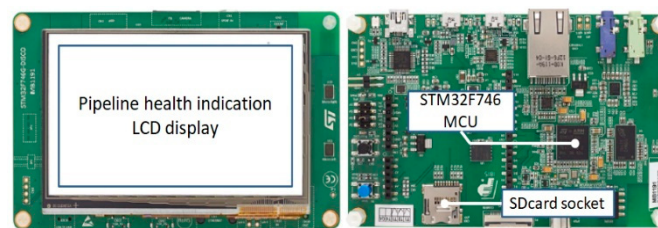


Figure 7. 32F746G-DISCOVERY board (top view: left side—bottom view: right side).

4.2.2. Real-Time Gas Leak Detection Implementation on the 32F746G-DISCOVERY Board

Due to the limitation of the MCU in internal memory and operating speed, we use integer instead of floating-point format for the feature calculation and the k -NN classification, thus utilizing the memory economically and lightening the computation load. In other words, a real feature value is multiplied by 10 before rounding it, which sustains a one-decimal point precision for the vectors of rounded features, while avoiding reduction in the classification quality.

A trained classifier leans heavily on its training datasets, while AE signals acquired from a pipeline are prone to variation because the inner flow rate and pressure change constantly. The signals also fluctuate according to the sensor installation location and the operating moment. To reconcile these differing environments, we must adjust the trained

leak detection model to its real and specific operational conditions. Therefore, the paper proposes updating the classifier by modifying the two parameters μ_{yn} and σ_{yn} related to the normal pipeline state in run-time, and which are employed in (5). Figure 8 shows the feature calculation and k -NN classification module of a real-time gas pipeline leak detection program implemented on the 32F746G-DISCOVERY board.

```

/* STE feature */
STE = 0;
for (i = 0; i < N; i++)
    STE += (x[i] * x[i]);
end
/* STD feature */
a = mean(x, N);
STD = 0;
for (i = 0; i < N; i++)
    STD += (x[i] - a) * (x[i] - a);
end
STD = sqrt(STD / (N - 1));

/* AVA feature */
AVA = 0;
for (i = 0; i < N; i++)
    AVA += abs(x[i]);
end
AVA = (AVA / N);
/* RMS feature */
RMS = 0;
for (i = 0; i < N; i++)
    RMS += (x[i] * x[i]);
end
RMS = sqrt(RMS / N);

/* Normalization */
y = [STE, AVA, RMS, STD]; // a feature vector
y = floor(10*(y-μyn)/σyn);
/* k-NN classification */
// Manhattan distance calculation
dist = dcal_func(v_train, M, y);
// Search for k minimum Manhattan distances
NN_index = find_nn_func(dist, k);
// Predict class
[n_NORM, n_LEAK] = count_func(label, NN_index);
if (n_NORM > n_LEAK) class = NORM;
else class = LEAK; end

```

/* x is an array of N raw AE signal samples; v_train is an array of M training feature vectors loaded from the sdcard; μ_{yn} and σ_{yn} are mean and standard deviation of normal state features, updated frequently in run-time */

Figure 8. Primary program module of real-time gas pipeline leak detection embedded in the 32F746G-DISCOVERY board.

5. Experimental Results

To evaluate the gas pipeline leak detection system quickly, we emulate a real data acquisition device (DAQ) using a computer program which dispatches recorded AE signal datasets, whose description is shown in Table 2, through an available communication channel to the 32F746G-DISCOVERY board. This does not affect the objective assessment because the datasets have been acquired from a practical pipeline testbed under various operating conditions. We here figure out three aspects: detection accuracy, real-time characteristic, and detection robustness, because those are key factors to apply a leak detection system for the real environment.

5.1. Detection Accuracy and Real-Time Characteristic

Figure 9 shows confusion matrices of experimental results returned by the leak detection program running on the 32F746G-DISCOVERY board, and Table 4 illustrates classification accuracy and execution time for evaluation scenarios. The accuracy, as averaged over the two sensor channels (R15i Ch1 and Ch2), and that of various pipeline states (L_0 , L_1 , L_2 , and L_3), is relatively high at better than 98% for every pressure condition (P_0 , P_1 , and P_2). Besides, the mean execution time ($t_E = 109$ s) is less than the total experimental dataset duration ($t_D = 123$ s). This demonstrates the real-time characteristic of the implemented detection system that does not miss any data and returns a timely result during the analysis operation. Furthermore, the ALEOR is monitored while examining dataset pairs (L_0, L_1), (L_0, L_2), and (L_0, L_3) subsequently (see Figure 10). This plot reveals the correct identification of pipeline states: normal (L_0), leaking (L_1, L_2 , and L_3), exploiting a threshold $\gamma = 10$ (see red dash line in Figure 10). The leaking state is decided only if ALEOR exceeds the threshold, despite the fluctuation below it. Therefore, no false alarm is reported in the experiment and the leaking state is also indicated punctually.

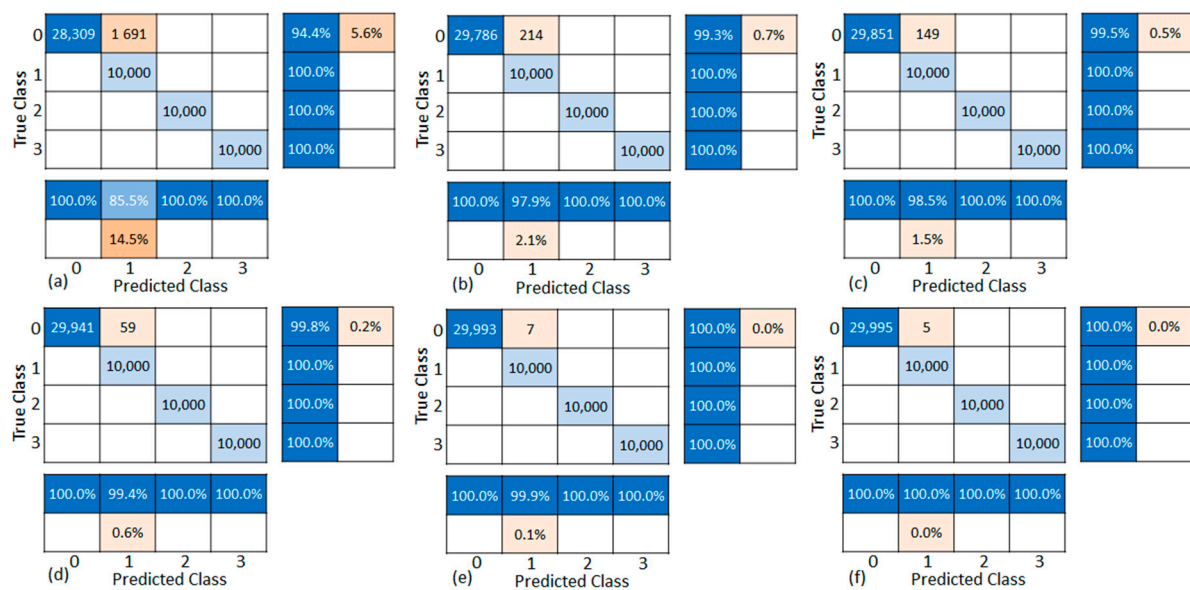


Figure 9. Confusion matrices resulting from experimental scenarios: R15i Ch1 (a) P₀, (b) P₁, (c) P₂; R15i Ch2 (d) P₀, (e) P₁, (f) P₂ (classes 0, 1, 2, 3, and 4 are L₀, L₁, L₂, and L₃, respectively).

Table 4. Classification accuracy and execution time.

		P ₀			P ₁			P ₂		
		A	t _D	t _E	A	t _D	t _E	A	t _D	t _E
R15i Ch1	L ₀	97.2	246	214	99.7	246	214	99.8	246	214
	L ₁	92.8	82	74	99.0	82	74	99.3	82	74
	L ₂	100	82	74	100	82	74	100	82	74
	L ₃	100	82	74	100	82	74	100	82	74
R15i Ch2	L ₀	99.9	246	214	100	246	214	100	246	214
	L ₁	99.7	82	74	100	82	74	100	82	74
	L ₂	100	82	74	100	82	74	100	82	74
	L ₃	100	82	74	100	82	74	100	82	74
Average		98.7	123	109	99.8	123	109	99.9	123	109

Where t_D and t_E are the total time of datasets and execution time, respectively, measured in seconds. A is classification accuracy given by: $A = 100 \times N_C / N_{FE}$ [%], N_C is the number of correctly classified frames.

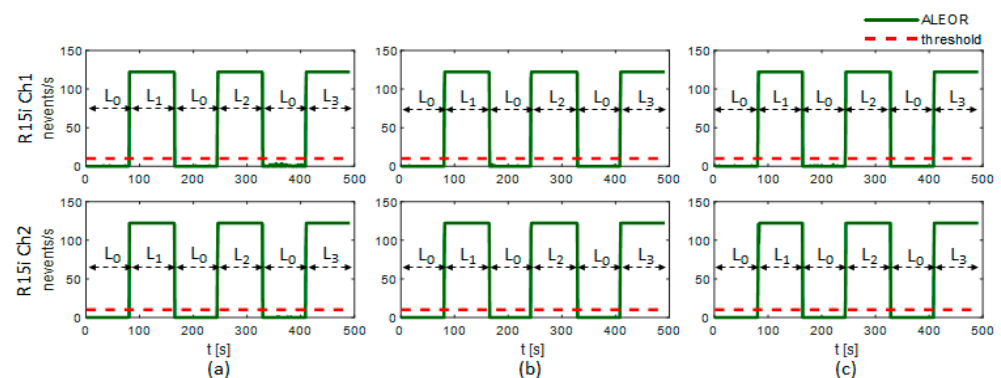


Figure 10. ALEOR under different pressure conditions: (a) P₀, (b) P₁, (c) P₂.

5.2. Detection Robustness

The result as exhibited in Table 4 and Figures 9 and 10 is obtained by using the test datasets under the same recording condition as the training datasets. As a result, the effectiveness of the proposed leak detection system may not be adequately demonstrated,

because in a real gas pipeline network, there are always irregular disturbances leading to AE signal modifications, such as operating mode variation (inner pressure or flow rate), noise interference, etc. Measurement of an AE sensor can be modelled as follows:

$$z = x + \eta \quad (6)$$

where z and x are measured and original signals, respectively, and η represents any signal modification including ambient noise and discrete events. We assume the normal distribution function for both x and η . According to the probability rule specified by [30], z distributes normally also, and its mean and standard deviation are sequentially:

$$\mu_z = \mu_x + \mu_\eta; \sigma_z = \sqrt{\sigma_x^2 + \sigma_\eta^2} \quad (7)$$

where $\mu_z, \sigma_z, \mu_x, \sigma_x, \mu_\eta, \sigma_\eta$ are means and standard deviations of z, x , and η , respectively. Equation (7) shows that the abnormal disturbance distorts original signals, thus deteriorating the signal-based leak detection model.

To verify the robustness of the proposed leak detection method, we add white noise to the experimental datasets prior to conducting the real-time leak detection on the 32F746G-DISCOVERY board. This noise is referred to as the signal disturbance η , simulated by an available function in the Matlab software with a rule below:

$$\mu_\eta = 0; \sigma_\eta = \rho \times \sigma_{xn} \quad (8)$$

where σ_{xn} is the standard deviation of normal state signal (acquired when the pipeline is healthy), and ρ is a proportion ratio. We set $\mu_\eta = 0$ in (8) because the mean parameter of a signal is mainly related to low frequency components of that signal, while the operating frequency range of R15i sensors is from 50 kHz to 400 kHz. The low frequency band (below 50 kHz) is not examined and the influence of μ_η is therefore relatively minor or $\mu_\eta \approx 0$. Figure 11 illustrates the signal distortion if adding a white noise η according to (6) and (8) where $\rho = 2$. We can easily realize that the distorted signal energy is greater than the original because of the added noise in Figure 11.

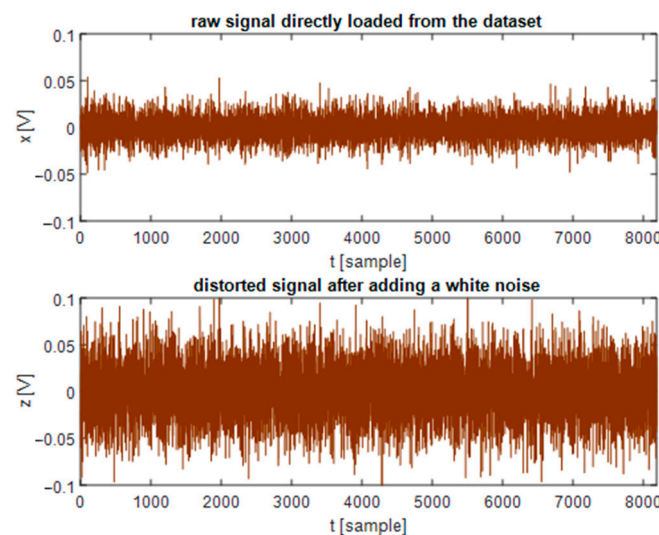


Figure 11. A signal after adding a white noise with $\rho = 2$.

We alter ρ and observe the performance deterioration of the trained classifier. Figure 12 shows the dependence of receiver operating characteristic (ROC) and average classification accuracy (ACA) on ρ . The computation is calculated on all the datasets of the two sensor channels in two cases: with updating μ_{yn} and σ_{yn} (see Section 4.2.2) and without updating. The classification performance substantially declines at slight values of ρ if we do not

adapt the model to the increasing added white noise (see Figure 12a and the blue dash dot line in Figure 12c). In contrast, the classifier can still work acceptably until $\rho = 70$ if we adjust μ_{yn} and σ_{yn} (see Figure 12b and the red solid line in Figure 12c). With $\rho = 10$, the resulting classification accuracy is above 90% (see Figure 12c) and the pipeline state can be exactly identified by the ALEOR with a threshold $\gamma = 10$, as shown in Figure 13 for every experimental condition. In short, the proposed methodology can ensure the robustness of the leak detection system.

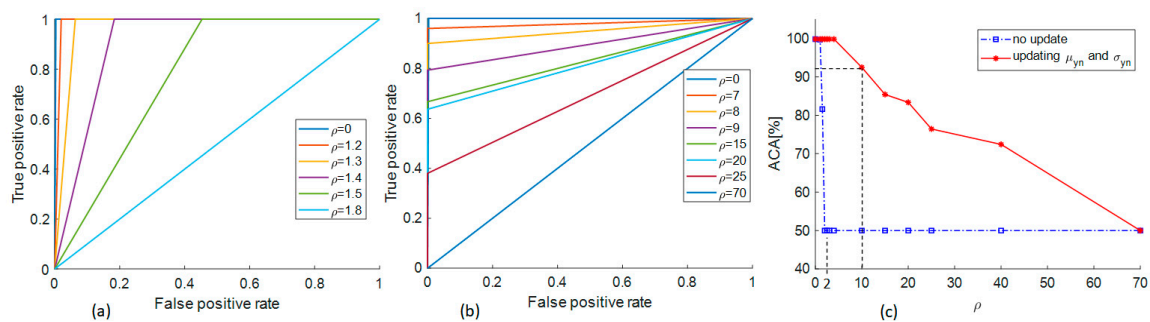


Figure 12. ROC and ACA according to ρ : (a) ROC without updating μ_{yn} and σ_{yn} , (b) ROC with updating μ_{yn} and σ_{yn} , (c) ACA reduction.

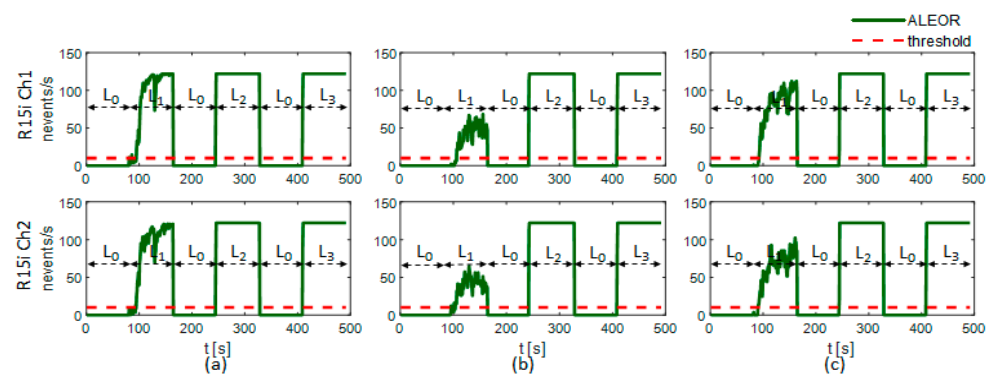


Figure 13. ALEOR under different pressure conditions: (a) P_0 , (b) P_1 , (c) P_2 (After adding a white noise with $\rho = 10$).

Although the proposed method can sustain a high classification performance with small values of ρ , the classification performance still deteriorates gradually according to the increase in ρ and the classifier cannot precisely operate with $\rho > 70$ which causes severe distortion of the acquired signals. Therefore, we should configure the testbed to resemble an applied real pipeline before gathering datasets for training the classifier, thus obtaining an adequate leakage detector. The greater the similarity between the testbed and the real pipeline, the more accurate the detection is.

6. Conclusions

A complete system is offered for real-time gas pipeline leak detection in the paper. First, the system offline analyzed recorded AE signals sampled at 1 MHz. The process configured a hybrid feature pool and normalized its elements using the mean and standard deviation of the set of feature observations related to normal pipeline health. Then, the pool was purified using the three-sigma rule and the Kullback–Leibler distance to obtain the most discriminative signatures. Next, the system identified the pipeline health states (normal/leaking) with an input vector of features, by exploiting a k -nearest neighbor classifier that seeks the purified feature pool for the signatures closest to the input vector, based on the Manhattan distance. To avoid issuing a false alarm, the system decided a pipeline state via monitoring the accumulative leaking event occurrence rate and a

predefined threshold. Finally, the total proposed leak detection method was embedded in a compact MCU-based hardware platform for real-time leak detection. The detection accuracy, the real-time characteristic, and the robustness of the introduced gas pipeline leak detection system have been evaluated. The experimental results showed that the system indicated pipeline health states robustly in a quick enough timeframe for real-time application. Thus, this system can be applied for inspecting pipeline health in a real gas pipeline network.

The testbed used in this paper for collecting AE signals is a part of a real gas pipeline network. Hence, the resulting AE signals are not simple signals generated by the pipeline leakage simulation in a laboratory. They do not only contain information about pipeline states (normal or leaking), but also depend on practical gas transportation and systematic behavior. Additionally, a noisy measurement location and wave attenuation could conceal symptoms of leakage in recorded signals. This challenges the signal investigation because the relation between the leakage phenomenon and AE signals is unclear in the initial analysis stages. Therefore, a short pipeline was chosen in the paper to easily separate signal classes related to pipeline states corresponding to different experimental scenarios, hence conveniently proposing a leak detection method as well as evaluating experimental results. However, it is believed that the proposed technique can effectively monitor a long pipeline in a real application. The pipeline length depends on the signal detection ability of the AE sensor—their sensitivity and a specific working environment. These parameters can be estimated by using pencil lead breaking tests.

Author Contributions: Conceptualization, T.B.Q. and J.-M.K.; data curation, T.B.Q.; formal analysis, T.B.Q.; funding acquisition, J.-M.K.; methodology, T.B.Q. and J.-M.K.; software, T.B.Q.; supervision, J.-M.K.; validation J.-M.K.; visualization, T.B.Q.; writing—original draft, T.B.Q.; and writing—review and editing, J.-M.K. All authors have read and agreed to the published version of 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. 20192510102510).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

References

1. McAllister, E.W. (Ed.) *Pipeline Rules of Thumb Handbook: Quick and Accurate Solutions to Your Everyday Pipeline Problems*, 7th ed.; Elsevier: Amsterdam, The Netherlands; Gulf Professional: Boston, MA, USA, 2009; ISBN 978-1-85617-500-5.
2. Antaki, G.A. *Piping and Pipeline Engineering: Design, Construction, Maintenance, Integrity, and Repair*; CRC Press: Boca Raton, FL, USA, 2003; ISBN 978-0-203-91115-0.
3. EGIG. *Gas Pipeline Incidents*; European Gas Pipeline Incident Data Group: Groningen, The Netherlands, 2018.
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. Miller, R.K.; Hill, E.K.; Moore, P.O.; American Society for Nondestructive Testing (Eds.) *Acoustic Emission Testing*, 3rd ed.; Nondestructive Testing Handbook; American Society for Nondestructive Testing: Columbus, OH, USA, 2005; ISBN 978-1-57117-106-1.
6. Li, Z.; Zhang, H.; Tan, D.; Chen, X.; Lei, H. A novel acoustic emission detection module for leakage recognition in a gas pipeline valve. *Process Saf. Environ. Prot.* **2017**, *105*, 32–40. [[CrossRef](#)]
7. Xu, Q.; Zhang, L.; Liang, W. Acoustic detection technology for gas pipeline leakage. *Process Saf. Environ. Prot.* **2013**, *91*, 253–261. [[CrossRef](#)]
8. Jin, H.; Zhang, L.; Liang, W.; Ding, Q. Integrated leakage detection and localization model for gas pipelines based on the acoustic wave method. *J. Loss Prev. Process Ind.* **2014**, *27*, 74–88. [[CrossRef](#)]
9. Song, Y.; Li, S. Leak detection for galvanized steel pipes due to loosening of screw thread connections based on acoustic emission and neural networks. *J. Vib. Control* **2018**, *24*, 4122–4129. [[CrossRef](#)]

10. Xiao, R.; Hu, Q.; Li, J. Leak detection of gas pipelines using acoustic signals based on wavelet transform and support vector machine. *Measurement* **2019**, *146*, 479–489. [[CrossRef](#)]
11. Sun, J.; Xiao, Q.; Wen, J.; Zhang, Y. Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis. *Measurement* **2016**, *79*, 147–157. [[CrossRef](#)]
12. Sun, J.; Xiao, Q.; Wen, J.; Wang, F. Natural gas pipeline small leakage feature extraction and recognition based on LMD envelope spectrum entropy and SVM. *Measurement* **2014**, *55*, 434–443. [[CrossRef](#)]
13. Zhu, S.-B.; Li, Z.-L.; Zhang, S.-M.; Liang, L.-L.; Zhang, H.-F. Natural gas pipeline valve leakage rate estimation via factor and cluster analysis of acoustic emissions. *Measurement* **2018**, *125*, 48–55. [[CrossRef](#)]
14. Datta, S.; Sarkar, S. A review on different pipeline fault detection methods. *J. Loss Prev. Process Ind.* **2016**, *41*, 97–106. [[CrossRef](#)]
15. 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*; European Standard: Brussels, Belgium, 2010.
16. Baroudi, U.; Al-Roubaiey, A.A.; Devendiran, A. Pipeline leak detection systems and data fusion: A survey. *IEEE Access* **2019**, *7*, 97426–97439. [[CrossRef](#)]
17. Gholizadeh, S.; Leman, Z.; Baharudin, B.T.H.T. A review of the application of acoustic emission technique in engineering. *Struct. Eng. Mech.* **2015**, *54*, 1075–1095. [[CrossRef](#)]
18. Wang, L.; Gao, X.; Liu, T. Gas pipeline small leakage feature extraction based on LMD envelope spectrum entropy and PCA-RWSVM. *Trans. Inst. Meas. Control* **2016**, *38*, 1460–1470. [[CrossRef](#)]
19. Xiao, Q.; Li, J.; Sun, J.; Feng, H.; Jin, S. Natural-gas pipeline leak location using variational mode decomposition analysis and cross-time–Frequency spectrum. *Measurement* **2018**, *124*, 163–172. [[CrossRef](#)]
20. He, P. Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law. *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.* **1998**, *45*, 114–125. [[CrossRef](#)]
21. Rienstra, S.W.; Hirschberg, A. An Introduction to Acoustics. *Eindh. Univ. Technol.* **2004**, *18*, 19.
22. Yan, Y.; Shen, Y.; Cui, X.; Hu, Y. Localization of multiple leak sources using acoustic emission sensors based on MUSIC algorithm and wavelet packet analysis. *IEEE Sens. J.* **2018**, *18*, 9812–9820. [[CrossRef](#)]
23. Dong, L.; Qiao, Z.; Wang, H.; Yang, W.; Zhao, W.; Xu, K.; Wang, G.; Zhao, L.; Yan, H. The gas leak detection based on a wireless monitoring system. *IEEE Trans. Ind. Inf.* **2019**, *15*, 6240–6251. [[CrossRef](#)]
24. Chraim, F.; Bugra Erol, Y.; Pister, K. Wireless gas leak detection and localization. *IEEE Trans. Ind. Inf.* **2016**, *12*, 768–779. [[CrossRef](#)]
25. Pukelsheim, F. The Three Sigma Rule. *Am. Stat.* **1994**, *48*, 88–91. [[CrossRef](#)]
26. Koutroumbas, S.T.K. *Pattern Recognition*, 4th ed.; Academic Press: Cambridge, MA, USA, 2008.
27. Mistras Group. *R151-AST Sensor*; Mistras Group, Inc.: Princeton Junction, NJ, USA, 2015.
28. Oppenheim, A.V.; Schafer, R.W. *Discrete-Time Signal Processing*, 3rd ed.; Pearson: Upper Saddle River, NJ, USA, 2010; ISBN 978-0-13-198842-2.
29. Ranganayakulu, S.V.; Goud, B.S.; Sastry, P.V.; Kumar, B.R. Calibration of acoustic emission system for materials characterization. *Univers. J. Mater. Sci.* **2015**, *3*, 62–69. [[CrossRef](#)]
30. Lemons, D.S.; Langevin, P. *An Introduction to Stochastic Processes in Physics: Containing “On the Theory of Brownian Motion” by Paul Langevin, Translated by Anthony Gythiel*; Johns Hopkins University Press: Baltimore, MD, USA, 2002; ISBN 978-0-8018-6866-5.