

Article

# Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks

Yang Liu <sup>1</sup>, Xuehui Ma <sup>1</sup>, Yuting Li <sup>1</sup>, Yong Tie <sup>1,\*</sup>  and Yinghui Zhang <sup>1,2</sup> and Jing Gao <sup>3</sup>

<sup>1</sup> College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China; yangliu@imu.edu.cn (Y.L.); xuehuimaimu@163.com (X.M.); yutingli@mail.imu.edu.cn (Y.L.); zhangyinghui\_imu@163.com (Y.Z.)

<sup>2</sup> Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

<sup>3</sup> Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China; jing401@126.com

\* Correspondence: yongtieimu@163.com

Received: 23 October 2019; Accepted: 19 November 2019; Published: 21 November 2019



**Abstract:** The detection of water pipeline leakage is important to ensure that water supply networks can operate safely and conserve water resources. To address the lack of intelligent and the low efficiency of conventional leakage detection methods, this paper designs a leakage detection method based on machine learning and wireless sensor networks (WSNs). The system employs wireless sensors installed on pipelines to collect data and utilizes the 4G network to perform remote data transmission. A leakage triggered networking method is proposed to reduce the wireless sensor network's energy consumption and prolong the system life cycle effectively. To enhance the precision and intelligence of leakage detection, we propose a leakage identification method that employs the intrinsic mode function, approximate entropy, and principal component analysis to construct a signal feature set and that uses a support vector machine (SVM) as a classifier to perform leakage detection. Simulation analysis and experimental results indicate that the proposed leakage identification method can effectively identify the water pipeline leakage and has lower energy consumption than the networking methods used in conventional wireless sensor networks.

**Keywords:** leakage detection; wireless sensor networks; machine learning; leakage triggered networking

## 1. Introduction

Water provides a material basis for human life and the survival of all living things, and it is an indispensable natural resource needed for the development of human society. Due to population growth, economic development, and changing consumption patterns, the demand for water resources is increasing rapidly, and this increasing demand is expected to accelerate greatly during the next 20 years [1]. However, the wastage of water resources caused by leakage from water pipelines is an important problem [2]. A study conducted by the World Bank indicated that the leakage from water pipeline exceeds 48.6 billion m<sup>3</sup> annually, and the corresponding annual economic losses are approximately US\$14.6 billion [3]. According to the 2012 Green City Index statistics, water pipeline leakage rates exceed 10% in one third of countries worldwide. For example, the average water pipeline leakage rate is 23% in the EU, 13% in the U.S. and Canada, 22% in Asia, 35% in Latin America, and 30% in Africa [4,5]. Accordingly, research on high performance water pipeline leakage detection technologies has great significance for the protection of water resources and the promotion of economic development.

The main causes of water pipeline leakage include the corrosive nature of the soil, deficient pipe material quality, the temperature and pressure, failure to employ standard pipe laying methods, geological changes, and human damage [6–8]. For instance, deterioration caused by the pressure of the soil may lead to uneven pipeline stress bearing, which can result in leakage from ruptures and breakage. Leakage may also result from damage to pipe junctions caused by external impacts or ruptures caused by the failure to weld junctions during construction. Nevertheless, because the vast majority of water pipelines are located deep underground, leakage is typically not promptly discovered, and when the amount of leakage is large, it is usually not discovered until water begins flowing from the ground surface. Furthermore, since most leaks are relatively small or undetectable, leakage causes substantial waste of water resources [9,10]. Therefore, it is extremely important to detect leakage from underground water pipelines effectively and accurately.

Academic researchers and the industry have conducted major research campaigns and developed numerous effective detection methods. One of the earliest detection methods is the listening system in which detection personnel listen to changes in the volume and sound quality of leakage noise coming from equipment and locate areas of leakage based on these observations [11,12]. This method not only depends on the experience of the detection personnel, but is highly labor intensive and unreliable due to the large areas of water pipeline networks. Ground penetrating radar can determine the locations of pipeline leaks through the detection of soil voids caused by water leakage. However, because of the complex differences in the geological structure in different areas, this method has poor applicability and is very expensive [13,14]. Other studies have presented leakage detection inspired by the changes in the internal pressure of a water pipeline, such as the pressure gradient method, negative pressure wave method, and flow rate balance method [15–17]. Although these methods are relatively sensitive to a pipeline's flow rate and pressure, they tend to give false positive results when the flow rate fluctuations are large because the flow rate in water pipeline networks will fluctuate continuously. It has been found that the spectrum of leakage signals is concentrated, and the pipeline vibration frequency is correlated with the leakage state [18]. This characteristic can be employed in conjunction with the spectral analysis of signals from piezoelectric accelerometers to perform leakage detection. However, when the environmental noise has a similar frequency spectrum to that of the leakage signals, this method also has a tendency to give false positives. Another study used the linear predictive coding coefficient (LPCC) of acoustic leakage signals and a hidden Markov model (HMM) to improve the ability to distinguish the leakage signal from interfering environmental signals [19]. However, the detection error of this method tends to increase with the length of time that the system is in use. Furthermore, in [20–22], some detection methods based on acoustic signals, such as through hydrophones, have proven that effectively leak detection can be achieved in a longer range of detection in water pipelines. These methods exhibit a high sensitivity to leaks in the range of interest, appearing appropriate for early leak detection. Additionally, the works in [23–25] have applied machine learning to pipeline monitoring and have achieved excellent results. Moreover, research on large scale water pipeline network systems has used methods such as real-time modeling to compare measured pipeline network data with flow rate model predictions [26–29], but problems such as high modeling difficulty and high computational loads in real applications limit the use of these methods. Researchers have achieved impressive results in research in the field of tap water pipeline leakage detection and location. However, the large scale of water pipeline networks and the extreme complexity of the network architectures, environmental conditions, and geological factors still pose significant challenges for the real-time monitoring of water supply pipeline networks and intelligent leakage detection.

Wireless sensor networks (WSNs) provide an efficient way to address these issues. Due to their outstanding sensing ability, communication protocols, processor speed, and data collection advantages, wireless sensor network technologies have been widely applied in the field of monitoring [30–34]. In [35], PipeNet, a system based on wireless sensor networks, was proposed. It aims to monitor water flow and detect leaks by attaching acoustic and vibration sensors to large bulk-water pipelines and pressure sensors to normal pipelines. In contrast to the PipeNet project, the PipeProbe

system does not assume that water pipe surfaces are exposed and accessible for sensor module attachment [36]. PipeProbe can be dropped into the source of a water pipeline. During its traversal of the pipeline, it collect the sensor readings necessary for the reconstruction of the 3D spatial layout of the traversed water pipelines. To demonstrate the application and control of a low cost wireless sensor network for a high data rate, WaterWiSe@SG, a wireless sensor network to enable real-time monitoring of a water distribution network in Singapore, was proposed in [37]. The goal of WaterWiSe@SG was to develop generic wireless sensor network capabilities to enable real-time monitoring of a water distribution network.

This paper proposes a water pipeline monitoring system based on wireless sensor networks and a leakage identification method based on support vector machine (SVM). Machine learning can simulate the acquisition of knowledge through human learning activities and can enable the automatic improvement of system performance. As a result, it is widely applied in speech and biological affect identification [38], physiological signal detection [39], body movement identification [40], signal feature detection and identification, etc. [41–44]. The proposed system employs ZigBee nodes serving as signal collection nodes and uses the 4G network to transmit the signals adopted by the sensors to the data processing center for processing. To address the high networking power consumption that affects conventional wireless sensor networks, we also propose a leakage triggered networking method able to network and perform data transmission from wireless sensor nodes in the vicinity of leakage points, which effectively reduces the network energy consumption and extends its life cycle. Based on the differences in the time-frequency features of leakage and non-leakage signals, we propose a leakage detection method that constructs feature matrices by employing the intrinsic mode function, approximate entropy, and principal component analysis (PCA) and that uses SVM as a classifier to identify the leaks. Our experiment is performed along an exposed aluminum-plastic composite pipe, with a diameter of 27 mm. The CT1010 acceleration sensor is used in this experiment due to its sensitivity. During our experiments, the water pressure is no less than 0.3 MPa. The detectable leaking flow rate is calculated to be approximately  $2.5 \text{ cm}^3/\text{s}$  based on the pipe pressure. Experimental and simulation results demonstrate that the proposed methods can effectively detect the leakage and prolong the lifetime of the wireless sensor network.

The rest of the paper is organized as follows. The implementation of the leakage monitoring system is introduced in Section 2. In Section 3, the leakage triggered networking solution for wireless sensor networks that are able to network sensors receiving leakage signals is developed. We propose a leakage detection method based on time-frequency features and SVM in Section 4. Experimental and simulations results are reported in Section 5. Conclusions are presented in Section 6.

## 2. Water Pipeline Leakage Monitoring System Based on ZigBee Technology

We designed a water pipeline monitoring system based on a ZigBee and 4G wireless communication system. It included a ZigBee wireless network, a gateway, and a data processing center. The system architecture is shown in Figure 1. The ZigBee wireless sensor was constructed from terminal nodes, coordinators, and routers. The terminal node employed an acoustic sensor to acquire signals from pipeline leakage, and the coordinator relied on serial ports and a gateway to establish a linkage and maintain the two-way transmission of data and control commands between the terminal nodes and the gateway. The gateway also took advantage of the 4G network to upload data collected by the sensors to the host for storage and processing.

In this system, the ZigBee nodes employed CC2530, and the Zstack protocol was used for data transmission. The system's gateway employed STM32F103VET6 as its chief control element and used a built-in 4G module to perform networking functions. The gateway and ZigBee coordinator communicated via a serial port, which enabled the data processing center to monitor and control the ZigBee wireless sensor network. The gateway used its built-in multithreading TCP server to monitor requests from the host and transmit data after establishing a connection. The host's functions included display of the remote pipeline monitoring system, interactive controls, and signal processing.

The software on the host was developed using C# and SQL, and the TCP protocol and TCP server on the gateway established network connections and received control commands and data transmissions from the ZigBee wireless sensor network. The CC2530 has a nominal maximum operating distance of 75 m, and research has indicated that ZigBee nodes have a reliable operating range of 60 m in a line-of-sight (LOS) environment [45,46]. The data acquisition card used in this paper was NI’s MCC BTH-1208LS, which was used to save data to the host computer for analyzing signals. CT1010 acceleration sensor was used in this experiment due to its sensitivity, which was matched with data acquisition card. Figure 2 shows the results of testing the packet loss rate of the system’s terminal nodes in a non-line-of-sight (NLOS) environment as the interval between the nodes changes. To ensure the reliability of the system’s data transmissions, the intervals between the CC2530 nodes should be kept at approximately 30 m.

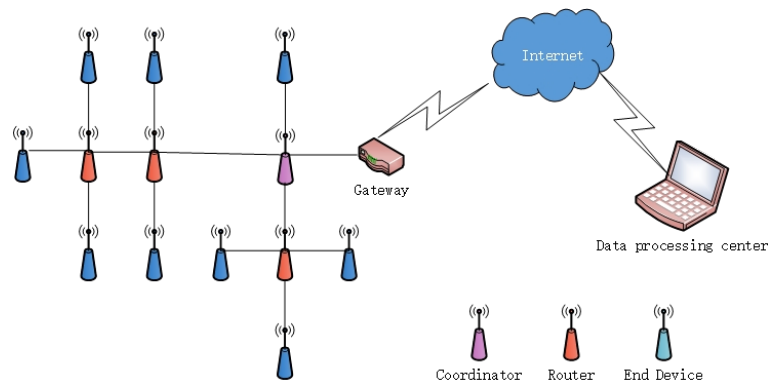


Figure 1. System architecture for pipeline monitoring.

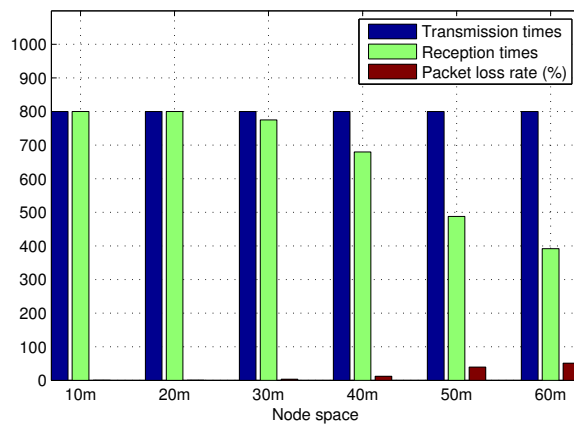


Figure 2. CC2530 packet loss rate under NLOS conditions.

### 3. Leakage Triggered ZigBee Networking

In actual water pipeline monitoring environments, a large number of terminal sensors needs to be installed along pipelines. Since the pipeline leakage is a low probability event, all of the sensors working at the same time will result in a significant waste of energy. Furthermore, leakage in water supply pipelines occurs randomly, and pipeline systems must be monitored in real time. To reduce energy consumption and increase network lifetime, we propose a leakage triggered networking method.

The ZigBee networking method includes initialization networking and triggered networking. The topological structure of the network provides an important basis for ZigBee networking; in view of the structural characteristics and distribution of water supply pipelines, this paper employed a network topology. To achieve initialization networking, the first step was to determine the coordinator nodes and set their signal channels and network ID numbers, which would initialize the network.

Non-coordinator nodes were then added to the network. Figure 3 mainly interprets how the nodes join the network. To ensure that the number of terminal nodes installed at each relay node (i.e., router node) was balanced, it would add received signal strength indicator (RSSI) information to each Beacon\_request frame when each terminal node in this solution sends a network join request. In accordance with the RSSI values, routing nodes provided joining service to the terminal nodes. To collect leakage signals effectively and reliably, it is necessary to set reasonable threshold values for the RSSI. If the RSSI values of the terminal nodes are smaller than the threshold, the routing nodes will not process the request of terminal nodes. If the RSSI values are greater than the threshold, the routing nodes will record the terminal nodes' information and ensure that they can join the network.

After the network was initialized, to reduce the network power consumption and prolong the network lifetime, this paper designed three types of control frames (i.e., join frame, active frame, and wave frame), which can be used to trigger the network according to the leakage detection results. The structures of the join frame, active frame, and wave frame are shown in Tables 1–3. The Sou\_address field is the original address, which is the routing node address. The nodes rely on this field to determine whether they have been activated and are working. The Des\_address field is the destination address and constitutes the address of a routing node that has joined the network. The PANID is the network address. In the join frames, the join\_result field is the result of networking; a result of zero indicates that the networking of a routing node has failed, and a result of one indicates that it has been successful. In the active frame, the Act\_address field is the active address, and child nodes will join the network after receiving this information and then perform data sampling. In the wave frames, the Position field records the actual physical address information. Because this paper mainly considered the networking solution and its performance, the Position field was not used for node location.

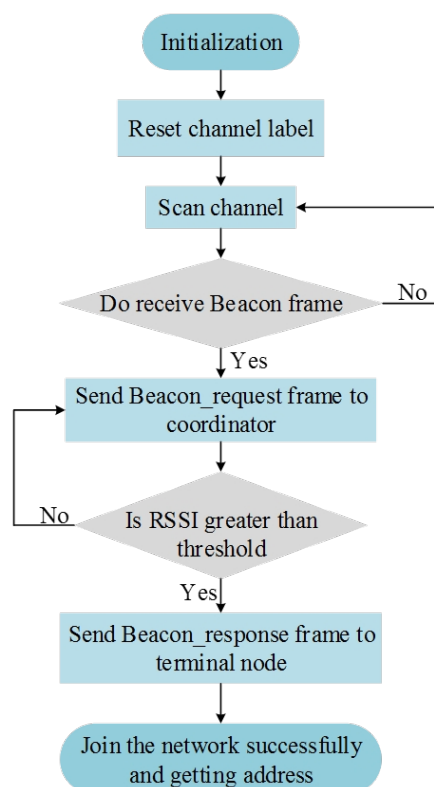


Figure 3. Flowchart of the nodes joining the network.

**Table 1.** Join frame.

Field Name	PANID	Des_address	Sou_address	Join_result	Channel
Length	16 bits	16 bits	16 bits	1 bit	32 bits
Instructions	Network ID	Destination	Source	Results	Channel

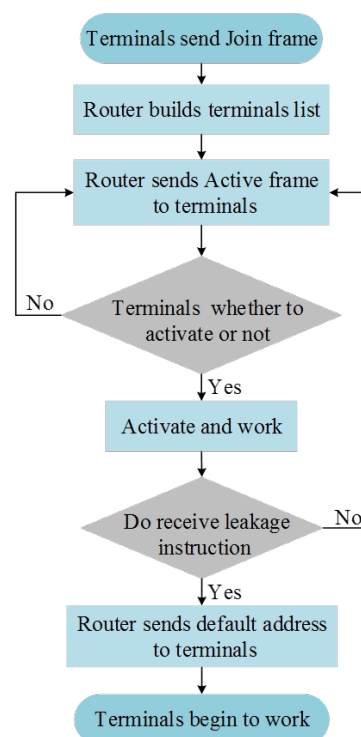
**Table 2.** Active frame.

Field Name	PAN ID	Sou_address	Act_address
Length	16 bits	16 bits	16 bits
Instructions	Network ID	Source	Active Node

**Table 3.** Wave frame.

Field Name	PAN ID	Des_address	Sou_address	Position	Value
Length	16 bits	16 bits	16 bits	64 bits	128 bits
Instructions	Network ID	Destination	Source	Leakage Point	Information

Figure 4 shows the leakage triggered networking flowchart. The terminal nodes send Join frames to the routing nodes to compile a list of terminal nodes. After the routing-terminal relationship has been constructed, the routing nodes will send Active frames to the terminal nodes. When receiving Active frames, the terminal nodes will then determine whether they are the destination nodes. If a node is a destination node, it will activate itself and perform data sampling. If a node is not a destination node, it will enter a sleep monitoring state and wait for the next Active frame. Once a node in the working state determines that a signal it has received is a leakage signal, the routing node will send the preset leakage triggered address to the nodes on the terminal nodes list. All of the nodes on the list receiving that address will activate themselves and initiate signal sampling and transmission. This method enables routing nodes to control the working status of the terminal nodes. After the networking is completed, data transmission is performed using the ZigBee routing protocol.

**Figure 4.** Flowchart of the ZigBee network leakage triggered networking.



## 4. Leakage Detection by Using Machine Learning and Time-Frequency Features

### 4.1. Time-Frequency Analysis of the Acoustic Leakage Signal

#### 4.1.1. Spectrum Density Feature

Many studies have indicated that the components of the spectra of leakage signals are chiefly concentrated within specific bands. As a result, the differences between signal spectra can be employed as pipeline leakage identification characteristics. To extract the differences of the signal spectral density, we used empirical mode decomposition (EMD) to perform a time-frequency analysis of pipeline signals and proposed a frequency domain feature for detection.

EMD can selectively decompose the signal as the sum of a finite number of intrinsic mode functions (IMFs) [47], enabling multiple IMFs to be used in the multiscale analysis of the spectrum density of leakage signals. An analytic function  $z(t)$  can be expressed as:

$$z(t) = x(t) + j\hat{x}(t) = a(t)e^{j\Phi(t)}, \quad (1)$$

where  $\hat{x}(t)$  is the Hilbert transform of a signal  $x(t)$  and  $\Phi(t)$  is given by:

$$\Phi(t) = \arctan \frac{\hat{x}(t)}{x(t)}. \quad (2)$$

Finding the derivative of the phase function with respect to time enables the signal analysis instantaneous frequency function to be determined as:

$$f(t) = \frac{1}{2\pi} \frac{d\Phi(t)}{dt}. \quad (3)$$

The definition of the instantaneous frequency shows that although the instantaneous frequency may take the form of a meaningless negative frequency under certain situations, if the instantaneous frequency is positive throughout a certain period of time, then the  $x(t)$  can be termed the IMF. Accordingly, an IMF must satisfy the following two conditions:

- (1) The number of extreme points is  $N_e$  (including the minimum and maximum values), which is the same as or no more than one from the number of zero crossing points  $N_s$ ,

$$(N_s - 1) \leq N_e \leq (N_s + 1). \quad (4)$$

- (2) At an arbitrary time  $t_i$  within the time period, the mean of the upper envelope determined by the local maximum and the lower envelope determined by the local minimum is zero,

$$[x_{max}(t) + x_{min}(t)]/2 = 0, \quad t_i \in [t_a, t_b]. \quad (5)$$

Generally, a signal can include multiple IMFs. The EMD method can be used to extract the IMFs from a signal. To derive the spectrum density features of a leakage signal, we used the EMD method to process the signal, which will yield the IMFs of that signal. All the extreme points of the original signal  $x(t)$  are connected with a cubic spline curve, yielding the upper and lower envelopes of  $x(t)$ , which causes the signal to be contained between the two envelopes. We assumed that the function formed by means of the two envelopes is  $m(t)$ . Subtracting  $m(t)$  from the original signal  $x(t)$ ,

$$h_1(t) = x(t) - m(t). \quad (6)$$

We now check whether  $h_1(t)$  satisfies the two conditions of the IMF. If it does not satisfy them, repeat (6) until the IMF conditions are satisfied.  $h_1(t)$  at this time is expressed as  $c_1(t)$ , where  $c_1(t)$  is the first IMF of the signal  $x(t)$ ,

$$c_1(t) = h_1(t). \quad (7)$$

Next,  $c_1(t)$  is subtracted from the original signal  $x(t)$  to obtain the new signal  $r_1(t)$ ,

$$r_1(t) = x(t) - c_1(t). \quad (8)$$

Repeating (6) until the IMF conditions are satisfied, the first IMF of  $r_1(t)$  is obtained, which is the second IMF of  $x(t)$  and is denoted as  $c_2(t)$ . Continuing in the same manner, we can progressively derive the  $m^{\text{th}}$  IMF  $c_m(t)$  of the signal  $x(t)$  and the remainder  $r_m(t)$ .

Using the foregoing steps, the original signal  $x(t)$  can be decomposed into the sum of IMFs and a remainder,

$$x(t) = \sum_{i=1}^m c_i(t) + r_m(t). \quad (9)$$

In general, the IMF condition (2) is difficult to satisfy; thus, a stopping criterion is generally established. When the stopping criterion has been satisfied, Condition (2) can be considered to have been met. For this reason, we set the standard deviation between two consecutive processing results as  $S_d$ ,

$$S_d = \sum_{n=0}^{N-1} \frac{|h_{k-1}(n) - h_k(n)|^2}{h_k^2(n)}, \quad (10)$$

where  $N$  is the observed signal length and  $h_{k-1}(n)$  and  $h_k(n)$  are two consecutive processing results in the process of the IMF derivative. When the standard deviation  $S_d$  reaches the preset threshold value, we can assume that Condition (2) has been satisfied. It has been indicated in [47] that the threshold value of the standard deviation  $S_d$  is typically taken as 0.2–0.3.

After the IMFs of a signal have been obtained, we can further obtain the discrete Fourier transform  $C_i(k)$  of the IMF components  $c_i(n)$  resulting from EMD decomposition,

$$C_i(k) = \sum_{n=0}^{N-1} c_i(n) e^{-j \frac{2\pi}{N} kn}. \quad (11)$$

Deriving the modulus square of  $C_i(k)$  yields the IMF power spectrum of the signal,

$$P_i(k) = \frac{1}{N} |C_i(k)|^2. \quad (12)$$

We then obtain the mean of Equation (12),

$$\bar{P}_i = \frac{\sum_{k=0}^{N-1} |C_i(k)|^2}{N^2}. \quad (13)$$

This paper uses the mean value of the IMF power spectrum as the frequency domain feature of the leakage signal.

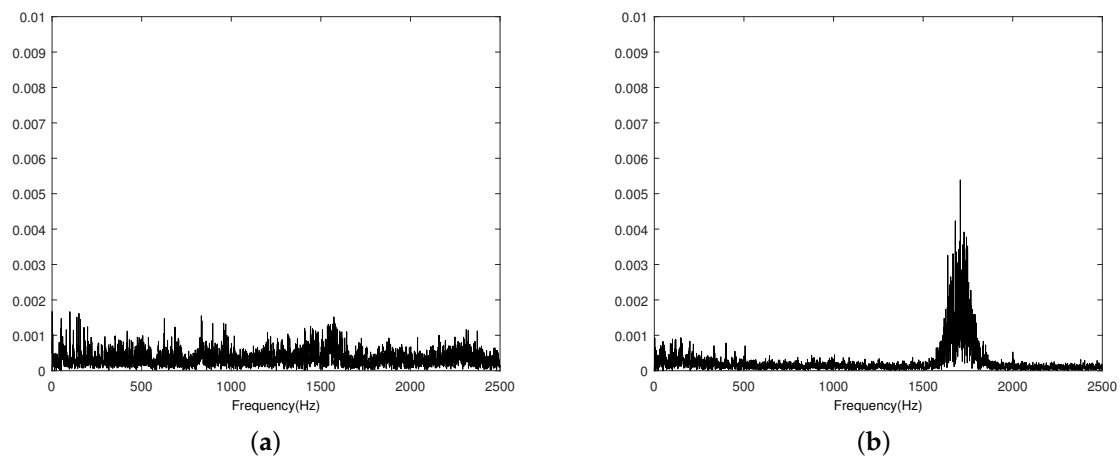
Figure 5 shows the power spectra of the signals from a pipeline with and without leakage. The results indicate that the frequencies of the acoustic signals leakage were chiefly concentrated near 1.6 kHz, which is consistent with previous research [48]. Figure 6 is the spectra of the first four groups of IMF components obtained by EMD of the pipeline leakage signal and the pipeline non-leakage signal. In this experiment, the threshold value of the standard  $S_d$  was set as 0.3. Comparing the IMF of each layer of the leakage signal and the non-leakage signal, it can be seen that the main spectrum of the signal was in the first layer IMF. The spectrum of the leakage signal mainly distributed between



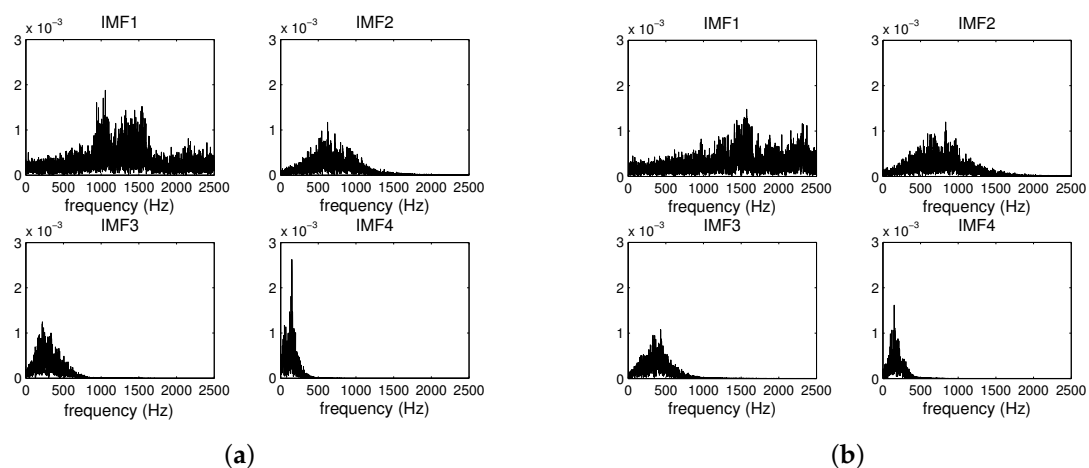
1000 Hz and 2000 Hz. The non-leakage signal spectrum was more random and mostly distributed over the entire band.

#### 4.1.2. Signal Complexity Feature

Because pipeline leakage is a localized and low probability event, there should be differences in the time domain composition of leakage signals and non-leakage signals, and the composition of leakage signals should be more complex. Accordingly, the differences of signal composition in time domain can be used to identify leakage signal. The approximate entropy (ApEn) is the conditional probability when similarity is maintained after the dimensions of a similarity vector are increased from  $m$  to  $m + 1$  and is the probability of the new mode when the number of dimensions changes [49]. The greater the probability of the new mode is, the more complex the signal and the greater the corresponding ApEn. Therefore, we used the average ApEn as a time domain feature to quantify signal complexity.



**Figure 5.** Power spectra of signals from a pipeline. (a) Normal signal power spectrum. (b) Leakage signal power spectrum.



**Figure 6.** The first four intrinsic mode function (IMF) spectra of the pipeline leakage and non-leakage signal. (a) Leakage signal. (b) Non-leakage signal.

For a sequence  $u(1), u(2), \dots, u(N)$ , two sequences with length  $m$  can be used to construct as  $x(i) = [u(i), u(i + 1), \dots, u(i + m - 1)]$  and  $x(j) = [u(j), u(j + 1), \dots, u(j + m - 1)]$ , where  $i, j \leq N - m + 1$ . We then calculate the distance between  $x(i)$  and  $x(j)$ ,

$$d[x(i), x(j)] = \max_{k=1,2,\dots,m} [|u(i + m - 1) - u(j + k - 1)|]. \tag{14}$$

Assuming a threshold value  $r$ , we determine the number of  $d[x(i), x(j)] \leq r$  (which is set as  $L$ ) for every  $i < N - m + 1$  and calculate the ratio of  $L$  and the number of vectors,

$$C_i^m(r) = \frac{L}{N - m + 1}. \tag{15}$$

For all  $i$  values, we derive the mean  $\phi^m(r)$  of  $\ln C_i^m(r)$ ,

$$\phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N - m + 1}. \tag{16}$$

Increasing  $m$  by one, we repeat Steps (14)–(16) to obtain  $\phi^{m+1}(r)$ , and in accordance with  $\phi^{m+1}(r)$  and  $\phi^m(r)$ , we can obtain the ApEn value as:

$$\text{ApEn}(m, r) = \phi^m(r) - \phi^{m+1}(r). \tag{17}$$

The results showed that the ApEn was a dimensionless scalar quantity, and its value was related to  $m$  and  $r$ . To ensure that the ApEn had reasonable statistical characteristics, based on experience,  $m = 2$  is usually employed, and  $r$  was set as 0.1–0.3-times the standard deviation (SD) of the sequence [49]. Figure 7 shows the ApEn of signals before and after leakage for different threshold values. In the experiment, 50 datasets were obtained in each of the two situations, the length of each dataset being 5000, with  $r = 0.3SD$ ,  $r = 0.2SD$ ,  $r = 0.1SD$ , respectively. The results in Figure 7 show that when  $r = 0.2SD$ , seven leakage signals were discriminated as normal signals and 10 normal signals were discriminated as leakage signals, with an accuracy of 83%; while in the other two cases, the accuracies were 81% and 80%, respectively. Therefore, it was appropriate to set the threshold value as  $r = 0.2SD$ . The above analysis indicates that the complexity of the leakage signal should be higher than that of the non-leakage signal, and the complexity can therefore be used to identify leakage.

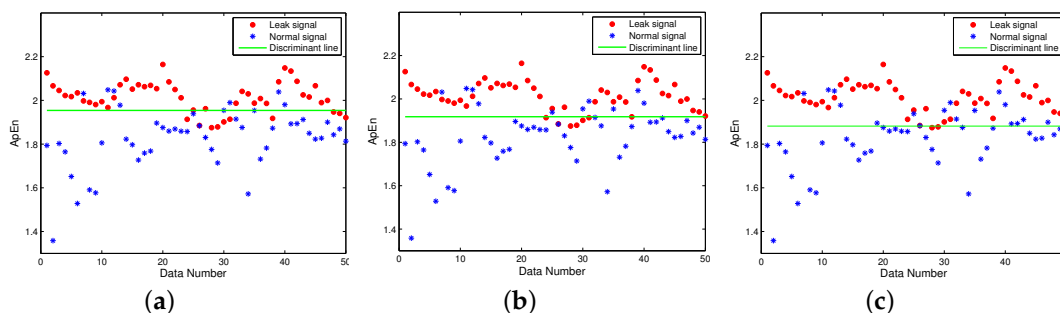


Figure 7. ApEn of signals for different threshold values. (a)  $r = 0.3SD$ . (b)  $r = 0.2SD$ . (c)  $r = 0.1SD$ .

### 4.1.3. Signal Principal Component Feature

PCA is a classical feature extraction method that involves the reduction of dimensionality and converts variables into a smaller number of aggregate variables (the principal components). Each principal component is a linear combination of the original variables, and the individual principal components are not mutually correlated. The principal components can convey a vast majority of the information contained in the original variables, and this information is not mutually

overlapping. In this paper, we used PCA to analyze the differences between pipeline leakage signals and non-leakage signals.

Assuming that  $n$  samples are obtained each time from pipeline signals, this can be expressed as  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T$ . If we have  $m$  sets of data  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ , we can construct an  $n \times m$  matrix  $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_m]$  as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}. \quad (18)$$

We used the signal principal components to construct a  $n \times l$  ( $0 < l \leq m$ ) component signal matrix  $\mathbf{Y} = [\mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_l]$ , and then constructed a matrix  $\mathbf{G}$  based on the internal product  $g_{ji} = [\mathbf{y}_j, \mathbf{x}_i]$  of the principal component signal matrix and original signal matrix,

$$\mathbf{G} = \mathbf{Y}^T (\mathbf{X} - E[\mathbf{X}]). \quad (19)$$

We further chose  $g_j = [g_{j1} g_{j2} \dots g_{jm}]$ ,  $0 < j \leq l$  as the feature for the identification of leakage.

#### 4.2. Machine Learning Inspired Water Pipeline Leakage Detection

Although the features in Section 4.1 have different characteristics in connection with the identification of pipeline leakage, the use of a single feature for identification is inefficient. For example, if the spectra of leakage and non-leakage signals are significantly different, the mean spectra of the IMFs will have excellent identification ability. However, when there is interference in the same band, this method tends to yield many false results. In addition, when the leakage from a pipeline is relatively small, the mean ApEn will have a poor ability to differentiate between leakage and non-leakage signals.

To increase the accuracy of leakage detection, this paper took advantage of the time-frequency features to construct identification feature sets and used SVM to classify the signal features and thereby determine pipeline leakage. The SVM is an advantageous means of solving small sample problems, nonlinear problems, and problems involving high-dimensional data, e.g., data forecasting, data fitting, and model identification. Assume that  $(\mathbf{x}_i, y_i)$  constitutes a training set data sample, where  $1 \leq i \leq N$ ,  $\mathbf{x}_i \in R^d$  for each sample,  $d$  is the dimensionality of the input space, and  $y_i \in \{-1, 1\}$  is the classification label. The training set can be linearly delimited by a hyper-plane that can be expressed as  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , where  $\mathbf{w}$  and  $b$  are locations that determine the hyper plane. A sample satisfying the following conditions is termed a support vector:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1. \quad (20)$$

In fact, the optimal classification of a sample is the solution for the optimal classification hyper plane,

$$\begin{cases} \min \phi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{i=1}^N \xi_i \\ \text{s.t. } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \end{cases}, \quad (21)$$

where  $\mathbf{w}$  is the coefficient vector of the classification hyper-plane in the feature space,  $b$  is the threshold value of the classification plane,  $\xi_i$  ( $\xi_i \geq 0$ ) is a relaxation factor included to account for the classification error, and  $C$  is a penalty factor for the misclassified sample. The optimal classification hyper-plane obtained can be expressed as:

$$\mathbf{w}_0 \cdot \mathbf{x} + b_0 = 0. \quad (22)$$

In nonlinear separable situations, a projection function (termed the kernel function) is used to project an input space  $R^d$  with low dimensionality into a feature space  $H$  with high dimensionality,

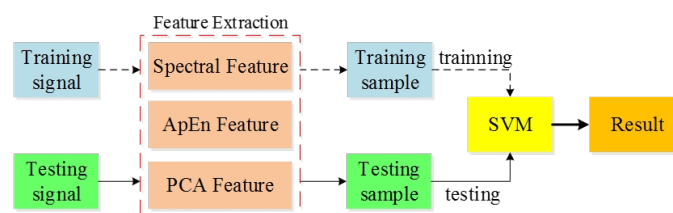
which converts the training sample from a linear inseparable problem with low dimensionality into a higher dimensional linear separable problem. At this time, the optimized dual problem is:

$$\begin{cases} \max_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad \sum_{i=1}^N y_i \alpha_i = 0, 0 \leq \alpha_i \leq C \end{cases}, \quad (23)$$

where  $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$  is the kernel function. Equation (23) shows that an SVM model with a suitable kernel function  $K(\cdot)$  must be chosen in the case of a nonlinear separable problem. The decision function corresponding to the use of Equation (23) is:

$$f(\mathbf{x}) = \text{sgn}[\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b]. \quad (24)$$

To improve the accuracy of leakage detection, we must use the training sample and testing sample to optimize the SVM, and the optimization processes are shown in Figure 8. Due to the effects of environmental factors on underground water pipelines, it is necessary to perform signal sampling during different times and at different places to compile sample sets including leakage signals and non-leakage signals. At first, the feature set of the training sample is used to perform SVM training, which creates a preliminary identification model. The feature set of the testing sample is then used to test the trained SVM model. The SVM model is optimized further based on testing results until the accuracy of the test output meets the requirements, which results in an SVM pipeline leakage identification model.



**Figure 8.** Schematic diagram of the SVM identification model training and optimization.

The theoretical analysis underlying Equations (23) and (24) indicates that the main factors that affect the SVM model's performance include the kernel function and the penalty factor  $C$ . According to the characteristics and ability of the SVM, we took the radial basis kernel as the kernel function in this paper [50–52]. The radial basis kernel function is expressed as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2). \quad (25)$$

In this case, the SVM model's performance is determined by the parameters  $C$  and  $\gamma$ . To achieve accurate identification results, the optimization process shown in Figure 8 must use the training sample and testing sample to optimize  $(C, \gamma)$ . Research has shown that the exponential sequences obtained using  $C$  and  $\gamma$  can achieve good results. In this paper, we obtained the following parameter values based on the parameter value range:  $C = 2^x (x \in [-5, 15])$  and  $\gamma = 2^y (y \in [-15, 5])$  [53,54]. Based on the cross-validation grid-search method, we then optimized the SVM parameters. We used the training sample and testing sample to perform testing of an SVM model with different  $2^x$  and  $2^y$  combinations and thus obtained the testing accuracy. The final step was to select the  $C$  and  $\gamma$  values that yielded the optimal cross-validation accuracy, which were selected for identification.

## 5. Simulation Results

### 5.1. Leakage Triggered Networking

In this section, OPNET Modeler14.5 was used to simulate the proposed leakage triggered networking method. The simulation process requires the design and configuration of three different layers. The node layer defines the node behavior and controls the data flow between the different modules in one node. The process layer uses the protocol to perform state conversion for the state machines. The network layer establishes the network topological structure and network layers.

Generally, a ZigBee node model includes an application layer, network layer, MAC layer, and a wireless transceiver. To compile the network power consumption, we added an energy calculation module to the node model. The energy calculation module kept track of the transceiver's standby, receiving, and transmitting energy consumption via monitoring of the transceiver status. Since the code for the application layer module and network layer module in OPNET was not available, we redesigned the application layer and network layer modules. During the simulation, the coordinator nodes, routing nodes, and terminal nodes had the same node model. The application layer included the source module and sink module. The source module employed the simple\_source model, which was a data packet generation module and was responsible for generating data packets with the specified packet size in accordance with the specified packet interval. The sink module was a data packet destruction module and was responsible for destroying data packets that had been transmitted to the destination node, which released internal storage dynamically assigned by the program. The network layer consisted of the network\_layer module and mainly served to drive the completion of networking procedures by the MAC module, complete initialization networking and leakage triggered networking, and perform data packet routing in accordance with the AODVjrouting protocol. The MAC layer employed an 802\_15\_4\_mac module and had a CSMA/CA competitive algorithm. The 802\_15\_4\_mac module performed some networking, multiple access, and sleep management functions via an added sleep state machine. The physical layer employed a wireless\_tx/wireless\_rx module as a wireless transceiver.

The water supply pipeline network shown in Figure 9 was designed for the simulation. The area was 1500 m × 1500 m and contained a total of 644 ZigBee nodes, which included six coordinator nodes and 638 routing and terminal nodes. The coordinator nodes were considered as sink nodes. The distance between adjacent nodes was 10 m; the distance between routing nodes was 50 m; and four terminal nodes were located between each pair of routing nodes. Each of the routing and terminal nodes was also a sensor. The terminal nodes completed the data collection and the detection of the water leakage signal and sent the data to the routing nodes. Then, the routing nodes transmitted the information collected by themselves and the terminal nodes to the sink nodes through multi-hop routing. Finally, the sink nodes sent the information to the background control center to realize the monitoring of the entire network. An RxGroup Configmo module was used to configure the nodes' single-hop link distance in the simulation scenario, and channel fading employed a free space loss model. The simulation parameters for each layer are shown in Table 4.

The simulation time was 1200 s. First, leakage signal information functions were established in the MAC layer, and a leakage triggered networking experiment was performed via the establishment of leakage point coordinates, signal attenuation coefficients, and the leakage signal detection threshold using the functions. The leakage coordinate was (319, 753), which indicated that the leakage point was located between Nodes 10 and 11, as shown in Figure 10. In the simulation, Node 3 was a coordinator node, Nodes 8, 13, and 18 were routing nodes, and the remaining nodes were terminal nodes. According to the simulation settings, leakage occurred at 700–720 s, and the signal attenuation coefficient and leakage signal detection threshold settings ensured that Routing Nodes 8 and 13 could receive the leakage signal. We also monitored the active and sleep status of the MAC layer to track the nodes' networking status.

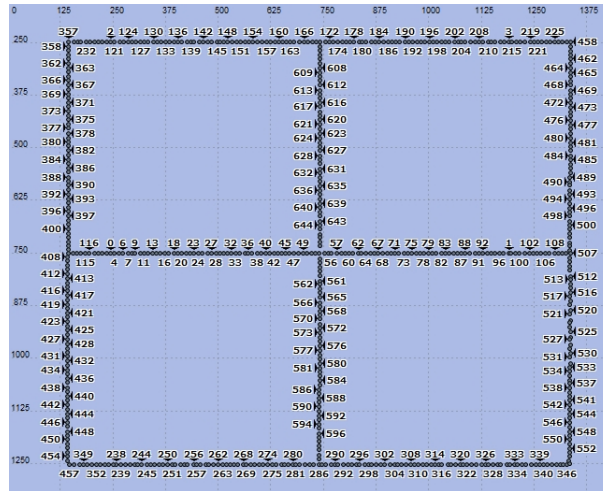


Figure 9. Arrangement of nodes.

Table 4. Simulation parameter settings.

Application Layer Parameters	
Packet Size	Constant (512)
Packet Interarrival Time	Exponential (5)
Start Time	120 s
Stop Time	Infinity
MAC Layer Parameters	
ACK Wait Duration	0.05 s
Minimum Backoff Exponent	3
Maximum Number of Backoffs	4
Channel Sensing Duration	0.5
Physical Layer Parameters	
Transmission Bands	2450 MHz Band
Data Rate	240 kbps
Packet Reception-Power Threshold	−85 dBm
Transmission Power	1 mW
Receive Power	0.4 mW
Idle Power	0.1 mW

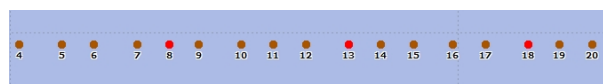
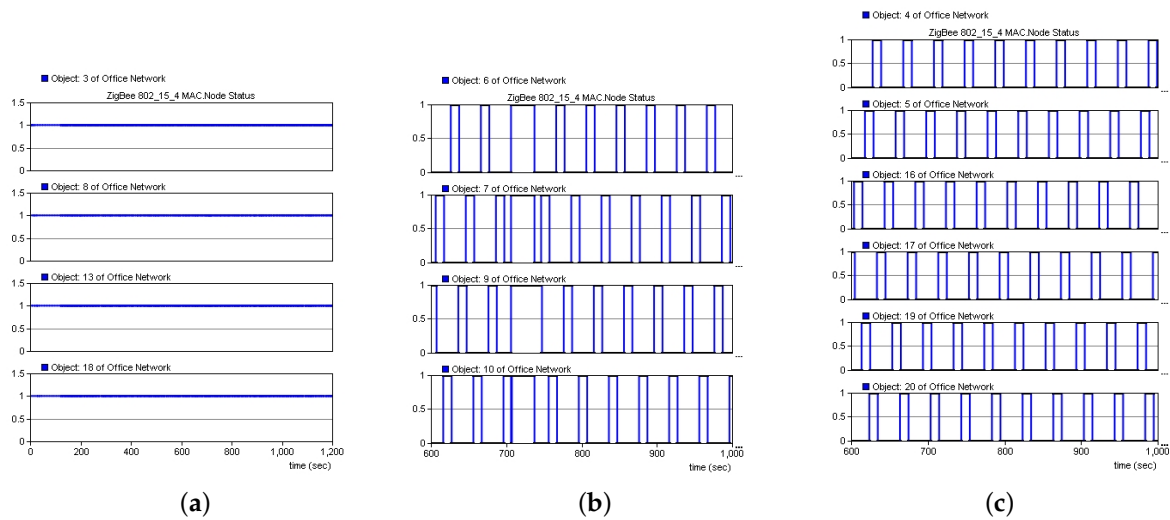


Figure 10. Arrangement of wireless sensor nodes in the vicinity of a leakage point.

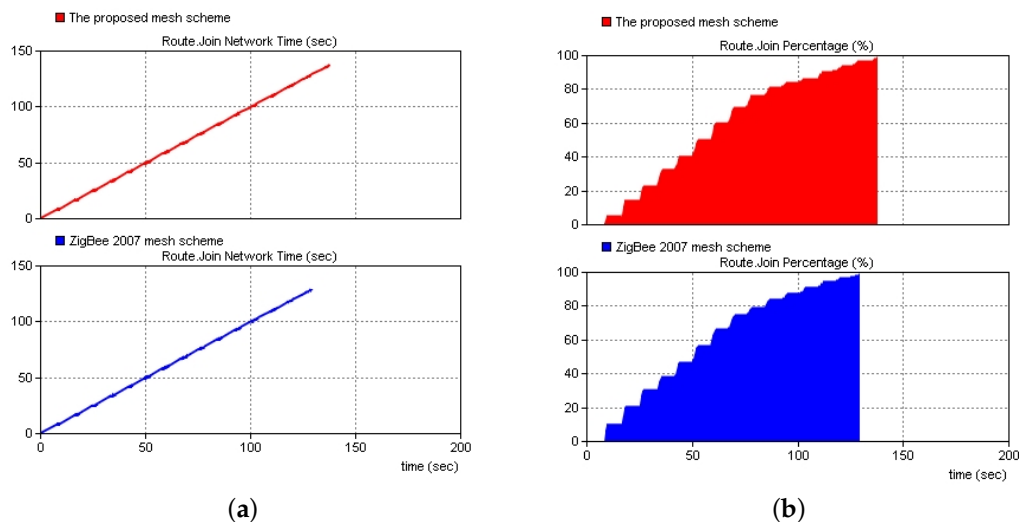
Figure 11 shows the working statuses during the 0–1200 s period. The coordinator and routing nodes were consistently in the working state, and the terminal nodes within the routing node network were sequentially working and sleeping. Leakage occurred when the simulation time reached 700 s, and all of the terminal nodes within the network formed by Routing Nodes 8 and 13 entered the working state at that time. Leakage ceased when the simulation time reached 720 s, and the terminal nodes within the network formed by Routing Nodes 8 and 13 resumed the normal working status. The routing nodes on both sides of the leakage point could detect the leakage signal and performed networking when the leakage occurred, whereas the other routing nodes remained in a normal working state. Simulation results indicated that the proposed solution achieved leakage triggered networking by the sensor nodes on both sides of a leakage point, which could further provide data to determine the location of the leakage point.





**Figure 11.** Node status. (a) Coordinator and routing nodes. (b) Networking nodes. (c) Nodes in the normal working state.

Figure 12 shows a comparison of the networking time between the proposed networking solution and the ZigBee 2007 networking solution. The simulation results indicated that the networking time of the proposed solution was slightly greater than that of the ZigBee 2007 solution because of the addition of the RSSI threshold value. However, the networking time of all the nodes increased only by approximately 1.39%. Therefore, the proposed solution could be used in a large scale networking environment.

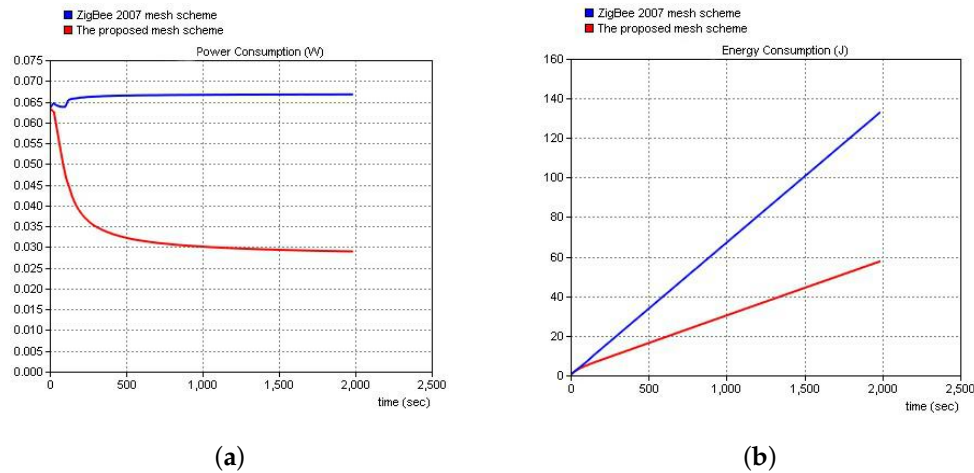


**Figure 12.** Comparisons of the networking time and proportions of networking nodes for the proposed solution and the ZigBee 2007 solution. (a) Networking time. (b) Proportion of networking nodes.

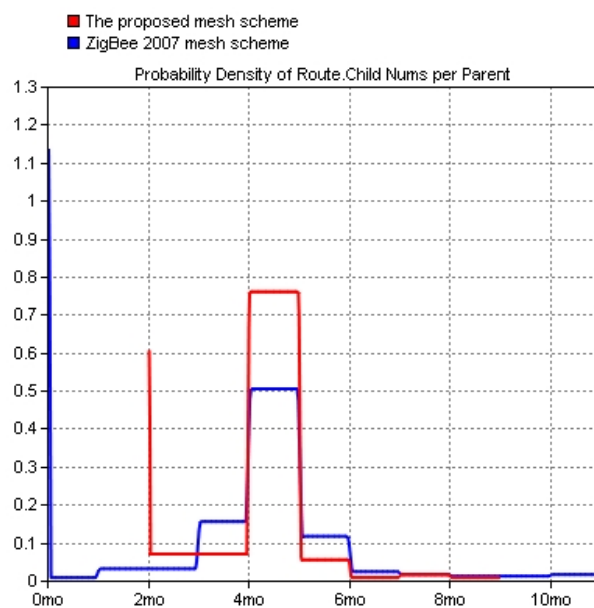
Figure 13 shows a comparison of the power and energy consumptions of the proposed solution and the ZigBee 2007 solution. Simulation results clearly demonstrated that the proposed solution could reduce the network power and energy consumption and increase the network lifetime through controlling of the polling work of the terminal nodes within their networks.

Figure 14 shows the percentage of terminal nodes carried by all the routing nodes when the transmission power of the sensor nodes was 1 mW and the RSSI threshold was  $-68$  dBm. In accordance with the channel loss model, the signal transmission distance controlled by the threshold value was approximately 25 m. Because the nodes were spaced at intervals of 10 m in the simulation,

four terminal nodes were carried by each routing node. The results shown in Figure 14 indicate that the proposed solution can ensure the number of terminal nodes carried by the routing nodes was more uniform than that in the original solution and could thereby ensure more stable network coverage.



**Figure 13.** Comparisons of the networking power and energy consumption for the proposed solution and the ZigBee 2007 solution. (a) Power. (b) Energy.

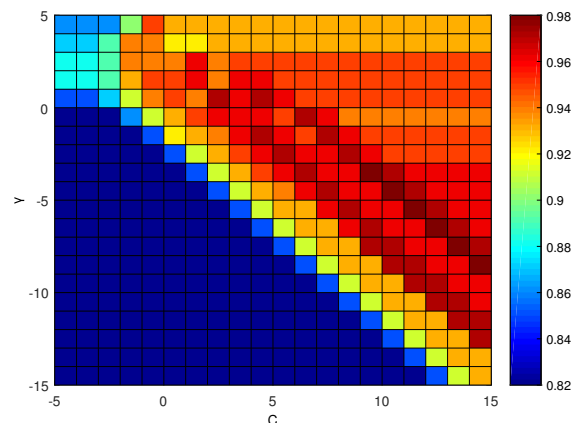


**Figure 14.** Ratio of terminal nodes carried by routing nodes for the proposed solution and the ZigBee 2007 solution.

## 5.2. Leakage Identification

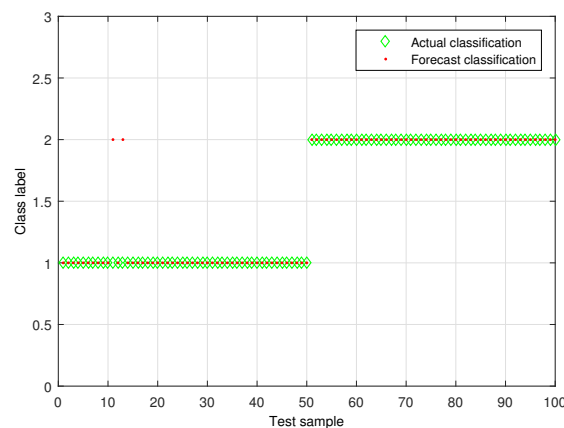
The experiment was performed along an exposed aluminum-plastic composite pipe, with a diameter of 27 mm. Moreover, the water-tap was used as the leakage sound source, then the flow rate was adjusted, and the sensor was placed at a distance of 20 cm away from the water-tap. One hundred datasets of leakage signals and non-leakage signals were sampled respectively. Each dataset had a length of 5000, and the data were used to train and optimize the SVM model. In addition, 100 datasets of leakage and non-leakage signals were sampled during the early quiet morning hours, respectively. Therefore, the simulated noise was employed to verify the effectiveness of the proposed leakage detection.

At first, 50 datasets were extracted from each of the leakage and non-leakage signals and used to create a training set. The remaining samples were then used to create a testing set. The SVM parameters  $(C, \gamma)$  were set to an integer power of two; the range of  $C$  was set as  $C \in [2^{-5}, 2^{15}]$ ; and the range of  $\gamma$  was set as  $\gamma \in [2^{-15}, 2^5]$ . The grid-search method was used, and  $21 \times 21 = 441$  for  $(C, \gamma)$  parameter combinations were used to perform the model training. The detection accuracy is shown in Figure 15. The results indicated that the highest identification accuracy achieved by the proposed algorithm was 98%. In addition, when  $C \geq 2^2$ ,  $\gamma \leq 2^0$ , and  $2^1 \leq C \times \gamma \leq 2^7$ , the SVM model based on the radial basis kernel provided good pipeline leakage signal identification performance.



**Figure 15.** Identification accuracy under different  $(C, \gamma)$  parameter combinations.

To verify the effectiveness of the proposed leakage detection, the experiment was performed in which 100 datasets of leakage signals and non-leakage signals were sampled, respectively. Figure 16 shows the detection results for  $(C, \gamma) = (2^9, 2^{-4})$  parameter combinations, where the leakage label was 2, and the non-leakage label was 1. The identification results showed that the proposed method only determined two leakage signals to be non-leakage signals and made accurate determinations in all other cases, indicating that the classification accuracy achieved by the proposed algorithm was 98%. Table 5 shows the results obtained by using the proposed algorithm to perform leakage identification after artificial Gaussian noise and impulsive noise were added to leakage signals obtained during a quiet period of time. The results indicated that the proposed water supply pipeline leakage detection method based on the time-frequency features of the signal and SVM could effectively detect pipeline leakage.



**Figure 16.** Pipeline leakage and non-leakage detection results.

**Table 5.** Leakage signal identification rate in an environment containing Gaussian noise and impulse noise.

		SNR(dB)									
		(c,g)	−12	−9	−6	−3	0	3	6	9	12
Gaussian Noise	$(2^9, 2^{-4})$	34	42	60	83	85	88	91	93	96	
	$(2^{10}, 2^{-5})$	34	42	60	82	85	87	91	92	96	
	$(2^{11}, 2^{-6})$	34	42	60	82	85	87	89	93	96	
	$(2^{12}, 2^{-7})$	37	42	60	82	84	87	88	93	95	
	$(2^{13}, 2^{-8})$	34	42	60	82	85	87	91	93	96	
	$(2^{14}, 2^{-9})$	34	42	59	82	85	87	91	93	96	
	$(2^{15}, 2^{-10})$	34	42	59	82	85	87	91	83	96	
Impulsive Noise	$(2^9, 2^{-4})$	31	32	56	84	85	88	90	94	96	
	$(2^{10}, 2^{-5})$	31	32	54	82	84	87	89	93	95	
	$(2^{11}, 2^{-6})$	31	32	59	81	84	87	90	93	95	
	$(2^{12}, 2^{-7})$	31	32	56	81	85	87	90	93	95	
	$(2^{13}, 2^{-8})$	31	32	55	81	84	87	90	93	95	
	$(2^{14}, 2^{-9})$	31	33	55	81	84	87	91	93	96	
	$(2^{15}, 2^{-10})$	31	32	55	80	84	87	91	93	95	

## 6. Conclusions

In this paper, an experimental water pipeline leakage detection system based on machine learning and wireless sensors networks was presented. The system employed ZigBee and 4G to acquire and transmitted signals. In addition, a leakage triggered networking method was further proposed to reduce the WSN energy consumption effectively and prolong the system life. To improve the accuracy of water pipeline leakage detection, the proposed system made better use of EMD, ApEn, and PCA of the leak signal and SVM to identify the leakage signal intelligently. Simulation analysis and experimental results indicated that the proposed leakage identification method could effectively identify the water pipeline leakage.

**Author Contributions:** The algorithms proposed in this paper were conceived by Y.L. (Yang Liu), X.M., Y.L. (Yuting Li), and Y.T. Y.L. (Yang Liu), Y.T., J.G., and Y.Z. designed the experiments. X.M., Y.L. (Yuting Li), Y.Z., and J.G. performed the experiments and analyzed the results. Y.L. (Yang Liu), Y.T., and X.M. wrote the paper.

**Funding:** The authors would like to acknowledge the National Natural Science Foundation of China for its support of this research. This work was supported by the National Science Foundation of China under Grant 61761033 and 61461036, in part by the Natural Science Foundation of Neimongol Autonomous Region of China under Grant 2019MS06033, and in part by the High College Research Project of Neimongol Autonomous Region of China under Grant NJYT-20-A11.

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

## Abbreviations

The following abbreviations are used in this manuscript:

WSNs	Wireless sensor networks
SVM	Support vector machine
LPCC	Linear predictive coding coefficient
HMM	Hidden Markov model
PCA	Principal component analysis
LOS	Line-of-sight
NLOS	Non-line-of-sight
RSSI	Received signal strength indicator
EMD	Empirical mode decomposition
IMFs	Intrinsic mode functions
ApEn	Approximate entropy
SD	Standard deviation

## References

1. Tindall, J.A.; Campbell, A.A. Water security-national and global issues. *Sex. Relatsh. Ther.* **2010**, *30*, 314–324.
2. 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](#)]
3. Cataldo, A.; Cannazza, G.; Benedetto, E.D.; Giaquinto, N. A new method for detecting leaks in underground water pipelines. *IEEE Sens. J.* **2012**, *12*, 1660–1667. [[CrossRef](#)]
4. Fang, D.; Chen, B. Ecological network analysis for a virtual water network. *Environ. Sci. Technol.* **2015**, *49*, 6722–6730. [[CrossRef](#)] [[PubMed](#)]
5. Moore, N.M. The Green City: Sustainable homes, sustainable suburbs. Nicholas Low, Brendan Gleeson, Ray Green, and Darko Radović. *Urban Geogr.* **2009**, *30*, 927–928. [[CrossRef](#)]
6. Martini, A.; Troncossi, M.; Rivola, A. Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements. *Shock Vibrat* **2015**, *2015*, 1–13. [[CrossRef](#)]
7. Kim, H.; Shin, E.S.; Chung, W.J. Energy demand and supply, energy policies, and energy security in the Republic of Korea. *Energy Policy* **2011**, *39*, 6882–6897. [[CrossRef](#)]
8. 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](#)]
9. Wang, F.; Lin, W.; Liu, Z.; Wu, S.; Qiu, X. Pipeline leak detection by using time-domain statistical features. *IEEE Sens. J.* **2017**, *17*, 6431–6442. [[CrossRef](#)]
10. Lay-Ekuakille, A.; Vergallo, P. Decimated signal diagonalization method for improved spectral leak detection in pipelines. *IEEE Sens. J.* **2014**, *14*, 1741–1748. [[CrossRef](#)]
11. Thompson, M.; Chapman, C.J.; Howison, S.D.; Ockendon, J.R. Noise generation by water pipe leaks. In Proceedings of the 40th European Study Group with Industry, Keele, UK, 9–12 April 2001; pp. 1–4.
12. Yang, J.; Wen, Y.; Li, P. Leak acoustic detection in water distribution pipelines. In Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, 25–27 June 2008; pp. 3057–3061.
13. Goulet, J.A.; Coutu, S.; Smith, I.F.C. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Adv. Eng. Inform.* **2013**, *27*, 261–269. [[CrossRef](#)]
14. Cabrera, D.A.; Herrera, M.; Izquierdo, J.; Levario, S.J.O.; Garcia, R.P. GPR-based water leak models in water distribution systems. *Sensors* **2013**, *13*, 15912–15936. [[CrossRef](#)]
15. Benkherouf, A.; Allidina, A.Y. Leak detection and location in gas pipelines. *IEE Proc. D Control. Theory Appl.* **1988**, *135*, 142–148. [[CrossRef](#)]
16. Chen, H.; Ye, H.; Chen, L.V.; Su, H. Application of support vector machine learning to leak detection and location in pipelines. In Proceedings of the 21st IEEE Instrumentation and Measurement Technology, Como, Italy, 18–20 May 2004; pp. 2273–2277.
17. Ellul, I.R. Advances in pipeline leak detection techniques. *Pipes Pipelines Int.* **1989**, *34*, 7–12.
18. Wan, Q.; Koch, D.B.; Morris, K. Multichannel spectral analysis for tube leak detection. In Proceedings of the Southeastcon'93, Charlotte, NC, USA, 4–7 April 1993; 4p.
19. Ai, C.S.; Zhao, H.; Ma, R.J.; Dong, X. Pipeline damage and leak detection based on sound spectrum LPCC and HMM. In Proceedings of the Sixth International Conference on Intelligent System Design and Applications, Jinan, China, 16–18 October 2006; pp. 829–833.
20. 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.
21. Hunaidi, O.; Chu, W.T. Acoustical characteristics of leak signals in plastic water distribution pipes. *Appl. Acoust.* **1999**, *58*, 235–254. [[CrossRef](#)]
22. 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*, 1–10. [[CrossRef](#)]
23. Kang, J.; Park, Y.; Lee, J.; Wang, S.; Eom, D. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4279–4289. [[CrossRef](#)]
24. Chraim, F.; Erol, Y.B.; Pister, K. Wireless gas leak detection and localization. *IEEE Trans. Ind. Inform.* **2016**, *12*, 768–779. [[CrossRef](#)]
25. Yoon, S.; Ye, W.; Heidemann, J.; Littlefield, B.; Shahabi, C. SWATS: Wireless sensor networks for steamflood and waterflood pipeline monitoring. *IEEE Netw.* **2011**, *25*, 50–56. [[CrossRef](#)]

26. Jing, L.; Li, Z.; Li, Y.; Murch, R. Channel characterization of acoustic waveguides consisting of straight gas and water pipelines. *IEEE Access* **2018**, *6*, 6807–6819. [[CrossRef](#)]
27. Poulakis, Z.; Valougeorgis, D.; Papadimitriou, C. Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probabilistic Eng. Mech.* **2003**, *18*, 315–327. [[CrossRef](#)]
28. Ferrante, M.; Brunone, B.; Rossetti, A.G. Harmonic analysis of pressure signal during transients for leak detection in pressurized pipes. In Proceedings of the 4th International Conference on Water Pipeline Systems, York, UK, 28–30 March 2001; pp. 28–30.
29. Wang, X.J.; Lambert, M.F.; Simpson, A.R.; Liggett, J.A.; Vítkovský, J.P. Leak detection in pipelines using the damping of fluid transients. *J. Hydraulic Eng.* **2002**, *128*, 697–711. [[CrossRef](#)]
30. Almazyad, A.S. A proposed scalable design and simulation of wireless sensor network-based long-distance water pipeline leakage monitoring system. *Sensors* **2014**, *14*, 3557–3577. [[CrossRef](#)] [[PubMed](#)]
31. Ali, S. SimpliMote: A wireless sensor network monitoring platform for oil and gas pipelines. *IEEE Syst. J.* **2018**, *12*, 778–789. [[CrossRef](#)]
32. Hodge, V.J.; O’Keefe, S.; Weeks, M.; Moulds, A. Wireless sensor networks for condition monitoring in the railway industry: A survey. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1088–1106. [[CrossRef](#)]
33. Miao, Y.; Li, W.; Tian, D.; Hossain, M.S.; Alhamid, M.F. Narrowband internet of things: Simulation and modelling. *IEEE Internet Things J.* **2018**, *5*, 2304–2314. [[CrossRef](#)]
34. Zhang, T.; Fan, H.; Loo, J.; Liu, D. User preference aware caching deployment for device-to-device caching networks. *IEEE Syst. J.* **2019**, *13*, 226–237. [[CrossRef](#)]
35. Stoianov, I.; Nachman, L.; Madden, S.; Tokmouline, T. PIPENET: A wireless sensor network for pipeline monitoring. In Proceedings of the 6th international conference on Information processing in sensor networks, Cambridge, MA, USA, 25–27 April 2007; pp. 264–273.
36. Chang, Y.C.; Lai, T.T.; Chu, H.H.; Huang, P. Pipeprobe: Mapping spatial layout of indoor water pipelines. In Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 18–20 May 2009; pp. 391–392.
37. Whittle, A.J.; Girod, L.; Preis, A.; Allen, M.; Lim, H.B.; Iqbal, M.; Goldsmith, D. WATERWISE@SG: A testbed for continuous monitoring of the water distribution system in Singapore. In Proceedings of the 12th Annual Conference on Water Distribution Systems Analysis (WDSA), Tucson, AZ, USA, 12–15 September 2010; pp. 1362–1378.
38. Jang, E.H.; Park, B.J.; Kim, S.H.; Chung, M.A.; Park, M.S.; Sohn, J.H. Emotion classification based on bio-signals emotion recognition using machine learning algorithms. In Proceedings of the 2014 International Conference on Information Science, Electronics and Electrical Engineering, Sapporo, Japan, 26–28 April 2014; pp. 1373–1376.
39. Siryani, J.; Tanju, B.; Eveleigh, T.J. A machine learning decision support system improves the internet of things’ smart meter operations. *IEEE Internet Things J.* **2017**, *4*, 1056–1066. [[CrossRef](#)]
40. Pławiak, P.; Sośnicki, T.; Niedźwiecki, M.; Tabor, Z.; Rzecki, K. Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms. *IEEE Trans. Ind. Informat.* **2016**, *12*, 1104–1113. [[CrossRef](#)]
41. Ye, H.; Liang, L.; Li, G.Y.; Kim, J.; Lu, L.; Wu, M. Machine learning for vehicular networks: Recent advances and application examples. *IEEE Veh. Technol. Mag.* **2018**, *13*, 94–101. [[CrossRef](#)]
42. Malfante, M. Machine learning for volcano seismic signals: Challenges and perspectives. *IEEE Singal Process. Mag.* **2018**, *35*, 20–30. [[CrossRef](#)]
43. Prieto, M.D. Bearing fault detection by a novel condition monitoring scheme based on statistical time features and neural networks. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3398–3407. [[CrossRef](#)]
44. Becari, W.; de Oliveira, A.M.; Peres, H.E.M.; Correra, F.S. Microwave-based system for non-destructive monitoring water pipe networks using support vector machine. *IET Sci. Meas. Technol.* **2016**, *10*, 910–915. [[CrossRef](#)]
45. Gomaa, R.I. Real-time radiological monitoring of nuclear facilities using Zigbee technology. *IEEE Sens. J.* **2014**, *14*, 4007–4013. [[CrossRef](#)]
46. Zheng, K. Energy efficient localization and tracking of mobile devices in wireless sensor networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 2714–2726. [[CrossRef](#)]
47. Huang, N.E.; Shen, Z.; Long, S.R. The empirical mode decomposition and the Hilbert spectrum for nonlinear non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]



48. Yang, J.; Wen, Y.; Li, P. Leak location using blind system identification in water distribution pipelines. *J. Sound Vibrat.* **2008**, *310*, 134–148. [[CrossRef](#)]
49. Pincus, S.M. Approximate entropy: A complexity measure for biological time series data. In Proceedings of the 1991 IEEE Seventeenth Annual Northeast Bioengineering Conference, Hartford, CT, USA, 4–5 April 1991; pp. 35–36.
50. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
51. Müller, K.R.; Mika, S.; Rätsch, G.; Tsuda, K.; Schölkopf, B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.* **2001**, *12*, 181–201. [[CrossRef](#)]
52. Ayat, N.E.; Cheriet, M.; Suen, C.Y. Automatic model selection for the optimization of SVM kernels. *Pattern Recognit.* **2005**, *38*, 1733–1745. [[CrossRef](#)]
53. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [[PubMed](#)]
54. Huang, C.L.; Wang, C.J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **2006**, *31*, 231–240. [[CrossRef](#)]



© 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/>).