Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2020, 7(9):51-68

Research Article ISSN: 2394 - 658X

Water Leakage Detection Using Machine Learning & Wireless Sensors Network (IoT)

Siva Sathyanarayana Movva

*sivasathya@gmail.com

ABSTRACT

Detecting water pipeline leaks is crucial for ensuring the safe operation of water supply networks and conserving water resources. To overcome the limitations of traditional leak detection methods, this study presents a novel approach utilizing machine learning and wireless sensor networks (WSNs). The system deploys wireless sensors along pipelines to gather data and leverages 4G networks for remote data transmission. An energy-efficient networking method triggered by leaks is introduced to optimize WSN energy consumption and prolong system longevity. To enhance detection accuracy and intelligence, a leakage identification methodology is proposed, incorporating intrinsic mode function, approximate entropy, and principal component analysis to construct a comprehensive signal feature set. The detection process uses support vector machine (SVM) as a classifier. Simulation and experimental results validate the efficacy of the proposed method in accurately identifying water pipeline leaks while demonstrating lower energy consumption compared to conventional WSN networking methods.

Key words: leak detection, wireless sensor networks, machine learning, SVM

INTRODUCTION

Water serves as the foundational substance supporting human life and the existence of all living organisms. It is an essential natural resource crucial for the advancement of human society. The demand for water resources is surging due to population growth, economic development, and shifting consumption patterns, with expectations of a significant acceleration in the next two decades. However, a critical issue lies in the wastage of water resources caused by leakage from water pipelines. Research by the World Bank revealed an annual leakage from water pipelines exceeding 48.6 billion m3, leading to economic losses of approximately US$14.6 billion. Globally, water pipeline leakage rates surpass 10% in one-third of countries, reaching 23% in the EU, 13% in the U.S. and Canada, 22% in Asia, 35% in Latin America, and 30% in Africa. Consequently, the exploration of high-performance water pipeline leakage detection technologies holds immense significance for safeguarding water resources and fostering economic development. The primary causes of water pipeline leakage encompass the corrosive nature of soil, substandard pipe material quality, temperature and pressure variations, non-adherence to standard pipe laying methods, geological changes, and human-induced damage. Detecting these leaks is challenging as most water pipelines are situated deep underground, making prompt discovery difficult. Leakage, often subtle or undetectable, results in significant water resource wastage. Thus, effective and accurate detection of underground water pipeline leakage is of utmost importance. Both academic researchers and industry experts have conducted extensive research and developed various detection methods. Among the earliest methods is the listening system, relying on personnel to detect changes in volume and sound quality of leakage noise for pinpointing leakage areas. Ground-penetrating radar and methods based on internal pressure changes have also been explored, each presenting its own set of challenges. Additionally, studies incorporating

spectral analysis, piezoelectric accelerometers, linear predictive coding coefficients, and hidden Markov models have been undertaken, although facing limitations. Notably, wireless sensor networks (WSNs) offer a promising solution to address these challenges. Leveraging superior sensing capabilities, communication protocols, processor speed, and data collection advantages, WSN technologies have found widespread applications in monitoring. For instance, PipeNet, a wireless sensor network-based system, has been proposed to monitor water flow and detect leaks by employing acoustic and vibration sensors on large bulk-water pipelines and pressure sensors on normal pipelines. In contrast to the PipeNet initiative, the PipeProbe system operates without the assumption that water pipe surfaces are exposed and accessible for sensor module attachment. PipeProbe can be deployed into the source of a water pipeline, traversing it while collecting the necessary sensor readings for reconstructing the 3D spatial layout of the traversed water pipelines. A wireless sensor network named WaterWiSe@SG was proposed to showcase the application and control of a low-cost wireless sensor network with high data rates, enabling real-time monitoring of a water distribution network in Singapore. The goal of WaterWiSe@SG was to develop generic wireless sensor network capabilities for real-time monitoring of a water distribution network.This paper introduces a water pipeline monitoring system based on wireless sensor networks and a leakage identification method using support vector machines (SVM). Machine learning, which simulates knowledge acquisition through human learning activities, facilitates automatic system performance improvement. Its wide application includes speech and biological affect identification, physiological signal detection, body movement identification, and signal feature detection and identification. The proposed system utilizes ZigBee nodes as signal collection nodes and employs the 4G network to transmit signals from sensors to the data processing center. To address high networking power consumption in conventional wireless sensor networks, we propose a leakage-triggered networking method capable of networking and performing data transmission from wireless sensor nodes near leakage points, effectively reducing network energy consumption and extending its life cycle.Distinguishing between leakage and non-leakage signals based on time-frequency features, we propose a leakage detection method that constructs feature matrices using intrinsic mode function, approximate entropy, and principal component analysis (PCA). SVM serves as a classifier to identify leaks. Experiments are conducted along an exposed aluminum-plastic composite pipe with a diameter of 27 mm, using the CT1010 acceleration sensor due to its sensitivity. The water pressure during experiments is maintained at no less than 0.3 MPa, and the detectable leaking flow rate is calculated to be approximately 2.5 cm3/s based on pipe pressure. Experimental and simulation results demonstrate the effectiveness of the proposed methods in detecting leaks and prolonging the lifetime of the wireless sensor network.

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 host's software was crafted using C# and SQL, incorporating the TCP protocol and TCP server on the gateway to establish network connections. It received control commands and data transmissions from the ZigBee wireless sensor network. Operating within the specified parameters, the CC2530 boasted a nominal maximum operating distance of 75m, while research confirmed the reliable range of ZigBee nodes to be 60m in a line-of-sight (LOS) environment [45,46]. Employing NI's MCC BTH-1208LS data acquisition card, data was stored on the host computer for signal analysis. The experiment utilized the CT1010 acceleration sensor for its sensitivity, aligning seamlessly with the data acquisition card. In Figure 2, the results display the packet loss rate

of the system's terminal nodes in a non-line-of-sight (NLOS) environment, varying with the intervals between nodes. To uphold the reliability of the system's data transmissions, it is imperative to maintain intervals between CC2530 nodes at approximately 30m.

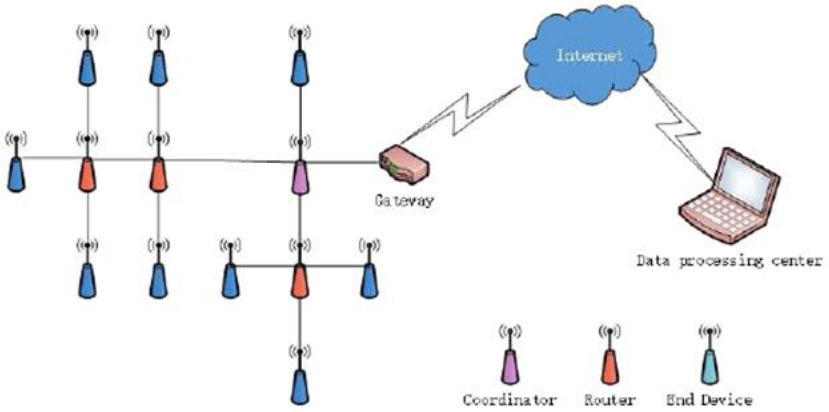

Figure 1: System architecture for pipeline monitoring

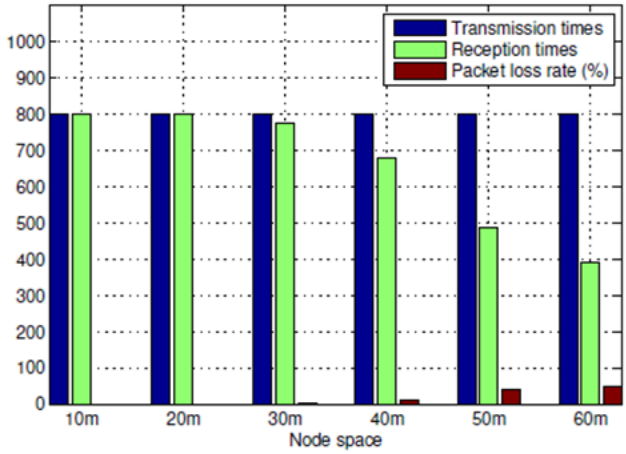

Figure 2: CC2530 packet loss rate under NLOS conditions

A. Leakage Triggered ZigBee Networking

First, in practical water pipeline monitoring settings, it is essential to deploy a multitude of terminal sensors along pipelines. Given the low probability of pipeline leakage occurrences, simultaneous operation of all sensors leads to a considerable waste of energy. Additionally, water supply pipeline leaks happen sporadically, necessitating real-time monitoring of pipeline systems. To mitigate energy consumption and enhance network longevity, we propose a leakage-triggered networking approach. The ZigBee networking method comprises initialization networking and triggered networking. The network's topological structure is a crucial foundation for ZigBee networking. Considering the structural characteristics and distribution of water supply pipelines, this study adopts a network topology. For initialization networking, the first step involves determining coordinator nodes, setting their signal channels and network ID numbers, thereby initializing the network. Subsequently, non-coordinator nodes are integrated into the network. Figure 3 illustrates how nodes join the network, emphasizing the need for balanced terminal node distribution among relay nodes (i.e., router nodes). To achieve this balance, the solution incorporates Received Signal Strength Indicator (RSSI) information into each Beacon_request frame when a terminal node sends a network join request. Routing nodes then offer joining services to terminal nodes based on RSSI values. For effective and reliable collection of leakage signals, reasonable threshold values for RSSI are crucial. If terminal node RSSI values fall below the threshold, routing nodes disregard their requests. Conversely, if RSSI values surpass the threshold, routing nodes record terminal node information, ensuring successful network integration. Once the network is initialized, this study introduces three types of control frames (join frame, active frame, and wave frame) to reduce network power consumption

and prolong its lifespan. These frames trigger network actions based on leakage detection results. Tables 1–3 depict the structures of the join frame, active frame, and wave frame, highlighting key fields such as Sou_address, Des_address, PANID, and specific result indicators.

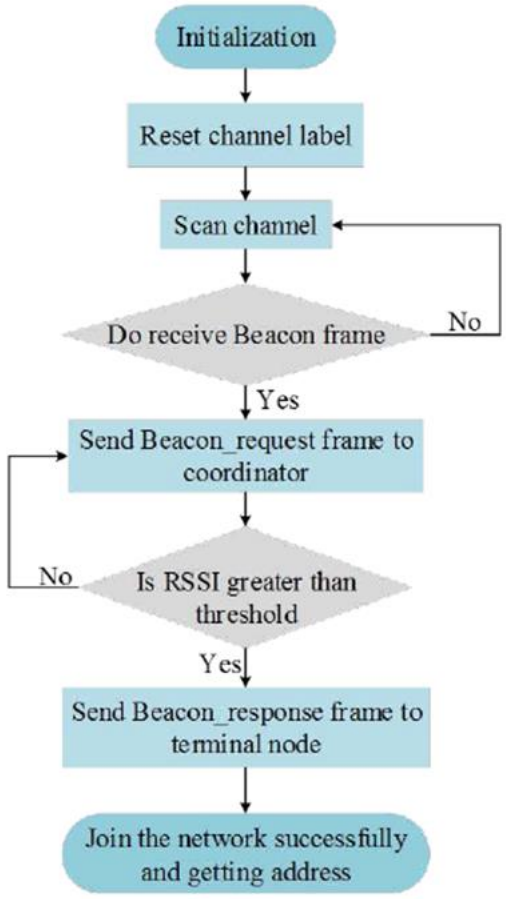

Figure 3: Flowchart of the nodes joining the network

Table 1: Join Frame

Field Name	PANID	Des_address	Sou_address	Join_result	Channel
Length	16 bits	16 bites	16 bits	1 bit	32 bits
Instruction	Network	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

Figure 4 illustrates the operational sequence of the networking flow in response to leakage triggers. The terminal nodes initiate the process by dispatching Join frames to the routing nodes, leading to the compilation of a comprehensive list of terminal nodes. Once the routing-terminal relationships are established, the routing nodes subsequently transmit Active frames to the terminal nodes. Upon reception of Active frames, the terminal nodes assess whether they serve as destination nodes. If a node identifies itself as a destination, it activates and proceeds with data sampling. Conversely, if a node is not a designated destination, it enters a sleep monitoring state, awaiting the next Active frame. When a node in the active state detects a received signal as a leakage signal, the routing node transmits the predetermined leakage-triggered address to the nodes listed in the terminal nodes compilation. All nodes receiving this address then activate themselves, initiating signal sampling and

transmission. This methodology empowers routing nodes to govern the operational status of terminal nodes. Upon completion of the networking process, data transmission is executed using the ZigBee routing protocol.

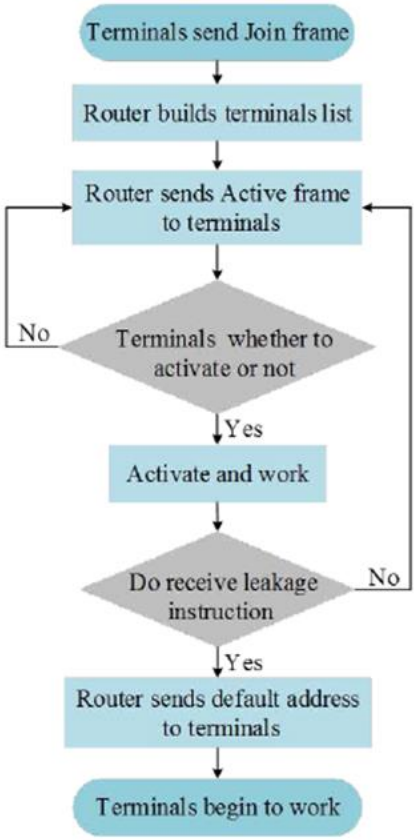

Figure 4: Flowchart of the ZigBee network leakage trigged networking

Leakage Detection by Using Machine Learning and Time-Frequency Features

Time-Frequency Analysis of the Acoustic Leakage Signal

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) 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:

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

where x ˆ (t) is the Hilbert transform of a signal x(t) and F(t) is given by:

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

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

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

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:
The number of extreme points is Ne (including the minimum and maximum values), which is the same as or no more than one from the number of zero crossing points Ns,

$$(N_s - 1) \leq N_e \leq (N_s + 1). \tag{4}$$

At an arbitrary time ti 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,

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

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),

$$h_1(t) = x(t) - m(t). \tag{6}$$

We now check whether h1(t) satisfies the two conditions of the IMF. If it does not satisfy them, repeat (6) until the IMF conditions are satisfied. h1(t) at this time is expressed as c1(t), where c1(t) is the first IMF of the signal x(t),

$$c_1(t) = h_1(t). \tag{7}$$

Next, c1(t) is subtracted from the original signal x(t) to obtain the new signal r1(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),
Next, c1(t) is subtracted from the original signal x(t) to obtain the new signal r1(t),

$$r_1(t) = x(t) - c_1(t). \tag{8}$$

Next, c1(t) is subtracted from the original signal x(t) to obtain the new signal r1(t), Repeating (6) until the IMF conditions are satisfied, the first IMF of r1(t) is obtained, which is the second IMF of x(t) and is denoted as c2(t). Continuing in the same manner, we can progressively derive the mth IMF cm(t) of the signal x(t) and the remainder rm(t).
Using the foregoing steps, the original signal x(t) can be decomposed into the sum of IMFs and a remainder,

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

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 Sd,

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

where N is the observed signal length and $h_{k-1}(n)$ and hk(n) are two consecutive processing results in the process of the IMF derivative. When the standard deviation Sd reaches the preset threshold value, we can assume that Condition (2) has been satisfied. It has been indicated in that the threshold value of the standard deviation Sd 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 Ci(k) of the IMF components ci(n) resulting from EMD decomposition,

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

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

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

We then obtain the mean of Equation (12),

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

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. 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 Sd 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.

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. 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.

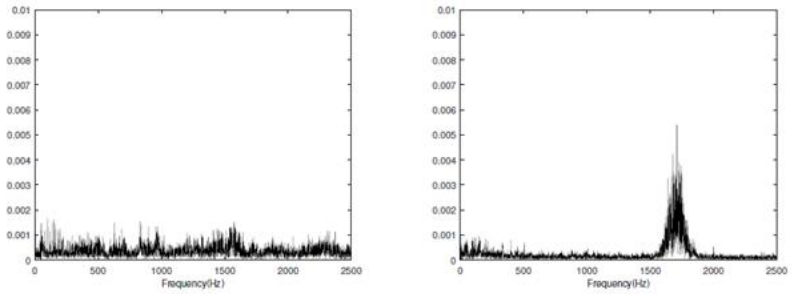

Figure 5: Power spectra of signals from a pipeline (a) Normal signal power spectrum (b) Leakage signal power spectrum

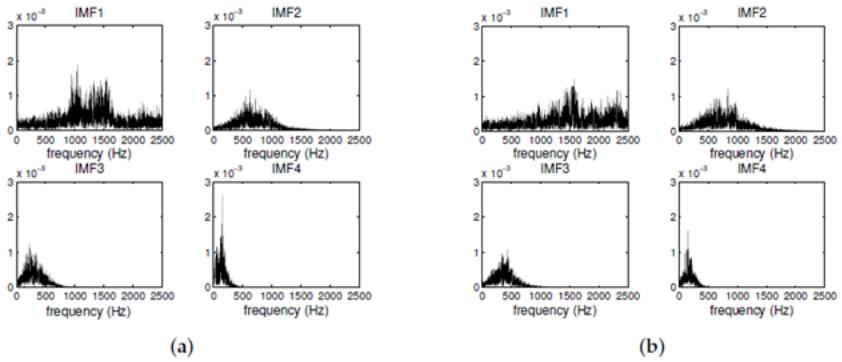

Figure 6. The first four intrinsic mode function (IMF) spectra of the pipeline leakage and non-leakage signal. (a) Leakage signal. (b) Non-leakage signal.

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

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

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

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

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

For all i values, we derive the mean fm(r) of ln C^m_i (r),

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

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

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

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.

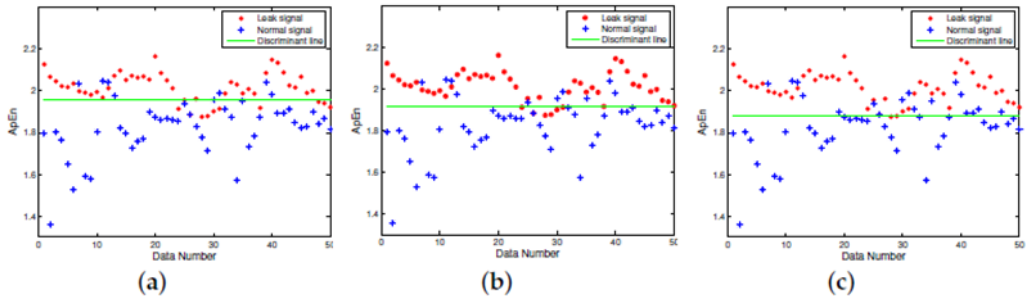

Figure 7: ApEn of signals for different threshold values (a) r=0,3SD, (b) r=0.2SD, (c) r=0.1SD

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 xi = (x1i , x2i , · · · , xni)T. If we have m sets of data x1, x2, · · · , xm, we can construct an n × m matrix X = [x1x2 · · · xm] as:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \tag{18}$$

We used the signal principal components to construct a n × l (0 < l ≤ m) component signal matrix Y = [y1y2 · · · yl], and then constructed a matrix G based on the internal product gji = [yj, xi] of the principal component signal matrix and original signal matrix,

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

We further chose gj = [gj1gj2 · · · gjm], 0 < j ≤ l as the feature for the identification of leakage.

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 (xi , yi) constitutes a training set data sample, where 1 ≤ i ≤ N, xi ∈ Rd for each sample, d is the dimensionality of the input space, and yi ∈ {−1, 1} is the classification label. The training set can be linearly delimited by a hyper-plane that can be expressed as w· x+ b = 0, where w and b are locations that determine the hyper plane. A sample satisfying the following conditions is termed a support vector:

$$y_i(\mathbf{w} \cdot x_i + b) = 1. \tag{20}$$

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

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

where w is the coefficient vector of the classification hyper-plane in the feature space, b is the threshold value of the classification plane, xi(xi ≥ 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:

$$\mathbf{w}_o \cdot \mathbf{x} + b_o = 0. \tag{22}$$

In nonlinear separable situations, a projection function (termed the kernel function) is used to project an input space Rd 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:

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

where K(xi , xj) = F(xi) · F(xj) is the kernel function. Equation (23) shows that an SVM model with a suitable kernel function K(·) must be chosen in the case of a nonlinear separable problem. The decision function corresponding to the use of Equation (23) is

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

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.

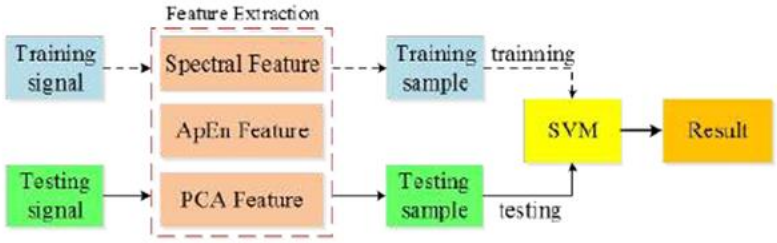

Figure 8: Schematic diagram of the SVM identification model training and optiimization

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. The radial basis kernel function is expressed as:

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

In this case, the SVM model's performance is determined by the parameters C and g. To achieve accurate identification results, the optimization process shown in Figure 8 must use the training sample and testing

sample to optimize (C, g). Research has shown that the exponential sequences obtained using C and g can achieve good results. In this paper, we obtained the following parameter values based on the parameter value range: $C = 2x(x \in [-5, 15])$ and $g = 2y(y \in [-15, 5])$. 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 2x and 2y combinations and thus obtained the testing accuracy. The final step was to select the C and g values that yielded the optimal cross-validation accuracy, which were selected for identification.

SIMULATION RESULTS

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 opological 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 AODVjrrouting 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.

Table 4: Simulation Parameter settings

Application Layer Parameters	
Packet Size	Constant(512)
Packet Interarrival Time	Exponential (5)
Start Time	120s
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 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.

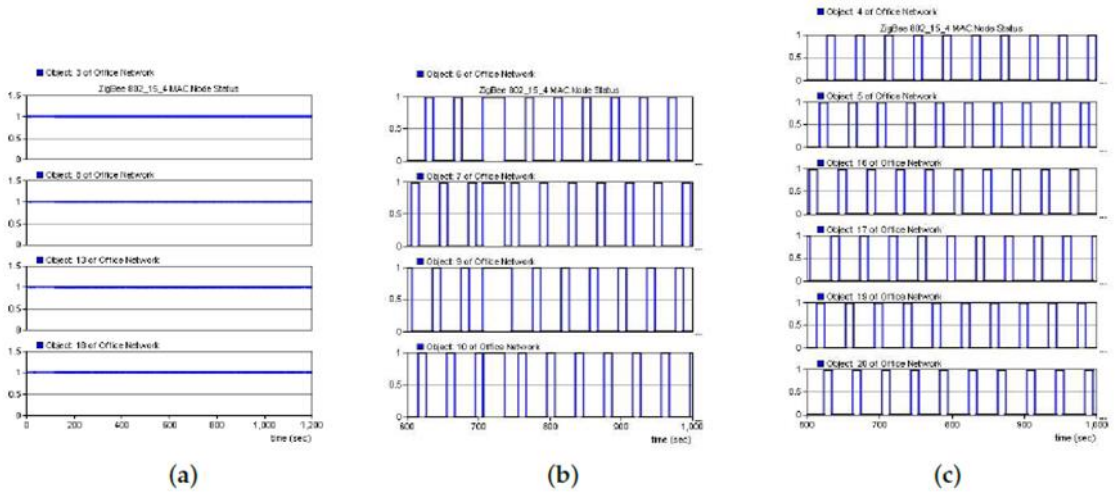

Figure 11: Node stauts (a) Coordinator and routing nodes. (b) Networking nodes. (c) Nodes in the normal working state

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.

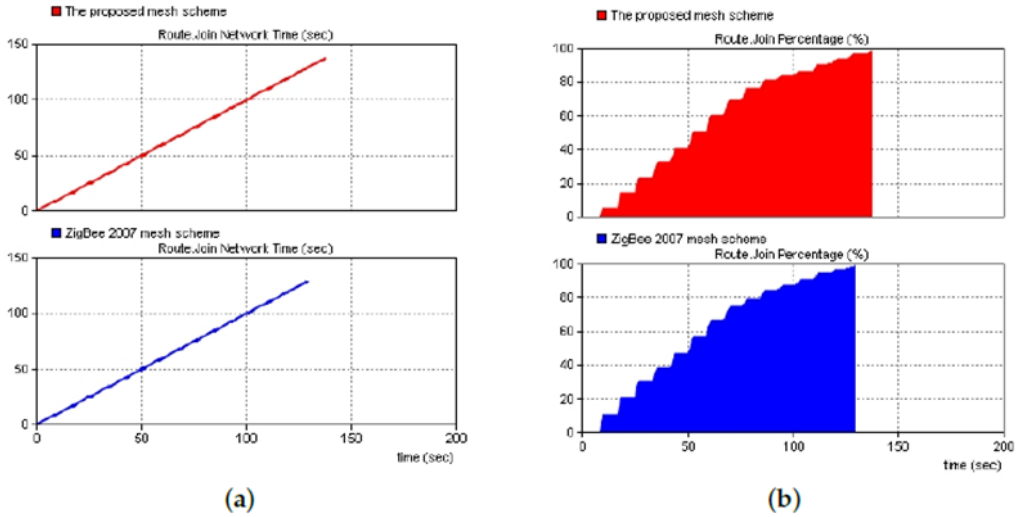

Figure 12: Comparisons of the networking time and propertions of networking nodes for the proposed solution and the ZigBee 2007 solution (a) Networking time (b) Proportion of networking nodes

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.

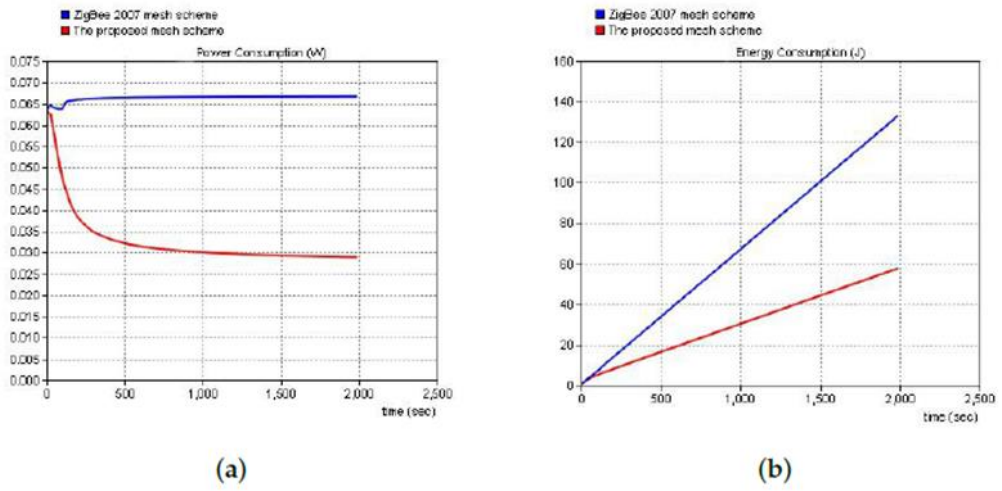

Figure 13: Comparisons of the networking power and energy consumption for the proposed solution and the ZigBee 2007 solution (a)Power (b)Energy

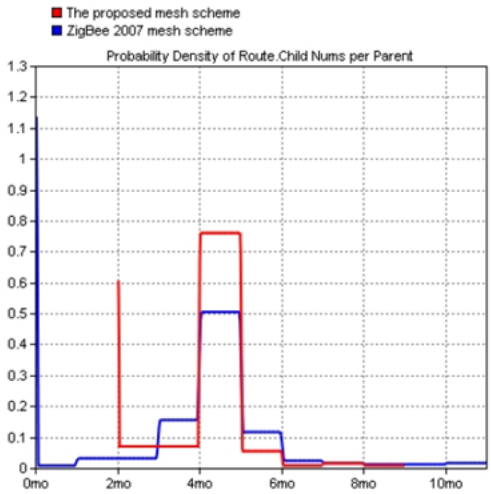

Figure 14: Ratio of terminal nodes carried by routing nodes for the project solution and the ZigBee 2007 solution

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, g) were set to an integer power of two; the range of C was set as $C \in [2^{-5}, 215]$; and the range of g was set as $g \in [2^{-15}, 25]$. The grid-search method was used, and $21 \times 21 = 441$ for (C, g) 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 $C \geq 22$, $g \leq 20$, and $21 \leq C \times g \leq 27$, the SVM model based on the radial basis kernel provided good pipeline leakage signal identification performance.

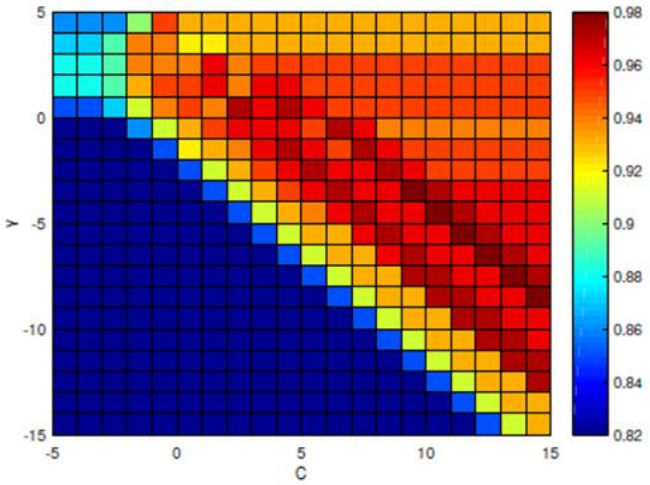

Figure 15: Identification accuracy under different(C,y) parameter combinations

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, g) = (29, 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.

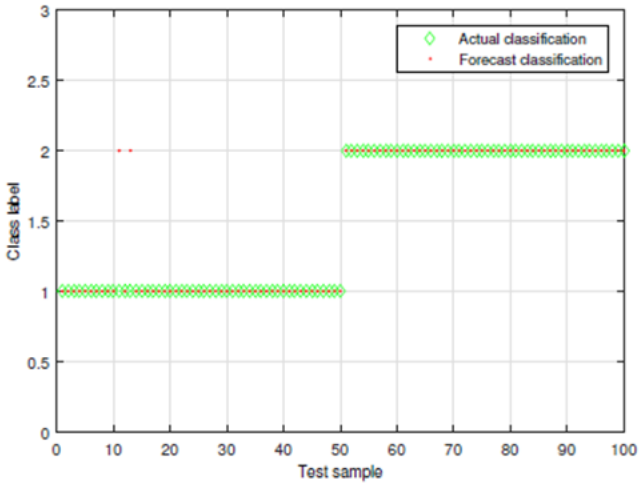

Figure 16: Pipeline leakage and non leakage detection results

Table 5: Leakage signal identification role in an enviroment containing Gaussian noise and impular noise

(c,g)\SNR(DB)		-12	-9	-6	-3	0	3	6	9	12
Gaussian Noise	$(2^{9},2^{-4})$	34	42	60	83	85	88	91	92	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	96
	$(2^{13},2^{-8})$	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	32	55	81	84	87	91	93	95
	$(2^{15},2^{-10})$	31	32	55	80	84	87	91	93	95

CONCLUSION

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.

ABBREVIATION

The following abbreviations are used in this paper:

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.
[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.
[4]. Fang, D.; Chen, B. Ecological network analysis for a virtual water network. Environ. Sci. Technol. 2015, 49, 6722–6730.
[5]. Moore, N.M. The Green City: Sustainable homes, sustainable suburbs. Nicholas Low, Brendan Gleeson, Ray Green, and Darko Radovi´c. Urban Geogr. 2009, 30, 927–928.
[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.
[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.
[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.
[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.
[10]. Lay-Ekuakille, A.; Vergallo, P. Decimated signal diagonalization method for improved spectral leak detection in pipelines. IEEE Sens. J. 2014, 14, 1741–1748.
[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.
[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.
[15]. Benkherouf, A.; Allidina, A.Y. Leak detection and location in gas pipelines. IEE Proc. D Control. Theory Appl. 1988, 135, 142–148.
[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,
[17]. Como, Italy, 18–20 May 2004; pp. 2273–2277.

[18]. Ellul, I.R. Advances in pipeline leak detection techniques. Pipes Pipelines Int. 1989, 34, 7–12.
[19]. 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.
[20]. 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.
[21]. 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.
[22]. Hunaidi, O.; Chu, W.T. Acoustical characteristics of leak signals in plastic water distribution pipes. Appl. Acoust. 1999, 58, 235–254.
[23]. 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.
[24]. Kang, J.; Park, Y.; Lee, J.;Wang, S.; Eom, D. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water istribution systems. IEEE Trans. Ind. Electron. 2018, 65, 4279–4289.
[25]. Chraim, F.; Erol, Y.B.; Pister, K.Wireless gas leak detection and localization. IEEE Trans. Ind. Inform. 2016,12, 768–779.
[26]. 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. Sensors 2019, 19, 5086 20 of 21
[27]. 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.
[28]. Poulakis, Z.; Valougeorgis, D.; Papadimitriou, C. Leakage detection in water pipe networks using a Bayesian probabilistic framework. Probabilistic Eng. Mech. 2003, 18, 315–327.
[29]. 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 onWater Pipeline Systems, York, UK, 28–30 March 2001; pp. 28–30.
[30]. 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.
[31]. 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.
[32]. Ali, S. SimpliMote: A wireless sensor network monitoring platform for oil and gas pipelines. IEEE Syst. J. 2018, 12, 778–789.
[33]. 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.
[34]. 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.
[35]. 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.
[36]. 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.
[37]. 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.
[38]. 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.
[39]. 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.

[40]. 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.
[41]. Pławiak, P.; So´snicki, T.; Nied´zwiecki, 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.
[42]. 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.
[43]. Malfante, M. Machine learning for volcano seismic signals: Challenges and perspectives. IEEE Singal Process. Mag. 2018, 35, 20–30.
[44]. 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.
[45]. 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.
[46]. Gomaa, R.I. Real-time radiological monitoring of nuclear facilities using Zigbee technology. IEEE Sens. J. 2014, 14, 4007–4013.
[47]. Zheng, K. Energy efficient localization and tracking of mobile devices in wireless sensor networks. IEEE Trans. Veh. Technol. 2017, 66, 2714–2726.
[48]. 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. Sensors 2019, 19, 5086 21 of 21
[49]. Yang, J.; Wen, Y.; Li, P. Leak location using blind system identifi-cation in water distribution pipelines. J. Sound Vibrat. 2008, 310, 134–148.
[50]. 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.
[51]. Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297.
[52]. 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.
[53]. Ayat, N.E.; Cheriet, M.; Suen, C.Y. Automatic model selection for the optimization of SVM kernels. Pattern Recognit. 2005, 38, 1733–1745.
[54]. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. 2002, 13, 415–425.
[55]. 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.