



Pipeline Fault Diagnosis Using Wavelet Entropy and Ensemble Deep Neural Technique

Bach Phi Duong and Jong-Myon Kim^(✉)

School of Electrical Engineering, University of Ulsan, Ulsan 680-749, South Korea
duongbachphi@gmail.com, jongmyon.kim@gmail.com

Abstract. The maintenance of pipelines is essential for the safe and cost effective transport of important fluids such as water, oil, and gas. The early detection of pipeline faults is vital for avoiding material and economic losses, and more importantly for ensuring the safety of both human life and the environment. This paper proposes a methodology for early fault detection in pipelines using an acoustic emission (AE) based technique. The proposed method incorporates wavelet entropy analysis of the AE signals and ensemble deep neural networks for the effective detection of different types of faults in a pipeline. The proposed method is tested on an experimental testbed, and the results indicate that it can detect various faults in the pipeline with an average accuracy of 96%.

Keywords: Pipeline fault diagnosis · Acoustic emission · Wavelet entropy
Ensemble deep neural network

1 Introduction

The transport of fluids such as water, oil, and gas is primarily carried out through pipelines. The safe and reliable transport of these and many other fluids is central to many aspects of modern civilization, i.e., air conditioning, water supply and sanitation to name a few. Moreover, many industries such as construction, chemical and petroleum, process huge amounts of fluids that are transported primarily through pipelines. Pipelines are prone to many serious problems such as internal or external corrosion, the risk of cracks due to increase in fluid pressure, and welding defects; all of which can cause leakages. Leakages in pipes can have many detrimental effects including the disruption of an industrial process, the pollution of the environment, health and safety risks to human life, and economic losses. Moreover, they can lead to serious safety accidents that can cause loss of human life [1].

To mitigate the detrimental effects of pipeline leakages, the development of a reliable technique for the timely detection of leakage is essential. The common methods for the detection of a leakage involve the analysis of fluid flow parameters such as flow rate and pressure [2–4]. However, these methods detect the leakage after it has occurred. Ideally, it is more desirable to prevent leakages by detecting the development of cracks that can lead to leakages in pipes. In this paper, an attempt has been made to develop a technique that can help in preventing pipeline leakages by detecting developing cracks in the walls of a pipe. Developing cracks in a material manifest themselves by emitting high

frequency surface waves or acoustic emissions (AE) [1]. Thus, AE based methods involving the analysis of these acoustic emissions is a promising area for research into the detection and prevention of leakages in pipelines. The development of good AE sensors has made the collection of huge amounts of AE data for different pipeline conditions very easy, thereby facilitating the application of machine learning techniques for pipeline fault diagnosis. In recent years, many data-driven methods have been developed for fault diagnosis in industrial piping networks based on machine learning techniques such as k-nearest neighbors [5], hidden Markov models [6], and Support Vector Machine [7]. Although, these methods have demonstrated some potential, nevertheless, there are many unresolved issues. The time and frequency properties of AE signals are affected by many factors such as the fault size, the fault type, the operating conditions of the pipes, and the position of the AE sensor. Thus, it is difficult to get signal spectra that are truly representative of a leakage, a crack, or a non-leakage condition. Moreover, the extraction of feature vectors from AE signals that can be used to distinguish between different conditions of a pipe is also very challenging.

To address these problems, a new method is proposed that employs an improved wavelet packet algorithm with wavelet entropy analysis and an ensemble deep neural network (EDNN) to detect different types of faults in a pipeline. Since, in this work early stage pipeline fault detection is investigated as a classification problem, an improved wavelet packet transform with wavelet entropy analysis is proposed to extract features from the recorded AE signals. Then an EDNN is constructed as a classifier to recognize the different types of pipeline faults.

The remainder of this paper is organized as follows: In Sect. 2, the basics of AE signal based pipeline fault detection are introduced. Section 3 describes the details of the proposed data-driven algorithm for pipeline fault diagnosis. Section 4 presents the experimental testbed and evaluates the performance of the proposed method through data collected through the experimental testbed. Finally, Sect. 5 concludes this paper.

2 Diagnosis of Pipeline Faults Using AE Signals

The basic idea of AE based methods for the diagnosis of faults in pipelines is to measure the response of the pipeline to the input over time. Afterwards, this response is compared to the baseline measurements to determine the type of fault in the pipeline. For the detection of leakage in a pipeline, the principle of this method is based on the observation that when a leak happens in a pressurized pipeline, the flow of the fluid becomes turbulent around the leakage point. Both the turbulence and the loss of fluid through the leakage increase the stress on the pipeline wall around that region. This increasing stress on the leakage point results in elastic waves in the pipeline material that propagate as acoustic emissions originating from the leakage point. Acoustic emission sensors placed on the outer surface of pipelines can be used to measure the energy of these emissions. Since the AE wave propagates through the material, any unusual point (such as leak or crack) in the signal propagation path affects the amplitude and speed of the wave. These effects can be used to detect surface defects in the pipe. Another mechanism for producing the AE signal is the initiation and propagation of fracture in a pipe. However, when a pipeline

system operates longer than a certain amount of time, its characteristics can be measured dynamically with the help of vibration testing [8, 9]. When a fluid is flowing under pressure in a pipeline, the impact of fluid on the pipe walls causes vibration in the pipeline, which stimulates metal to metal contact in a developing crack and expands the plastic deformation due to local contact phenomenon at the surface of the fracture to generate acoustic emissions. However, the time-frequency characteristics and the statistical properties of these AE signals are non-stationary and therefore require the use of tools such as wavelet entropy analysis.

3 The Proposed Methodology for Pipeline Fault Diagnosis

3.1 Wavelet Entropy Analysis

Wavelet entropy (WE) measures the entropy of wavelet coefficients of an AE signal that are obtained through wavelet packet decomposition [10]. It has the benefits of evaluating the complexity of the non-stationary AE signals at multiple resolutions. In this work, it is used for feature extraction to identify the patterns associated with leaks and cracks. First, the AE signal is decomposed using the wavelet packet transform into multiple sub-bands. The wavelet coefficients in the k^{th} sub-band of the level j^{th} with $0 \leq k \leq 2^j - 1$ are denoted by $C_{j,k}(n)$. The wavelet energy at each level j is given by $E_j = (C_j)^2$. The total energy can, therefore, be given as $E_{\text{total}} = (1/N) \sum_j E_j$, where N is the number of the wavelet coefficients. Then, the relative wavelet energy is defined as $p_j = E_j/E_{\text{total}}$. Thus, according to the definition of Shannon entropy, the wavelet entropy can be calculated as follows:

$$WE = - \sum_j p_j \log p_j \quad (1)$$

The entropy can be used to measure the amount of information or complexity of the output signal. Pipeline leakage signals have properties of complexity, uncertainty, and non-linearity from which entropy can be used as a measure of signal complexity. In fact, a well-ordered process could be considered as a periodic signal with only one frequency. The wavelet representation of such a signal will mostly be determined by a single decomposition level, i.e., all relative wavelet energies will be nearly zero except for the wavelet resolution level that includes the representative signal frequency. For this level, the relative wavelet energy will be nearly one. And as a result, the total WE will be very low. A signal generated by a stochastic process can be represented as a disarranged response. Such a signal has a wavelet representation with significant contributions from all frequency bands. Moreover, all the contributions may be of the same order. Consequently, the relative wavelet energy is equal for all levels and the WE will have a high value. The WE is therefore used to evaluate different sub-bands of the AE signal and select the ones with the most fault information. Once the sub-band with the most information is determined, it is reconstructed and then divided into multiple segments. For each segment, two features, i.e., the root mean square value (RMS) and wavelet entropy (WE), are calculated. The values of these features for all the segments of the

reconstructed signal are merged into a feature vector, which is then used to train an EDNN. The trained EDNN is then used for the diagnosis of pipeline faults in unknown AE signals.

3.2 Ensemble Deep Neural Network with Optimized Model Using Genetic Algorithm

As mentioned earlier, the feature vectors obtained from the optimal reconstructed signals, i.e., sub-band with the most information, are used to train the ensemble deep neural network (EDNN). Each feature vector contains 1000 elements, i.e., RMS and WE values for each of the five hundred segments of the optimal reconstructed signal. These feature vectors are used to train an EDNN with six layers. To identify the number of units in each layer, this paper employs the genetic algorithm (GA) [11]. The configuration of the EDNN is thus represented by a chromosome consisting of multiple genes, where each gene corresponds to the number of neurons in each layer. The number of nodes for each layer are in a predefined range. To reduce the dimensions of the feature vector before the last soft-max layer, these ranges are decreased after each layer. The values of the number of neurons are arranged in arrays of the same length, which is helpful in encoding with the same number of bits. The index of each value is encoded by an n -bit encoder resulting in strings of bits, which are combined to construct chromosomes. In the same way, other chromosomes are randomly initialized from the predefined ranges to create a population of chromosomes. The accuracy of the EDNN training is used to evaluate the fitness of each chromosome or each candidate configuration of the EDNN. As the training of the EDNN is a stochastic process, therefore, for each configuration the EDNN is trained five times, and the average accuracy is computed as the fitness function. The selection of chromosomes is done through tournament selection with a tournament size of three. A two-point crossover operator is used, and genes are mutated by randomly flipping bits. The crossover and mutation rates are set to 0.5 and 0.2, respectively. The optimal individual from the GA is used to configure the architecture of the EDNN. The EDNN uses the Rectified Linear Unit (ReLU) as an activation function to avoid vanishing gradients, speed up the convergence of the training and yield better solutions. The definition of ReLU is $y = \max(0, a)$, where $a = Wx + b$. The constant gradient of ReLUs helps the EDNN in faster learning. Adam optimization, i.e., the algorithm for first-order gradient-based optimization that is an extension to the stochastic gradient descent, is used as the optimization mechanism to reduce the training time by training the EDNN with a larger effective step size. The output layer uses soft-max logistic regression. The normalization initiation is, therefore, necessary when initializing EDNN because of the multiplicative effect across two layers, and we suggest the following initialization strategy to sustain activation variances and back-propagated gradients variance for the forward and backward flow. The j^{th} weight of the i^{th} layer are initialized by normalized initialization with

$$W_{ij} \sim U \left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \quad (2)$$

where $U[-a, a]$ is the uniform distribution with values in the interval $(-a, a)$ and n is the size of the front layer. The normalized initialization can be quite helpful, presumably because the layer-to-layer transformations maintain magnitudes of activations and gradients. Other hyper-parameters of the network are as follows. A dropout rate of 0.7 is used, whereas the batch size is 10. The learning rate is $1e-3$. The total number of epochs is fixed at 100. The net output is the classification of different AE signal to different fault labels.

4 Experimental Setup and Results

4.1 Data Acquisition

The proposed method is tested on data collected through a pressurized water pipeline network that is designed to mimic field transmission pipelines as shown in Fig. 1. A pump is used to maintain a constant flow rate of water in the piping network. The pressure is held constant at 3 bar. The AE signals are recorded using two AE sensors of RTS WDI-AST type with an operating frequency range of 200–900 kHz. The two AE sensors are placed at both sides of the test pipe section and the AE signals are recorded under different conditions. A pre-amplifier with 96 dB gain is used to amplify the AE signals for subsequent processing. The AE sensors are installed on the outer surface of the pipe, and a data acquisition system with PCI-DAQ board is used to record the AE signals at a sampling rate of 1 MHz. The AE signals are then decimated to 250 kHz. The duration of each AE signal is one second. In this work, four types of pipeline conditions are considered including a normal pipeline, a pipeline with a 5 mm crack, a pipeline with a 10 mm crack, and a pipeline with a 10 mm leak hole. The valves are used to simulate

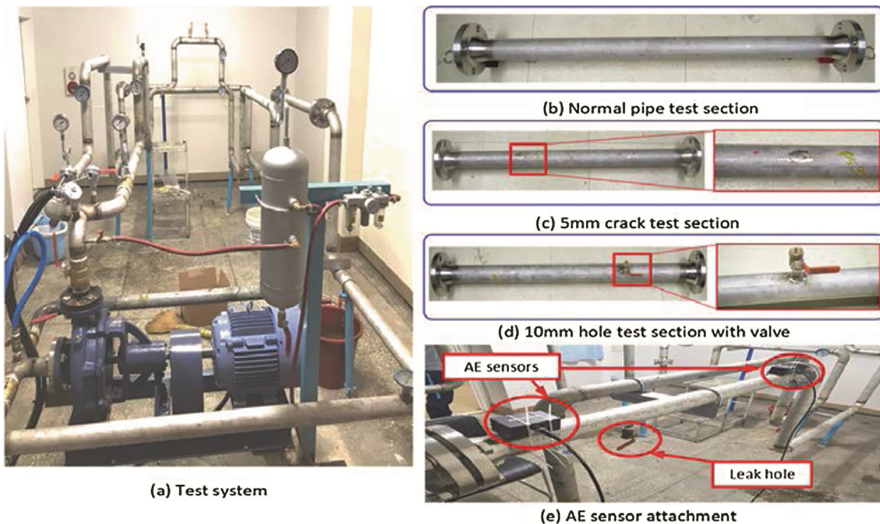


Fig. 1. Test rig set up: (a) Pipeline test system (b) Normal pipe test section (c) 5 mm crack test section (d) Sensor attachment

the leaks. The normal case is also recorded as the normal baseline signal. The EDNN is implemented using Google TensorFlow on a general computing platform with a GeForce GTX 1080 Ti GPU.

4.2 Experimental Results

To extract the features of AE signals that can be used to distinguish between different conditions of the pipeline, i.e., normal and faulty, wavelet entropy spectral analysis is performed on each AE signal. Figure 2 shows the wavelet entropy scalograms of AE signals for pipelines of different health conditions. It can be observed in Fig. 2 that the pattern of the wavelet entropy scalogram is different for different type of pipeline condition. The scalograms show the shift in the frequency band of the entropy energy, which is useful in discriminating different types of faults. Then the optimal sub-band is chosen based upon the maximum wavelet entropy. The optimal sub-band is then used to reconstruct the AE signal, which is then divided into multiple segments. Figure 3 shows the waveforms of a few segments of the optimal reconstructed signal for different pipeline conditions. Features extracted from these segments are then used to train the EDNN with the optimal configuration. Figure 4 shows the confusion matrices and the accuracy of the EDNN during testing. The rows of the matrices in Fig. 4 represent the actual labels of the classes, whereas the columns represent predicted states. The average classification accuracy of the proposed method using EDNN is 96%. The method is also compared with Support Vector Machine (SVM) and the Stacked Denoising Autoencoder (SDAE), both of which are trained using the same data. The proposed method with an average classification accuracy of 96%, outperforms both SVM and SDAE with accuracies of 93% and 82%, respectively. The results indicate that the proposed method using EDNN renders better results in comparison to SVM and the SDAE. Moreover, most of the misclassification occurs between the 10 mm crack and 10 mm hole classes, which is due to the relatively greater similarity between the signals for the two conditions as shown in Fig. 3.

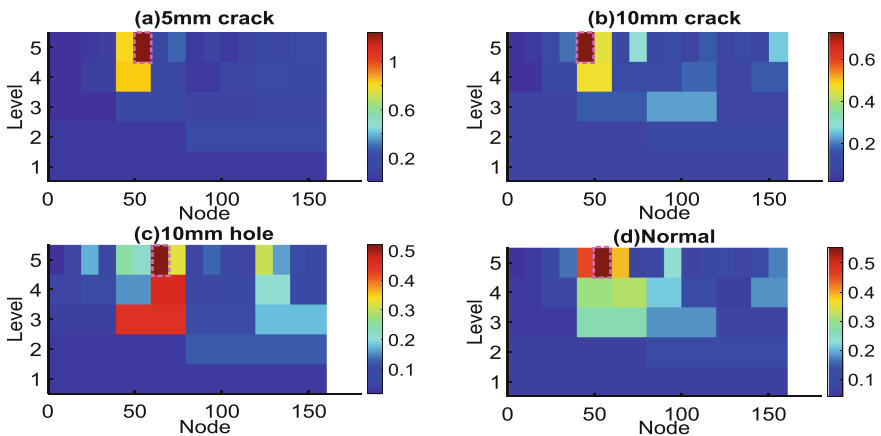


Fig. 2. Wavelet entropy scalogram of difference fault signal type

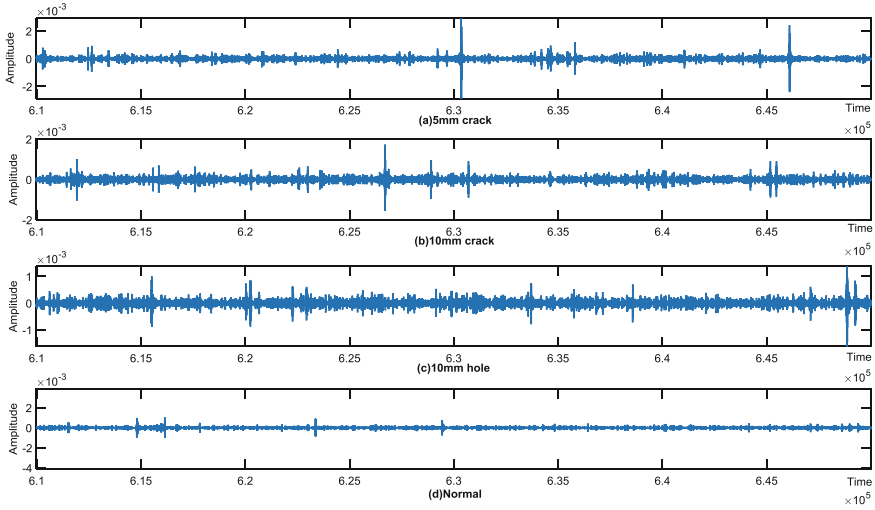


Fig. 3. Segments of optimal signal of (a) 5 mm crack (b) 10 mm crack (c) 10 mm hole (d) Normal

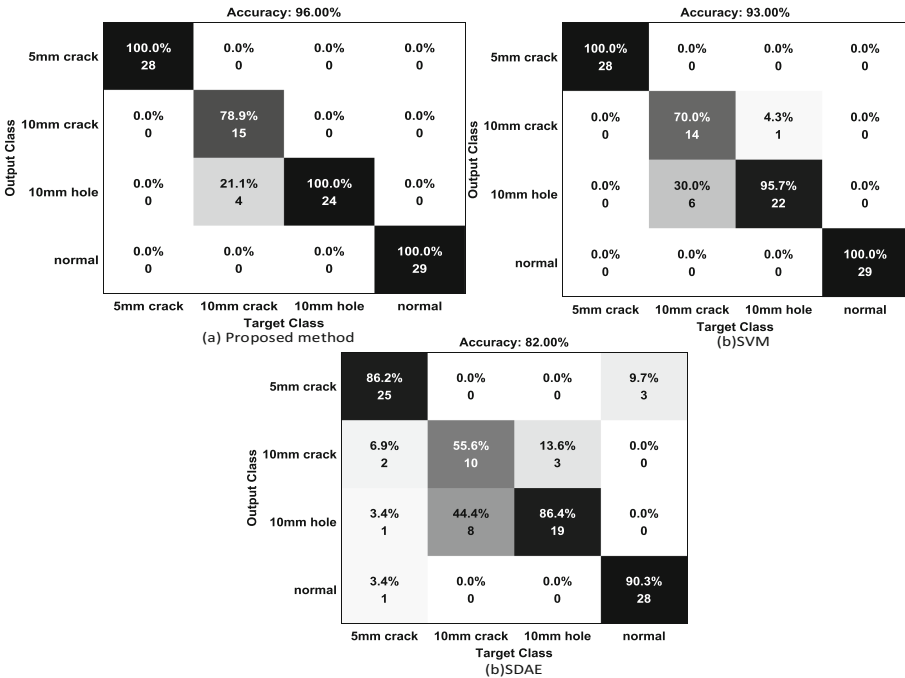


Fig. 4. Confusion matrix result of (a) Proposed method (b) SVM (c) SDAE

5 Conclusions

In this paper, a new method is proposed for early stage fault detection in pipelines. The proposed methodology employs wavelet entropy analysis to determine the optimal sub-band of the acoustic emission signals for the extraction of features to distinguish between different pipeline faults. The optimal sub-band is then used to reconstruct a signal, which is then divided into multiple segments. Features extracted from these segments are then used to train an EDNN. The trained EDNN is then used to classify unknown AE signals. Experiments on data obtained through an experimental testbed show that the proposed method can diagnose different types of pipeline faults with 96% accuracy.

Acknowledgements. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (Nos. 20162220100050, 20161120100350, 20172510102130). It was also funded in part by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2016H1D5A1910564), and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

References

1. Xu, Q., Zhang, L., Liang, W.: Acoustic detection technology for gas pipeline leakage. *Process Saf. Environ. Prot.* **91**, 253–261 (2013)
2. Murvay, P.-S., Silea, I.: A survey on gas leak detection and localization techniques. *J. Loss Prev. Process Ind.* **25**, 966–973 (2012)
3. Taghvaei, M., Beck, S.B.M., Staszewski, W.J.: Leak detection in pipelines using cepstrum analysis. *Meas. Sci. Technol.* **17**, 367 (2006)
4. Lee, P.J., Vítkovský, J.P., Lambert, M.F., Simpson, A.R., Liggett, J.A.: Leak location using the pattern of the frequency response diagram in pipelines: a numerical study. *J. Sound Vib.* **284**, 1051–1073 (2005)
5. Valizadeh, S., Moshiri, B., Salahshoor, K.: Leak detection in transportation pipelines using feature extraction and KNN classification. In: *Pipelines* (2009)
6. Ai, C., Zhao, H., Ma, R., Dong, X.: Pipeline damage and leak detection based on sound spectrum LPCC and HMM. In: *Sixth International Conference on Intelligent Systems Design and Applications*, pp. 829–833 (2006)
7. Sato, T., Mita, A.: Leak detection using the pattern of sound signals in water supply systems. In: *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2007*, p. 65292K. International Society for Optics and Photonics (2007)
8. Oterkus, E., Yang, Z.: Corrosion detection in pipelines based on measurement of natural frequencies. *Ann. Limnol. Oceanogr.* **2**, 1–6 (2017)
9. Yi-min, H., Yong-shou, L., Bao-hui, L., Yan-jiang, L., Zhu-feng, Y.: Natural frequency analysis of fluid conveying pipeline with different boundary conditions. *Nucl. Eng. Des.* **240**, 461–467 (2010)

10. Rosso, O.A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., Başar, E.: Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *J. Neurosci. Methods* **105**, 65–75 (2001)
11. Whitley, D.: A genetic algorithm tutorial. *Stat. Comput.* **4**, 65–85 (1994)