PAPER • OPEN ACCESS

Leak diagnosis of pipeline based on empirical mode decomposition and support vector machine

To cite this article: A Faysal *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1078** 012023

View the article online for updates and enhancements.

Leak diagnosis of pipeline based on empirical mode decomposition and support vector machine

A Faysal[1], M S N A Adhreena[2], E Vorathin[2], Z M Hafizi[2] and W K Ngui[1,*]

[1]College of Engineering, Universiti Malaysia Pahang (UMP), 26300 Gambang, Pahang, Malaysia.
[2]Advanced Structural Integrity and Vibration Research (ASIVR), Faculty of Mechanical & Automotive Engineering Technology, Universiti Malaysia Pahang (UMP), 26600 Pekan, Pahang, Malaysia.

*Corresponding author's e-mail: wkngui@ump.edu.my

Abstract. The pipeline is used as a medium of transportation in global gas and oil industries, providing the most efficient, convenient and transportation method for natural gas and oil from downstream to upstream production of the economical mode of the power station, refineries, and domestic needs. However, the pipeline leakages become a major concern as their failure may contribute to operational and economic loss as well as environmental pollution. This paper proposed a system to detect pipe fault at different locations. Empirical Mode Decomposition (EMD) was applied for feature extraction using energy and kurtosis. The one-against-one (OAO) and one-against-all (OAA) multiclass SVM with radial basis function (RBF), polynomial and sigmoid kernel functions were implemented in order to classify the multiple fault locations from the extracted features. RBF kernel function recorded the highest classification accuracy for both OAO and OAA approaches with 97.77% and 96.29%, respectively, followed by slightly reduced accuracy for sigmoid whereas significantly low accuracy for the polynomial kernel. The outputs were further analysed to justify the performance of the classifiers. From all the cases, it was observed that OAO-SVM with RBF kernel performed the best for pipe fault diagnosis.

Keywords: Fault Diagnosis; Condition Monitoring; Support Vector Machine.

1. Introduction

The use of pipelines for fluid transportation has been widely utilised across the world. However, over time, the pipelines are susceptible to leakage due to several factors such as corrosion, degradation of the material, poor quality of fittings and many more [1]. Leakage in pipelines will not only consequence in the operational loss but also to environmental pollution. Therefore, effective leak detection in the pipeline is essential. Over the years, leak detection using vibration analyses are the most widely reported.

Karkulali *et al.* [2] reported on the use of a piezo thin-film sensor to capture the leak signals and Fast Fourier Transform (FFT) signal processing was used to analyse the leak. As a result, the FFT spectrum capable of distinguishing between no leak and leak conditions. However, the authors did not further report on the classification of the leak in terms of size and distance. Wijayanto *et al.* [3] utilised the same signal processing method to determine the size of the leak in a pipeline. However, the results showed that the FFT spectrum could not differentiate between no leak and small leak conditions. The

frequency-domain was only significant when the leak was massive. According to Bentoumi *et al.* [4], the used of FFT signal processing was also associated with drawbacks in terms of noise, fault geometry and fluid nature. However, for various limitation of frequency domain analysis for vibration signals, time-frequency adaptive signal processing can effectively mitigate the discrepancies.

An unconventional, relatively new and a robust time-frequency analysis, EMD was first presented in 1998. Since then, EMD has been applied to stationery and nonstationary signals in numerous fault diagnosis research [5, 6]. The principle of EMD is based on the signal's local characteristics in the time scale. The time scale is divided into a set of orthogonal components known as intrinsic mode function (IMF). IMFs are determined by the signal, not some kernel. Therefore, EMD is self-adaptive and can be applied efficiently to any signal. Since EMD is so well fit for any signal, it has been applied affluently in condition monitoring and fault diagnosis research. Jiang *et al.* [7, 8] used an improved EMD for rolling element bearing fault diagnosis where the results showed that the proposed method can not only diagnose known faults but also monitor unknown faults with strong, robust performance. Lei *et al.* [7, 8] used locomotive rolling element bearings to demonstrate the performance of EMD. In this paper, EMD is applied to the pipe fault signal for effective fault feature extraction.

An effective fault diagnosis method would have the adaptability to accommodate to signal variations and different system. Machine learning algorithms meet these criteria and that makes them a helpful tool in fault diagnosis research. SVM is a powerful supervised learning algorithm which analyses data for regression analysis or non-probabilistic classification. SVM uses hyperplane to maximise the distance between the two classes. Gao *et al.* [9] applied wavelet packet transform (WPT) to extract features and used SVM for classification of the samples in reciprocating pumps fault diagnosis. Yuan and Chu [10] successfully applied SVM for fault diagnosis of the turbo-pump rotor. However, SVM has many kernel function and approach for multiclass classification where the hyperparameter tuning dominates the classification accuracy. So, it is important to investigate which multiclass SVM approach with the right kernel function performs the best with the feature extraction algorithm. In this paper, EMD was used for feature extraction, while SVM was used for classification.

2. Theoretical background

2.1. EMD
EMD is a new time-frequency adaptive method which uses the shifting process to decompose the signal into IMFs. The shifting process is conducted by connecting the local maxima and minima to form the upper and lower envelopes. The sifting process is repeated until the following two conditions satisfy:
(a) The number of extrema in the whole time series must differ with the zero-crossing value by at most one.
(b) The mean value of the upper and lower envelopes at any given location is zero.

An IMF, c_1, is produced when the above-mentioned conditions are satisfied. The residue is obtained by subtracting the IMF_1 from the main signal which is denoted by r_1 and is used to obtain the next IMFs. The decomposition is performed n times until the residue gets monotonic. By adding the IMFs and the remaining residue, the original signal can be obtained as follows:

$$X(t) = \sum_{i=1}^{n} c_i + r_n \tag{1}$$

where X(t) is the original signal, c_i is the ith IMF, and r_n is the nth residue.

2.2. SVM

SVM uses hyperplane to maximise the distance between the two classes. An optimal hyperplane is obtained that represents the margin of the vectors known as support vectors. This type of SVM uses a linear decision boundary and is called linear SVM and expressed as follows:

$$w.x + b = 0 \tag{2}$$

which indicates

$$y_i(w.x + b = 0) \geq 1, i = 1, \cdots, N \tag{3}$$

The goal of SVM is to obtain the highest separating hyperplane by minimizing $\|w\|$ using the conditions. Here $\|w\|$ the Euclidean norm of w, and $2/\|w\|$ denotes the distance between the hyperplane and the nearest data points of each class. By introducing Lagrange multipliers α_i, the SVM is trained to solve a convex quadratic problem (QP). A unique globally optimized result is the solution, which contains the following properties:

$$w = \sum_{i}^{N} \alpha_i y_i x_i \tag{4}$$

Only if corresponding $\alpha_i > 0$, these x_i are called support vectors.

The decision function of the trained SVM can be expressed as:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i (x \cdot x_i) + b\right) \tag{5}$$

When the observations are non-linearly separable, SVM performs a nonlinear mapping of the input vector x from the input space R^N into a higher dimensional Hilbert space, where the mapping is determined by the kernel function. Different types of kernels are suitable for different classification problems to reach the optimal classification error.

3. Experimental procedures and results

The test rig setup for the collection of the vibration signal from the pipe is shown in figure 1. An air storage tank with the length and diameter of 120 cm and 11.43 cm was used to replicate the leaks in a pipeline. A total of 5 plug fittings with a hole diameter of 1 mm were welded along the length of the pipe. The distance between each plug was fixed at 20 cm. A ball valve was fitted to the plug to control the leak. The storage tank was pressurised by using centralised compressed air system. The pressure inside the tank was monitored by using a pressure gauge. An accelerometer with the sensitivity of 100 mV/g was used to capture the leak signals. The accelerometer was attached at the distance of 15 cm away from leak 1. The leak signals were acquired by using National Instruments data acquisition model NI 9234. The signals were captured at a sampling rate of 25.6 kHz. Initially, the storage tank was pressurised with a pressure of 0.2 MPa and the signal was recorded for 1 minute which is labelled as no leak condition. Next, the ball valve at 15 cm was opened to induce the leak and the signals were recorded for 1 minute. The same procedures were repeated for leaks at 35 cm, 55 cm, 75 cm and 95 cm away from the accelerometer. The total leak type considered in this study is illustrated in table 1.
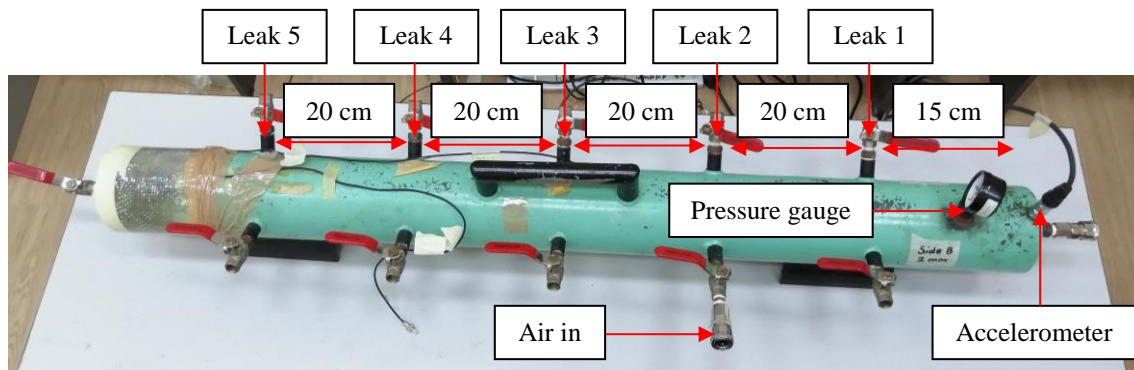
Figure 1. Experimental setup of pipe fault test rig.

Table 1. Pipe leak at different locations from the sensor.

Leak type	No leak	Leak 1	Leak 2	Leak 3	Leak 4	Leak 5
Fault class	C1	C2	C3	C4	C5	C6

3.1. Classification modelling

For feature extraction, 6250 data points were taken per sample. For each condition, 75 samples were obtained and for all six different conditions, the total number of samples is 450. EMD was applied to decompose the vibration signals into several IMFs. It is to be noted that the lower order IMFs carry the most information and for higher order, the information decreases. So, taking all the IMFs into consideration would increase the computational burden and decrease the accuracy. In this paper, only the first five IMFs are considered [11].

In this paper, two statistical parameters, energy and kurtosis are used to obtain the feature vector. The summation of the energy can obtain the total energy of a signal in each sub-band and energy feature in each band. Kurtosis is a dimensionless statistical measure that characterises the flatness of a signal's probability density function defined as the fourth-order moment of signal data and kurtosis of the sub-band coefficient [10.1007/s00500-013-1055-1]. From each condition, out of five IMFs, ten features were extracted which consist of energy and kurtosis for each IMF. Table 2 represents one feature vector row from each fault class chosen randomly for visualization of the features. The final size of the feature vector is 450x10 for the entire feature set. This feature vector was normalised between 0 to 1 and fed into the SVM for classification. OAO SVM and OAA SVM were implemented to classify the faults at different locations. In SVM, kernel function is crucial to get desired classification accuracy. In this study, the radial basis function (RBF), polynomial and sigmoid kernel functions are used. To evaluate the performance of the classifier, 70% of the data was used to train the classifier and the rest 30% to test the accuracy. So, the number of samples obtained for training and testing were 315 and 135, respectively. The data was partitioned randomly which results in uneven distribution of the number of samples in different fault class.

4. Results and discussion

All three combinations of the kernel were applied to OAO and OAA approach of SVM. From table 3, it is observed that the RBF kernel obtained the highest classification accuracy for both approaches, whereas the polynomial kernel obtained the lowest. The testing accuracy with RBF in OAO was slightly higher than OAA. In both cases of sigmoid, the testing accuracy was slightly lower than RBF, which is 94.81% and 94.07% for OAO and OAA approach, respectively. Since the accuracy of the polynomial kernel in both OAO and OAA was significantly low, no further analysis was conducted with it. The confusion matrix with precision, recall and F1 score for RBF and Sigmoid from both OAO and OAA approach are provided in table 4 and table 5.

Table 2. Representation of the feature vector for different class.

Fault Class	IMF1		IMF2		IMF3		IMF4		IMF5	
	Energy	Kurtosis	Energy	Kurtosis	Energy	Kurtosis	Energy	Kurtosis	Energy	Kurtosis
No leak	0.001228	0.000434	0.000214	0.000176	0.000148	2.3732	3.0511	3.2202	2.8091	4.0188
Leak 1	2.9025	0.48943	0.10441	0.030102	0.009964	2.3998	3.1116	3.8018	3.2697	2.7446
Leak 2	4.7834	0.66713	0.14955	0.039948	0.019658	2.3546	3.1181	3.7326	3.4483	2.381
Leak 3	2.1343	0.88299	0.069758	0.019215	0.008126	2.2583	3.0966	3.5044	3.7247	3.9072
Leak 4	2.3039	0.42037	0.050061	0.012056	0.006248	2.3876	3.0483	3.7251	2.9731	4.3754
Leak 5	2.6432	1.7895	0.11299	0.033748	0.010545	2.9254	3.033	5.2849	3.9484	4.6836

In all results of table 4 and 5, although the number of samples varies for individual class because of the random partition of feature set, the total observation is 135. Among all four cases, since OAO-SVM with RBF kernel obtained the highest accuracy, its precision, recall and F1 score were also the highest with 0.98, 0.97 and 0.98. respectively. In case of OAO-SVM, the no fault condition could obtain the perfect accuracy for both RBF and sigmoid kernel which is 1 for precision, recall and F-1 score. Only class C2 and C6 in OAO-SVM with RBF kernel could obtain the perfect accuracy too. As expected, the precision, recall and F1 score for the OAA-SVM with RBF kernel was the second highest which were 0.96 for all of them. The performance of sigmoid was lower for both SVM approach where OAO-SVM with sigmoid kernel was slightly more accurate than OAA-SVM with Sigmoid kernel. From table 5(a), it is observed that the first three classes obtained the perfect accuracy whereas in table 5(b) C3 and C6 could obtain it. These differences show why the classifier with RBF kernel was better. In all cases, OAO-SVM performed better than OAA-SVM for the same kernel whereas OAO-SVM with RBF kernel was found to be the best performing classifier in this study.

Table 3. SVM accuracy using different kernel functions.

Approach	Kernel	Test accuracy (%)
OAO	RBF	97.77
	Polynomial	74.81
	Sigmoid	94.81
OAA	RBF	96.29
	Polynomial	75.55
	Sigmoid	94.07

Table 4(a). Confusion matrix with precision, recall and F1 score for OAO-SVM with RBF kernel.

Class	C1	C2	C3	C4	C5	C6	Total	Precision	Recall	F1
C1	27	0	0	0	0	0	27	1	1	1
C2	0	20	0	0	0	0	20	1	1	1
C3	0	0	21	0	1	0	22	1	0.95	0.98
C4	0	0	0	17	1	0	18	0.94	0.94	0.94
C5	0	0	0	1	20	0	21	0.91	0.95	0.93
C6	0	0	0	0	0	27	27	1	1	1
Total							135	0.98	0.97	0.98

Table 4(b). Confusion matrix with precision, recall and F1 score for OAO-SVM with Sigmoid kernel.

Class	C1	C2	C3	C4	C5	C6	Total	Precision	Recall	F1
C1	27	0	0	0	0	0	27	1	1	1
C2	0	20	0	0	0	0	20	0.91	1	0.95
C3	0	1	21	0	0	0	22	0.91	0.95	0.93
C4	0	0	0	16	2	0	18	0.94	0.89	0.91
C5	0	1	0	1	19	0	21	0.90	0.90	0.90
C6	0	0	2	0	0	25	27	1	0.93	0.96
Total							135	0.94	0.95	0.94

Table 5(a). Confusion matrix with precision, recall and F1 score for OAA-SVM with RBF kernel.

Class	C1	C2	C3	C4	C5	C6	Total	Precision	Recall	F1
C1	24	0	0	0	0	0	24	1	1	1
C2	0	21	0	0	0	0	21	1	1	1
C3	0	0	19	0	0	0	19	1	1	1
C4	0	0	0	20	2	0	22	0.91	0.91	0.91
C5	0	0	0	2	17	1	20	0.89	0.85	0.87
C6	0	0	0	0	0	29	29	0.97	1	0.98
Total							135	0.96	0.96	0.96

Table 5(b). Confusion matrix with precision, recall and F1 score for OAA-SVM with Sigmoid kernel.

Class	C1	C2	C3	C4	C5	C6	Total	Precision	Recall	F1
C1	18	0	0	0	0	1	19	1	0.95	0.97
C2	0	24	0	1	0	0	25	1	0.96	0.98
C3	0	0	29	0	0	1	29	1	1	1
C4	0	0	0	14	4	0	18	0.82	0.78	0.80
C5	0	0	0	2	17	0	19	0.77	0.89	0.83
C6	0	0	0	0	0	25	25	1	1	1
Total							135	0.93	0.93	0.93

5. Conclusion

Throughout this study, EMD and SVM were successfully developed for early detection and localisation of small leaks occurred along the pipeline. The OAO and OAA SVM approaches were implemented to classify the multiple fault location with applying the kernel function of RBF, polynomial and sigmoid functions. Compared to others, RBF kernel function recorded the highest classification accuracy in both OAO and OAA approaches with a percentage of 97.77% and 96.29% respectively followed by a slightly lower performance of sigmoid, 94.81% and 94.07%, respectively. On the other hand, the performance of polynomial kernel was significantly poor in both SVM approach. The classification results with RBF and sigmoid were further analysed using confusion matrix, precision, recall and F1 score. It was observed that OAO-SVM with RBF kernel is the best combination since it can obtain the highest accuracy with the best precision, recall and F1 score.

References

[1] Ranginkaman M H, Haghighi A and Vali Samani H M 2016 Inverse frequency response analysis for pipelines leak detection using the particle swarm optimisation *Iran University of Science & Technology* **6** 1-12.

[2] Karkulali P, Mishra H, Ukil A and Dauwels J 2016 Leak detection in gas distribution pipelines using acoustic impact monitoring *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society* 412-6.

[3] Wijayanto A P and Miasa I M 2019 Effects of variations in pipe leakage diameter and pipe leakage distance to the FFT diagrams *Journal of Physics: Conference Series* **1351** 012055.

[4] Bentoumi M, Chikouche D, Mezache A and Bakhti H 2016 Wavelet dt method for water leak-detection using a vibration sensor: an experimental analysis *IET Signal Processing* **11** 396-405.

[5] Lei Y, Lin J, He Z and Zuo M J 2013 A review on empirical mode decomposition in fault diagnosis of rotating machinery *Mechanical systems and signal processing* **35** 108-26.

[6] He Q, Li P and Kong F 2012 Rolling bearing localised defect evaluation by multiscale signature via empirical mode decomposition *Journal of vibration and acoustics* **134.**

[7] Jiang F, Zhu Z, Li W, Chen G and Zhou G 2013 Robust condition monitoring and fault diagnosis of rolling element bearings using improved EEMD and statistical features *Measurement Science and Technology* **25** 025003.

[8] Lei Y, Liu Z, Ouazri J and Lin J 2017 A fault diagnosis method of rolling element bearings based on CEEMDAN *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **231** 1804-15.

[9] Gao J, Shi W, Tan J and Zhong F 2002 Support vector machines based approach for fault diagnosis of valves in reciprocating pumps *IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No. 02CH37373)* **3** 1622-7.

[10] Pan S J and Yang Q 2009 A survey on transfer learning *IEEE Transactions on knowledge and data engineering* **22** 1345-59.

[11] Cheng Y, Wang Z, Chen B, Zhang W and Huang G 2019 An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis *ISA transactions* **91** 218-34.