Pertanika J. Sci. & Technol. 28 (S1): 173 - 184 (2020)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Classification of Faults in Oil and Gas Pipelines using Support Vector Machines

Nagoor Basha Shaik[1]*, Srinivasa Rao Pedapati[1], Syed Ali Ammar Taqvi[2], Shazaib Ahsan[1] and Faizul Azly Abd Dzubir[3]

[1]Department of Mechanical Engineering, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia
[2]Department of Chemical Engineering, NED University of Engineering & Technology, Karachi City, Sindh 75270, Pakistan
[3]Department of Mechanical, Group Technical Solutions, Project Delivery and Technology Division, Petroliam Nasional Berhad, 50050 Kuala Lumpur, Malaysia

ABSTRACT

Leaks and breakdowns of pipelines can lead to catastrophic failures and cause economical losses worldwide. Currently, condition monitoring has become a challenging process because of various reasons such as fluctuating external conditions, natural hazards. Pipelines are installed in severe conditions and are subjected to degradation mainly due to corrosion and metal loss. This study attempted to classify different types of metal loss faults using historical inspection data of oil and gas fields. For this purpose, Support Vector Machines (SVM) were employed to classify and predict various types of metal loss faults which were affecting the life condition of a crude oil pipeline. The historical inspection data was acquired from a crude oil pipeline located in Sudan. Different types of SVM models were trained and quadratic SVM type was selected for the present study due to its high prediction accuracy. The performance evaluation of the proposed SVM model was done using the confusion matrix. The developed SVM model provides promising results with a prediction accuracy of 93.0%. As a result, the fault detection rate (FDR) for all faults is found to be 90.4%, while the misclassification rate (MR) is 9.6%. The prediction of metal loss fault type may help in condition assessment and maintenance schedule to take prior actions for the better life of pipeline which reduces the degradation rate of a pipeline.

Keywords: Classification, metal loss, oil and gas, pipeline, prediction, support vector machines

ARTICLE INFO

Article history:
Received: 26 February 2020
Accepted: 18 May 2020
Published: 16 September 2020

E-mail addresses:
nagoor_16000473@utp.edu.my (Nagoor Basha Shaik)
srinivasa.pedapati@utp.edu.my (Srinivasa Rao Pedapati)
aliammar@neduet.edu.pk (Syed Ali Ammar Taqvi)
shazaibahsan@hotmail.com (Shazaib Ahsan)
faizul.dzubir@petronas.com.my (Faizul Azly Abd Dzubir)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Pipelines are being inspected in various ways such as intelligent pigging, Magnetic flux leakage (MFL) detection techniques, ultrasonic techniques, on a timely basis since decades. Among all methods, smart pigs are widely used by pipeline industries due to their better performance and pinpoint information of faults (Isa & Rajkumar, 2009). Many features like metal loss anomalies, weld anomalies can be detected using these smart pigs which are affecting the life of a pipeline majorly. The rate of metal loss of a pipeline is typically interrelated with external and internal features (Cosham et al., 2007; Nešić, 2007). The metal loss may be of any type like pitting, general, biologically influenced and stray currents (Vanaei et al., 2017). However, it is difficult to know which type of metal loss is majorly affecting the life of the pipeline (Kishawy & Gabbar, 2010). Many attempts were made to predict different types of leakages and faults of a pipeline since decades (Breton et al., 2010; Mandal et al., 2012; Qu et al, 2010; Sun et al., 2014; Vanaei et al., 2017).

The health monitoring systems is crucial to reduce the risk and consequences of failure (Ahsan et al., 2019). Therefore, the reliability analysis of pipeline was carried out with corrosion defects by Teixeira et al. (2008) in which corrosion resulted as one of the major defects which were affecting the pipeline condition. Gloria et al. (2009) introduced a magnetic sensor that predicted the internal defects of corrosion for the inspection process and also had more advantages than MFL The detection of pipeline leakages and pre-warning monitoring using SVM techniques was presented by Qu et al. (2010). Some researchers developed statistical models that could predict the life condition of pipelines using inspection data from oil and gas fields by means of different machine learning techniques like artificial neural networks, regression analysis, Bayesian networks (Basha & Rao, 2018; El-Abbasy et al., 2014; Li et al., 2016; Sacluti et al., 1999). However, of all machine learning techniques, SVM is used in the present work because of its popularity in classifying different types of faults in a pipeline.

An illustration of an oil and gas pipeline undergoing a pig inspection located in a severe environmental condition is shown in Figure 1 (Vanaei et al., 2017). The intelligent pigs are generally used to detect and recognize abnormalities like pressure flow differences, pits, dents, wall thinning, and cracks.

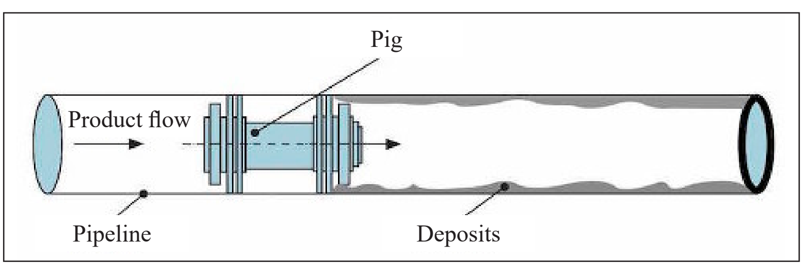

Figure 1. Corroded pipeline in a severe environment

The usage of Support vector machines (SVM) is continuously increasing due to their generalization ability and observed performance for the classification problems. SVM's have a unique architecture that predicts the classification of data in which the hyperplane separates the data of two or more categories. The detailed representation of optimal hyperplane is shown in Figure 2. The vectors that are located closest to the margins of a hyperplane are termed as support vectors. The hyperplane separating two types of metal loss anomalies is termed as optimal separating hyperplane as shown in Figure 2. The generalization control can be attained by maximizing the margin, which corresponds to the minimization of the weight vector (Gunn, 1998). SVM is an innovative learning technique that shows greater performance related to other machine learning techniques, which is derived from the statistical learning concept (Qu et al., 2010). The main contribution of the current work is to classify the types of metal loss anomalies using historical inspection data using SVM technique. The proposed model may help in predicting the type of metal loss defects affecting the life condition of an oil and gas pipeline.

METHODS

Data Acquisition

The historical inspection data was acquired from the crude oil industry located in Sudan. The pipeline section was prolonged to a length of 241.2 km. The pipeline was commissioned in the year 2006 and the inspection of a crude oil pipeline was done in the years 2009 and 2015 (Shaik et al., 2019). The historical data was collected and used in the present study. The recorded metal loss features were identified and selected for the development of the SVM model. Five types of metal loss features were recorded all over the pipeline during the inspection time such as pitting, general, circumferential grooving, axial grooving and circumferential slotting. It was found that the pitting type of metal loss was most reoccurring as compared to other metal loss types.

Since the available data was large, only a small portion of the data set was used for training SVM models to deal with computational memory requirements. To develop the SVM model, the parameters such as length, width, depth, pressure, wall thickness, weld girth are considered as predictors whereas metal loss type is considered as a response. Different types of SVM models were investigated for their accuracies in predicting the fault. Later, the best accurate SVM model type was opted for the prediction of metal loss defects for unseen data. The performance of the selected SVM model was analysed by means of correctness rate in prediction for fault type. The detailed methodology of the selected SVM type is given in the next section.

SVM Approach

SVMs can manage linear, modest, classification tasks as well as complex non-linear classification problems. The basic principle or idea behind SVM is presented in Figure 2, where the hyperplane is separating two classes of data so that the margin is maximized. In other words, the more is the margin, the best will be the classification of data. The formulation or the decision rule that is involved for a point to lie in a hyperplane can be given by the Equation (1) for a data $(x_i,y_i)^n_{i=1}$.

$$w.x_i + b = 0 \quad , \qquad \text{for } y_i = 0 \tag{1}$$

where, w is the weight of vector, x_i is a point and b is the bias.

For an optimization constrained problem, an optimal hyperplane can be attained as a solution by means of Equation (2) when the data sets are linearly separable (Vapnik, 2000).

$$\text{Minimize } \frac{1}{2}||w||^2 \tag{2}$$

Subjected to $y_i\ [w.x_i + b] - 1 \geq 0, i = 1, 2, \cdot \cdot k$

Nevertheless, the acquired data was overlapping, hence a soft constrained condition using a soft variable ε_i was used as given in Equation (3), Yan (2015).

$$y_i\,[w.x_i + b] - 1 + \varepsilon_i \geq 0, i = 1, 2, \cdot \cdot k \tag{3}$$

Where ε_i is soft variable

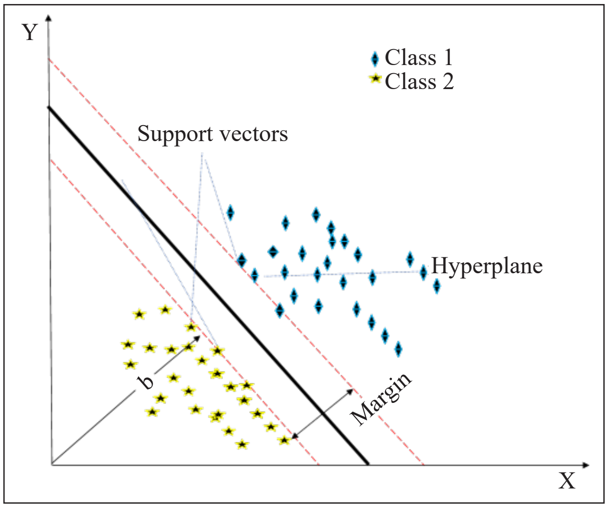

Figure 2. Optimal hyperplane architecture

The structure risk minimization principle was applied to the classification case and the SVM model was trained based on the Equations (4) and Equation (5), Yan (2015).

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}\varepsilon_i \tag{4}$$

$$\text{Subjected to } y_i\,[w.x_i + b] - 1 + \varepsilon_i \geq 0,\ \varepsilon_i \geq 0,\ i = 1, 2, \cdot \cdot k \tag{5}$$

where, C is the disciplinal parameter

Proposed SVM Model Training

The historical inspection data was used to develop SVM models for the prediction of metal loss of different types. Five types of metal loss were identified such as pitting, circumferential grooving, general, circumferential slotting and axial grooving which are affecting the crude oil pipeline in the available data. These metal loss anomalies were given as response while the variables length, width, depth, wall thickness, pressure and weld girth were given as inputs during the training stage for all types of SVM models. The SVM models were trained based on supervised classification principles by means of Equation (4) and Equation (5).

After the data acquisition process, a small portion of five metal loss types data sets were used for training different types of SVM models namely (i) linear SVM, (ii) quadratic SVM, (iii) cubic SVM, (iv) Fine Gaussian SVM, (v) Medium Gaussian SVM, (vi) Coarse Gaussian SVM. Classification toolbox in MATLAB® 2018 for the training of the SVM

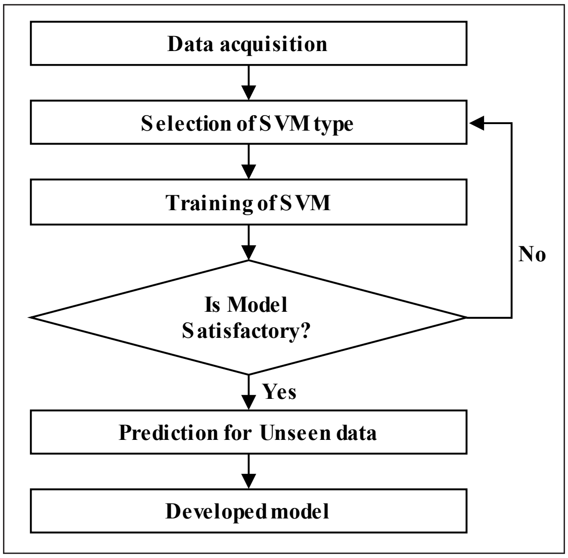

Figure 3. SVM approach

model. Later, the accuracy of all these models was compared. The best accurate SVM type was chosen for the prediction of unseen data. The workflow of the SVM approach is given in Figure 3.

Performance Evaluation Indices

The classification performance of proposed SVM had been evaluated using confusion matrix and various performance indices such as Accuracy, Recall, Precision and F1 Score were defined in Equation (6), Equation (7), Equation (8) and Equation (9), Taqvi et al. (2018). It is noted that the performance evaluation of a developed classification model is a compulsory step in model selection. Once the classification model has developed, the performance of the developed model has been evaluated. It is usually incorporated using predefined performance measure indices. These terms are defined as:

Accuracy: It is a ratio of correctly predicted samples to the total samples

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

Precision: It is the ratio of correctly predicted positive samples to the total predicted positive samples. The high precision from a classifier relates to the low false positive rate.

$$Precision\,(P) = \frac{TP}{TP + FP} \tag{7}$$

Recall: It is also known as sensitivity. It is the ratio of correctly predicted positive samples to all observations in the actual class.

$$Recall\,(R) = \frac{TP}{TP + FN} \tag{8}$$

F1 score: F1 score is the weighted average of precision and recall. Therefore, this score takes both false positives and false negatives into account.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{9}$$

Whereas, TN = True Negative, TP = True Positive, FP = False Positive and FN = False Negative

RESULTS AND DISCUSSION

The crude oil pipeline historical data was used to train all types of SVM models. The models were trained based on supervised classification principles using Equation (4) and Equation (5). The accuracies of all trained SVM models were then calculated and are presented in Table 1. The finest model in terms of accuracy among all models was quadratic SVM, with an accuracy of 87.4%. The proposed Quadratic SVM classified three (03) faults such as general, circumferential slotting and axial grooving most accurate (100% accuracy) among five (05) types of faults. The other two faults such as pitting and circumferential grooving were not found to be so accurate, the accuracies of true and false classification can be seen in the confusion matrix as shown in Figure 4. Notably, it is estimated that some pin locations of a crude oil pipeline were corroded due to more than one type of class, but one type of class was reported in the inspection data. The same SVM model was employed for further predictions on unseen data.

Table 1
Training prediction accuracy for each SVM type

SVM Type	Prediction Accuracy
Linear	84.5%
Quadratic	87.4%
Cubic	78.9%
Fine Gaussian	65.3%
Medium Gaussian	76.8%
Coarse Gaussian	78.5%

True class	Circum. Slotting	Axial Grooving	Circum. Grooving	General	Pitting	TP Rate	FN rate
Circum. Slotting	100%					100%	
Axial Grooving		100%				100%	
Circum. Grooving	5%		84%	5%	5%	84%	16%
General				100%		100%	
Pitting	11%		11%	11%	68%	68%	32%

Predicted class

Figure 4. Confusion Matrix of Quadratic SVM for training data

The trained quadratic SVM model had been applied to unseen data for the prediction and results found to be satisfactory. The trained SVM model was tested using a new dataset that has not been used in the training stage and allowed for the prediction of fault. Later, the prediction was made for the data irrespective of fault type, and results were found to be 93.0% accurate. Among these predictions, pitting fault type was least accurate which may be due to the occurrence of more than one fault type at the same pin location of a pipeline. The accuracies of all predictions of faulty type for unseen data are presented in Figure 5.

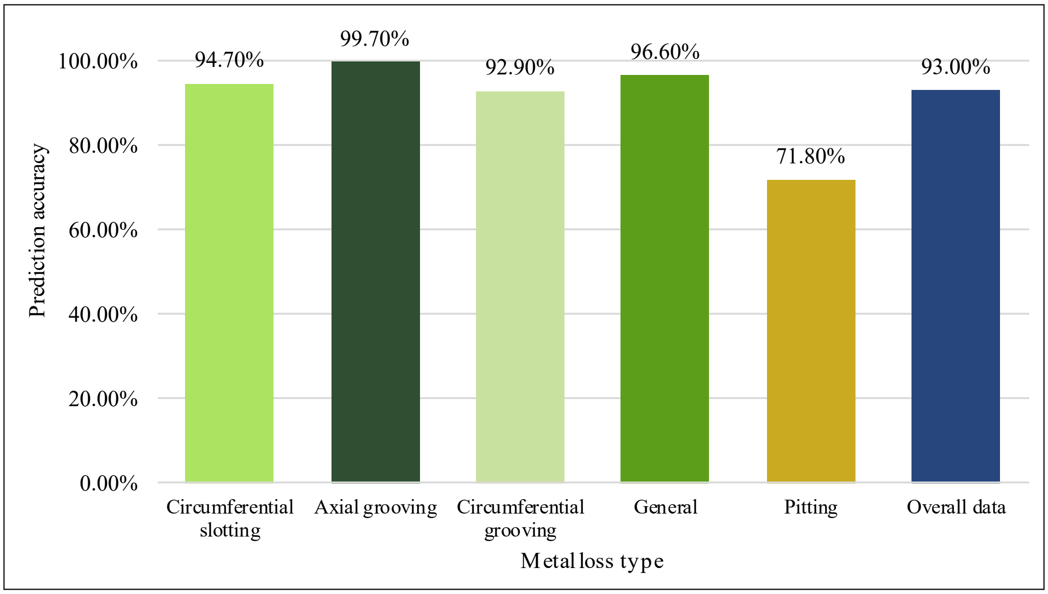

Figure 5. Accuracy predictions of metal loss type

Performance Evaluation

The classification performance of the developed SVM model had been evaluated using a confusion matrix as shown in Table 2. It can be observed that SVM had correctly classified axial grooving fault with the accuracy and F1 score of 100%. Similar results were obtained for type general, circumferential slotting and circumferential grooving fault with accuracies of 98%, 97.88% and 90.62%, F1 scores of 98.04%, 97.92% and 89.96 respectively. A very low classification accuracy can be observed in pitting fault as it has been 33% misclassified in other fault classes. As a result, the fault detection rate (FDR) for all faults was found to be 90.4%, while the misclassification rate (MR) was 9.6%. These results can also be compared with the previous study by (Taqvi et al., 2018) in which distillation column fault were classified using SVM. The FDR was found to 99.7% which is comparable with the current study. The high FDR is the previous study is due to the large number of faults used for the training of classifier.

Table 2
Classification performance

Type Of Fault	Accuracy	Recall	Precision	F1 Score
Circumferential Slotting	97.88	100	95.90	97.92
Axial Grooving	100	100	100	100
Circumferential Grooving	90.62	84	96.8	89.96
General	98.00	100	96.2	98.04
Pitting	83.37	68	98.2	80.35

Pareto Analysis

The data had been analysed by utilizing Pareto analysis to observe the most affecting fault type using historical data. The historical data that contained a type of metal loss was counted and arranged in descending order such that the most occuring type of metal loss on top. The cummulative percentage of fault type had been calculated and presented in Table 3. Figure 6 shows that pitting fault type has the highest impact according to the 80/20 rule as denoted by the dotted line. The Pareto analysis shows that the pitting fault type was majorly occuring when compared to other faults (metal loss) types. Therefore, it can be concluded that the pipeline under consideration in this study is most affected by pitting type of metal loss. The pitting fault type is highly affecting the pipeline and causing major deterioration due to harsh environmental conditions around the pipeline. Vanaei et al. (2017) described pitting corrosion to be the most observed metal loss in the transmission pipeline causing severe, localized deterioration to the surface area of the pipeline (Vanaei et al., 2017). Pitting corrosion can be caused due to several reasons that include: pipe material defects, penetration by chemicals, damage protective passive film or improper material used in the

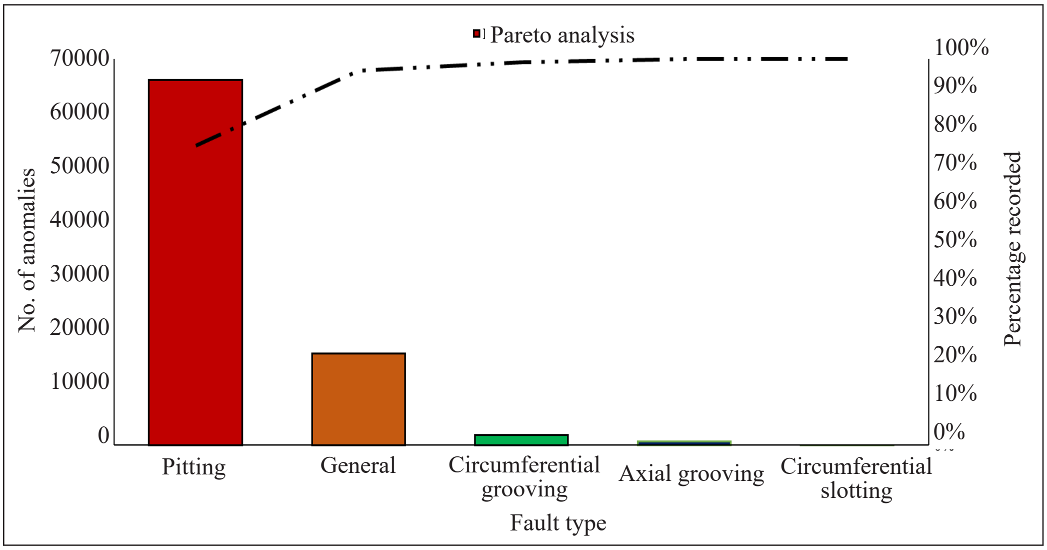

Figure 6. Pareto analysis

pipeline. To overcome this metal loss type, necessary actions like repair, a cathodic coating is done for better reliability of pipeline and smooth operation. Failure analysis for pitting corrosion for pipelines installed in Iran is done by Mansoori et al. (2017) which founds pitting mechanism to be the cause of failure in the pipeline. Necessary actions like repair, cathodic coating need to be done to avoid these issues for better reliability of pipelines and smooth operation.

Table 3
Contribution effect of individual type

Type of fault	Contribution (%)
Pitting	77.47869
Circumferential Grooving	2.197173
General	19.49041
Axial Grooving	0.810346
Circumferential Slotting	0.022217

CONCLUSION

This study presented an SVM based approach to classify various faults in crude oil pipeline based on historical inspection data. The data was utilized to develop an accurate SVM model for the classification purpose. The accuracy rates of all SVM models such as Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM and Coarse Gaussian SVM were calculated and presented. The quadratic SVM type was found to be more appropriate for this available data in terms of classifying the metal loss type and was selected for further predictions. The results were found to be satisfactory with an overall prediction rate of 93.0%. It was found that the data used for prediction was not classified as two faults at a single point which might be the cause of deviation in predictions. Furthermore, it was revealed that the pipeline under consideration was mostly affected by pitting type of metal loss with 77.47 % share in total recorded failures. The proposed approach may help the oil and gas or other pipeline industries in condition assessment and maintenance schedule, which can decrease the product losses, increase the life span of a pipeline and minimize the risk associated with failure.

ACKNOWLEDGEMENT

This work was supported by the YUTP grant with the cost centre 0153AA-E56.

REFERENCES

Ahsan, S., Lemma, T. A., & Gebremariam, M. A. (2019). Reliability analysis of gas turbine engine by means of bathtub-shaped failure rate distribution. *Process Safety Progress, 39*(S1), 1-10.

Basha, S. N., & Rao, P. S. (2018, September 18-19). *A simulated model for assesing the line condition of onshore pipelines.* In *UTP-UMP-VIT Symposium on Energy Systems 2018* (pp. 1-5). Pekan, Pahang, Malaysia.

Breton, T., Sanchez-Gheno, J., Alamilla, J., & Alvarez-Ramirez, J. (2010). Identification of failure type in corroded pipelines: A Bayesian probabilistic approach. *Journal of Hazardous Materials, 179*(1-3), 628-634.

Cosham, A., Hopkins, P., & Macdonald, K. (2007). Best practice for the assessment of defects in pipelines–Corrosion. *Engineering Failure Analysis, 14*(7), 1245-1265.

El-Abbasy, M. S., Senouci, A., Zayed, T., Mirahadi, F., & Parvizsedghy, L. (2014). Artificial neural network models for predicting condition of offshore oil and gas pipelines. *Automation in Construction, 45*, 50-65.

Gloria, N., Areiza, M., Miranda, I., & Rebello, J. (2009). Development of a magnetic sensor for detection and sizing of internal pipeline corrosion defects. *NDT & E International, 42*(8), 669-677.

Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS Technical Report, 14*(1), 5-16.

Isa, D., & Rajkumar, R. (2009). Pipeline defect prediction using support vector machines. *Journal of Applied Artificial Intelligence, 23*(8), 758-771.

Kishawy, H. A., & Gabbar, H.A. (2010). Review of pipeline integrity management practices. *International Journal of Pressure Vessels and Piping, 87*(7), 373-380.

Li, X., Chen, G., & Zhu, H. (2016). Quantitative risk analysis on leakage failure of submarine oil and gas pipelines using Bayesian network. *Process Safety and Environmental Protection, 103*, 163-173.

Mandal, S. K., Chan, F. T., & Tiwari, M. (2012). Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM. *Expert Systems with Applications, 39*(3), 3071-3080.

Mansoori, H., Mirzaee, R., Esmaeilzadeh, F., Vojood, A., & Dowrani, A. S. (2017). Pitting corrosion failure analysis of a wet gas pipeline. *Engineering Failure Analysis, 82*, 16-25.

Nešić, S. (2007). Key issues related to modelling of internal corrosion of oil and gas pipelines–A review. *Corrosion Science, 49*(12), 4308-4338.

Qu, Z., Feng, H., Zeng, Z., Zhuge, J., & Jin, S. (2010). A SVM-based pipeline leakage detection and pre-warning system. *Measurement, 43*(4), 513-519.

Sacluti, F., Stanley, S., & Zhang, Q. (1999, October 18-20). *Use of artificial neural networks to predict water distribution pipe breaks.* In *Proceedings of the 51st Annual Conference of the Western Canada Water and Wastewater Association* (p. 12). Saskatoon, Canada.

Shaik, N. B., Pedapati, S. R., & Dzubir, F. A. A. (2019). Remaining useful life prediction of crude oil pipeline by means of deterioration curves. *Process Safety Progress, 39*(S1), 1-6.

Sun, J., Xiao, Q., Wen, J., & Wang, F. (2014). Natural gas pipeline small leakage feature extraction and recognition based on LMD envelope spectrum entropy and SVM. *Measurement, 55*, 434-443.

Taqvi, S. A., Tufa, L. D., Zabiri, H., Maulud, A. S., & Uddin, F. (2018). Multiple fault diagnosis in distillation column using multikernel support vector machine. *Industrial & Engineering Chemistry Research, 57*(43), 14689-14706.

Teixeira, A. P., Soares, C. G., Netto, T. A., & Estefen, S. F. (2008). Reliability of pipelines with corrosion defects. *International Journal of Pressure Vessels and Piping, 85*(4), 228-237.

Vanaei, H. R., Eslami, A., & Egbewande, A. (2017). A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *International Journal of Pressure Vessels and Piping, 149*, 43-54.

Vapnik, V. (2000). *The nature of statistical learning theory* (2nd Ed). New York, USA: Springer-Verlag New York.

Yan, J. (2015). *Machinery prognostics and prognosis oriented maintenance management*. Singapore: John Wiley & Sons.