Decision Tree-Based Approach for Defect Detection and Classification in Oil and Gas Pipelines

Abduljalil Mohamed[1](✉), Mohamed Salah Hamdi[1], and Sofiene Tahar[2]

[1] Information Systems Department, Ahmed Bin Mohamed Military College, Doha, Qatar
{ajmaoham, mshamdi}@abmmc.edu.qa
[2] Electrical and Computer Engineering Department, Concordia University, Montreal, Canada
tahar@ece.concordia.ca

Abstract. Metallic pipelines are used to transfer crude oil and natural gas. These pipelines extend for hundreds of kilometers, and as such, they are very vulnerable to physical defects such as dents, cracks, corrosion, etc. These defects may lead to catastrophic consequences if not managed properly. Thus, monitoring these pipelines is an important step in the maintenance process to keep them up and running. During the monitoring stage, two critical tasks are carried out: defect detection and defect classification. The first task concerns with the determination of the occurrence of a defect in the monitored pipeline. The second task concerns with classifying the detected defect as a serious or tolerable defect. In order to accomplish these tasks, maintenance engineers utilize Magnetic Flux Leakage (MFL) data obtained from a large number of magnetic sensors. However, the complexity and amount of MFL data make the detection and classification of pipelines defects a difficult task. In this study, we propose a decision tree–based approach as a viable monitoring tool for the oil and gas pipelines.

Keywords: Defect detection and classification · Decision tree
Data mining · Pipeline monitoring and maintenance

1 Introduction

Oil and gas pipeline defect monitoring is an essential component of the pipeline maintenance process. In order to maintain the pipeline in a properly working order, different inspection tools such as magnetic flux leakage (MFL), ultrasonic waves, and closed circuit television (CCTV) are used to detect and classify pipeline defects [1–3]. The complexity and amount of data obtained by such diverse tools require the use of sophisticated defect detection and classification techniques. Most of the approaches reported in the literature [4] have been proposed for the purpose of either prediction of defect dimensions, detection of defects, or classification of defect types. To achieve

K. Arai et al. (Eds.): FTC 2018, AISC 880, pp. 490–504, 2019.
https://doi.org/10.1007/978-3-030-02686-8_37

these objectives, techniques such as machine learning [5–7], wavelets [8–13], and signal processing [14–16] are widely used.

The focus of this paper, however, is on developing a pipeline monitoring tool that incorporates the two tasks namely: defect detection and defect classification. The main inference engine for both tasks is a decision tree that takes as an input the crucial MFL depth and length parameters.

2 Pipeline Monitoring

In this paper, we propose a new monitoring approach for oil and gas pipelines. The general structure of the proposed approach is shown in Fig. 1.

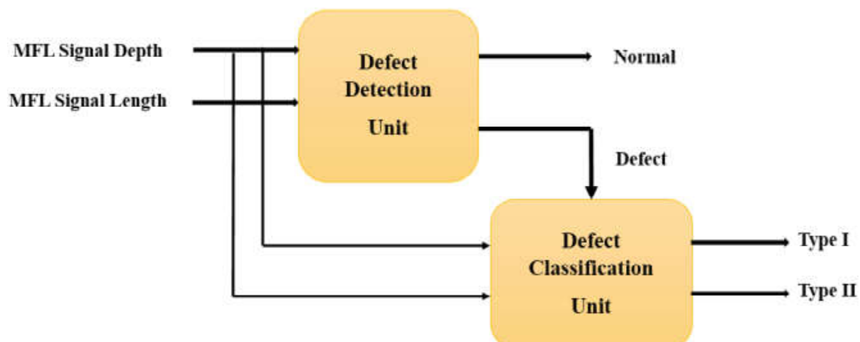


Fig. 1. The proposed monitoring approach for the oil and gas pipelines.

MFL Signals. MFL data are collected from autonomous devices known as intelligent pigs. An increase in flux leakage may indicate metal loss, which in turn, means the possibility of defect occurrence. Thus, at the location of the potential defect, the depth and length of the flux leakage are measured or estimated by using artificial neural networks.

Defect Detection. These two crucial MFL parameters are first entered into the defect detection unit. A decision tree is realized in this unit as defect detection technique. If no defect is detected, the monitoring process terminates. On the other hand, if a pipeline defect is detected, the two parameters will be passed on to the classification unit.

Defect Classification. In this unit, based on their severity level, the defect is classified into one of two categorizes: Type I or Type II. In this work, Type I is considered a very serious pipeline defect which requires an immediate action and reparation. Type II is considered less serious and can wait and be scheduled for defect maintenance.

3 Decision Tree-Based Approach for Defect Detection and Classification

The decision tree utilized in this work is derived from the simple divide-and-conquer algorithm. The decision tree is expressed recursively as described in the following sections.

MFL Signal Depth and Length Attributes. In order to detect/classify pipeline defects, the obtained MFL signals are first normalized and mapped into depth and length ranges. According to the industry standard [17], the depth range for the MFL signals is normalized between 0 and 1; and the length range for the MFL signals is normalized between 0 and 6. These two ranges constitute the MFL attributes, and are divided into different values as described below.

The MFL depth attribute values are:
Very high = [0.80 1.00],
High = [0.60 0.79],
Medium = [0.40 0.59],
Low = [0.20 0.39],
Very low = [0.00 0.19],
The MFL length attribute values are:
Large = [3.81 6.00],
Medium = [1.81 3.80],
Small = [0.61 1.80],
Very small = [0.00 0.60],

Defect Detection. Based on the information given in [17], the MFL attributes can now be used to identify the status of the MFL signals as shown in Table 1. The MFL signal can either be identified as abnormal (defect) or normal.

Constructing Decision Tree. To construct the decision tree for the defect detection, an attribute is first selected and placed at the root node, and make branch for each possible value. This splits up the MFL signals into subsets, one for every value of the attribute. The process is repeated recursively for each branch, using only those instances that actually reach the branch. If all instances at a particular node are all either abnormal or normal, then we stop developing that part of the tree. There are two possibilities for each split; and they produce two trees as shown in Figs. 2 and 3 for the depth and length attributes, respectively.

The number of 2 (abnormal) and 1 (normal) classes is shown at the leaves. Any leaf with only one class (i.e., 2 or 1) reaches the final split; and thus the recursive process terminates. In order to reduce the size of the trees, the information gain for each node is measured. Now the information for the two attributes is calculated and split is made on the one that gains the most information.

Tree Structure. The informational value of creating a branch on the MFL-depth attribute and the MFL-length attribute are then calculated. The number of normal and abnormal at the leaf nodes in Fig. 2 are [0 4], [1 3], [2 2], [2 2], and [4 0], respectively.

Table 1. MFL signal abnormal and normal status based on its depth and length range

MFL-depth					MFL-length				Status	
Very high	High	Medium	Low	Very low	Very small	Small	Medium	Large	Normal (1)	Abnormal (2)
YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES
YES	NO	NO	NO	NO	NO	YES	NO	NO	NO	YES
YES	NO	NO	NO	NO	NO	NO	YES	NO	NO	YES
YES	NO	NO	NO	NO	NO	NO	NO	YES	NO	YES
NO	YES	NO	NO	NO	YES	NO	NO	NO	YES	NO
NO	YES	NO	NO	NO	NO	YES	NO	NO	NO	YES
NO	YES	NO	NO	NO	NO	NO	YES	NO	NO	YES
NO	YES	NO	NO	NO	NO	NO	NO	YES	NO	YES
NO	NO	YES	NO	NO	YES	NO	NO	NO	YES	NO
NO	NO	YES	NO	NO	NO	YES	NO	NO	YES	NO
NO	NO	YES	NO	NO	NO	NO	YES	NO	NO	YES
NO	NO	YES	NO	NO	NO	NO	NO	YES	NO	YES
NO	NO	NO	YES	NO	YES	NO	NO	NO	YES	NO
NO	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO
NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	YES
NO	NO	NO	YES	NO	NO	NO	NO	YES	NO	YES
NO	NO	NO	NO	YES	YES	NO	NO	NO	YES	NO
NO	NO	NO	NO	YES	NO	YES	NO	NO	YES	NO
NO	NO	NO	NO	YES	NO	NO	YES	NO	YES	NO
NO	NO	NO	NO	YES	NO	NO	NO	YES	YES	NO

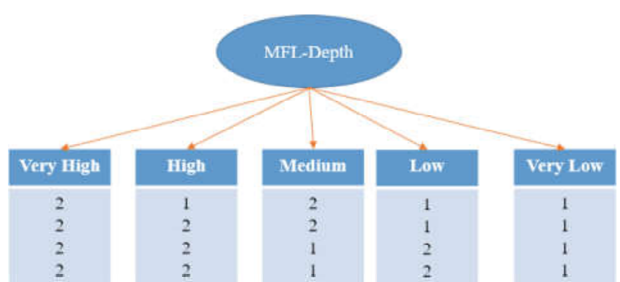


Fig. 2. The decision tree for the MFL depth attribute. The abnormal status is referred to by 2; while the normal status is referred to by 1.

The number of normal and abnormal at the leaf nodes in Fig. 3 are [4 1], [3 2], [1 4], and [1 4], respectively.

Calculating the information gain for each attribute yields the tree structure shown in Fig. 4. As described in Fig. 5, the decision tree basically uses three values of the MFL-depth attribute and four values of the MFL-length attribute. The values are Low, Medium, and High for the MFL-depth attribute, and Very Small, Small, Medium, and Large for the MFL-length attribute.

Fig. 3. The decision tree for the MFL length attribute. The abnormal status is referred to by 2; while the normal status is referred to by 1.

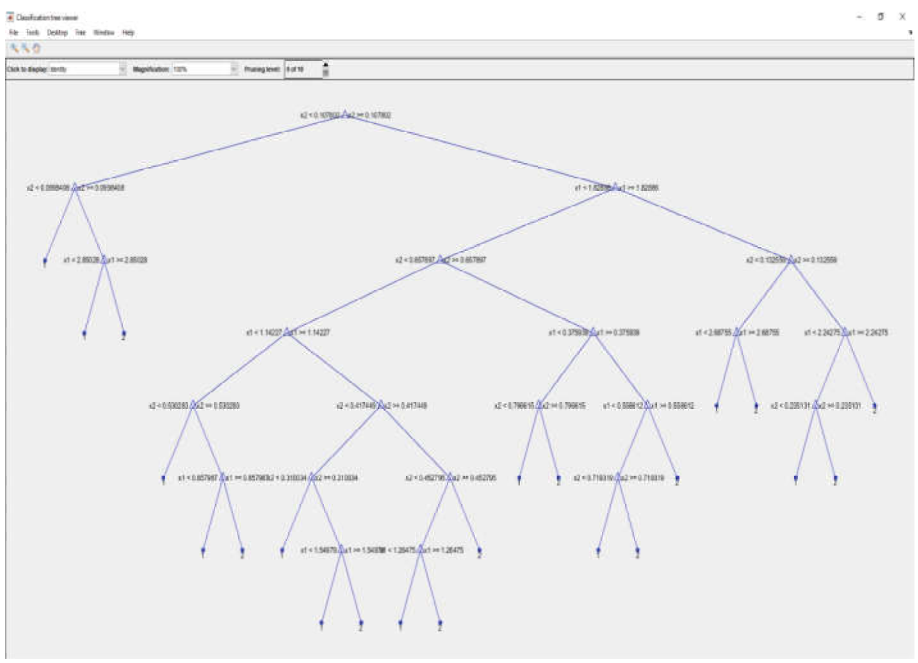


Fig. 4. The decision tree structure for the defect detection.

Defect Classification. The MFL data used for classifying the defect severity level is shown in Table 2. The table shows that the two attribute values can indicate either the defect level is of Type I, or the defect level is of Type II.

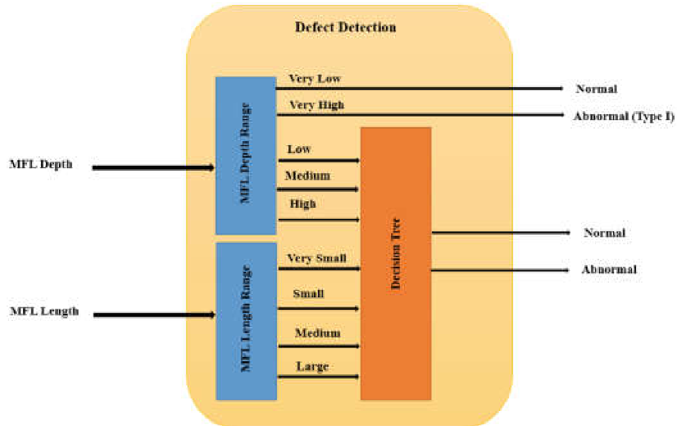


Fig. 5. The defect detection based on the two MFL attributes.

Table 2. MFL signal defect (i.e., Type I, Type II) status based on its depth and length range

MFL-depth			MFL-length			Defect	
High	Medium	Low	Small	Medium	Large	Type I (1)	Type II (2)
YES	NO	NO	YES	NO	NO	NO	YES
YES	NO	NO	NO	YES	NO	YES	NO
YES	NO	NO	NO	NO	YES	YES	NO
NO	YES	NO	YES	NO	NO	NO	YES
NO	YES	NO	NO	YES	NO	NO	YES
NO	YES	NO	NO	NO	YES	YES	NO
NO	NO	YES	YES	NO	NO	NO	YES
NO	NO	YES	NO	YES	NO	NO	YES
NO	NO	YES	NO	NO	YES	NO	YES

Constructing Decision Tree. The two trees produced by the two attributes are shown in Figs. 6 and 7. As was the case for the defect decision tree, the information gain for each node is measured, and split is made on the one that gains the most information.

Tree Structure. The informational value of creating a branch on the MFL-depth attribute and the MFL-length attribute are then calculated. The number of defect Type I and Type II at the leaf nodes in Fig. 6 are [2 1], [1 2], and [0 3], respectively. The number of defect type I and type II at the leaf nodes in Fig. 7 are [0 3], [1 2], and [2 1], respectively.

Calculating the information gain for each attribute yields the tree structure shown in Fig. 8. As described in Fig. 9, the decision tree basically uses three values of the MFL-depth attribute and three values of the MFL-length attribute. The values are Low, Medium, and High for the MFL-depth attribute, and Small, Medium, and Large for the MFL-length attribute.

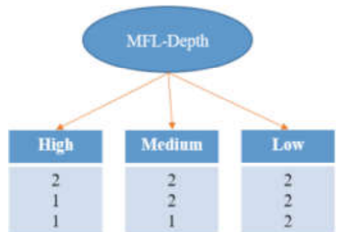


Fig. 6. The decision tree for the MFL-depth attribute. The defect status of type I is referred to by 1; while type II is referred to by 2.

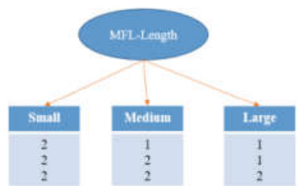


Fig. 7. The decision tree for the MFL-length attribute. The defect status of type I is referred to by 1; while type II is referred to by 2.

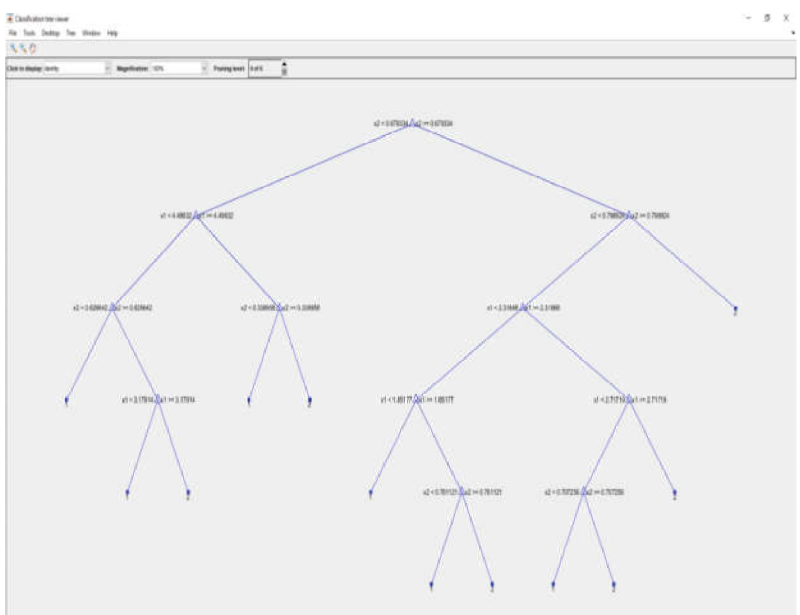


Fig. 8. The decision tree structure for the defect classification.

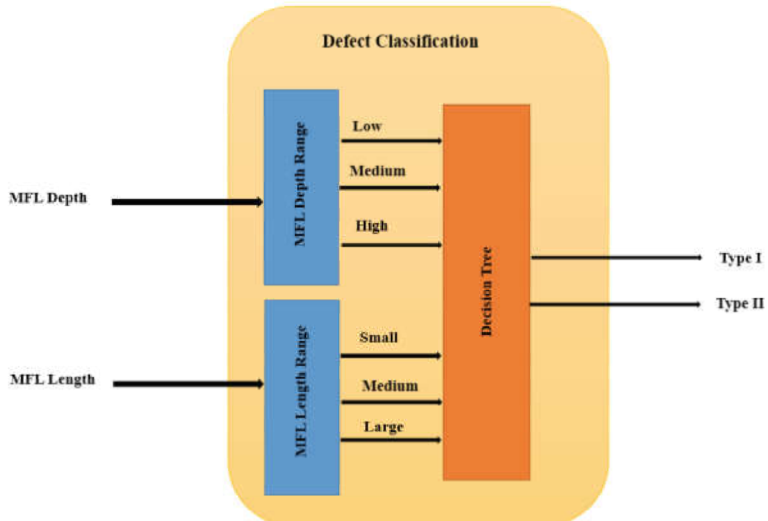


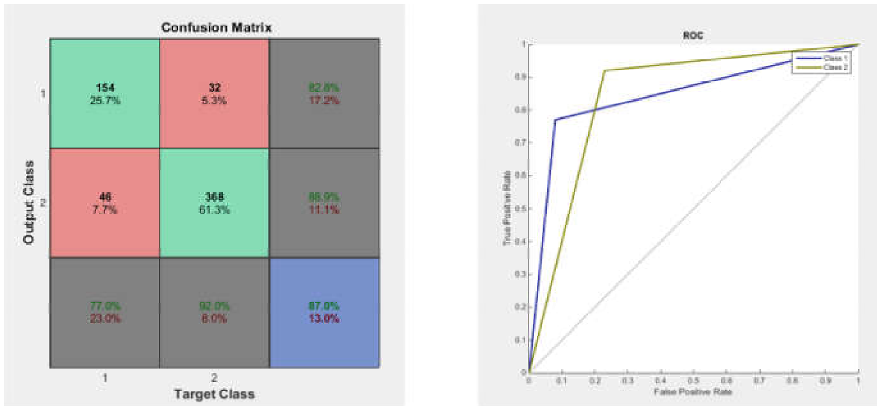
Fig. 9. The defect classification based on the two MFL attributes.

4 Performance Evaluation

The performance of the proposed approach is measured by two important criteria: the receiver operating characteristics (ROC) curves and the confusion matrices. In ROC, the true positive rates (sensitivity) are plotted against the false positive rates (1-specificity) for different cut-off points. For a specific severity class, the closer its ROC curve is to the left upper corner of the graph, the higher its classification accuracy is. In the confusion matrix plot, the rows correspond to the predicted class (output class), and the columns show the true class (target class). In the defect detection and classification, the proposed approach is compared with the four well-known classifiers, namely the Naive Bayesian (NB) classifier, k-nearest neighbor (KNN) classifier, Artificial Neural Network (ANN) classifier, and the Support Vector Machine (SVM) classifier.

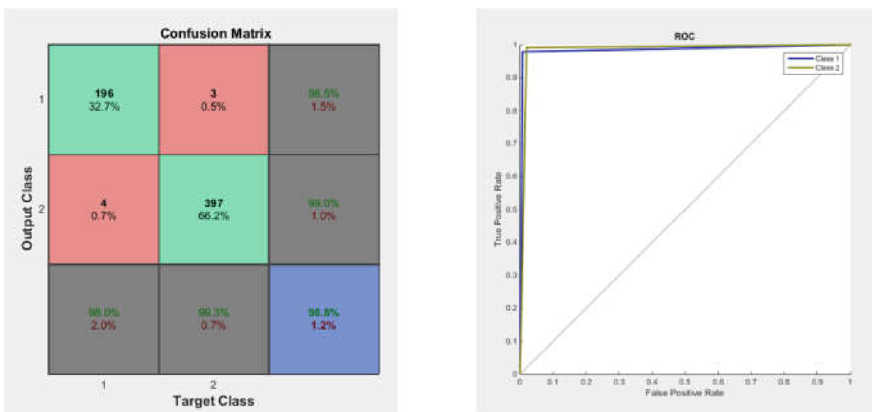
Data. The available MFL dataset used in the experimental work is categorized as follows. For the defect detection, there are 907 samples of normal status, and 2721 samples of the abnormal status. For the defect classification, there are 907 samples for each type of defects. The data samples have been further divided as follows: 70% for training, 15% for validation, and 15% for testing.

Defect Detection. The confusion matrix and the ROC curves for each detector model are shown in Figs. 10, 11, 12, 13 and 14 for the models NB, KNN, ANN, SVM, and the proposed decision tree (DT). In these figures, the normal status of the MFL signal is referred to by Class 1, and abnormal status is referred to by Class 2.



(a) (b)

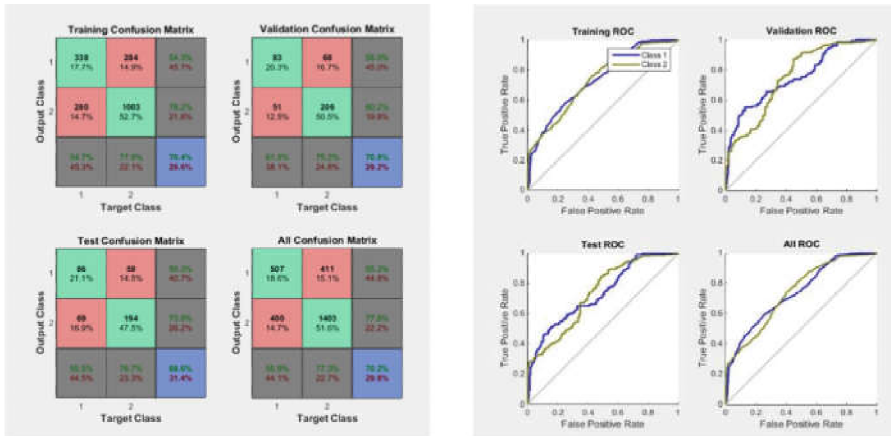
Fig. 10. The defect detection confusion matrix (a) and ROC curves (b) for the NB model.



(a) (b)

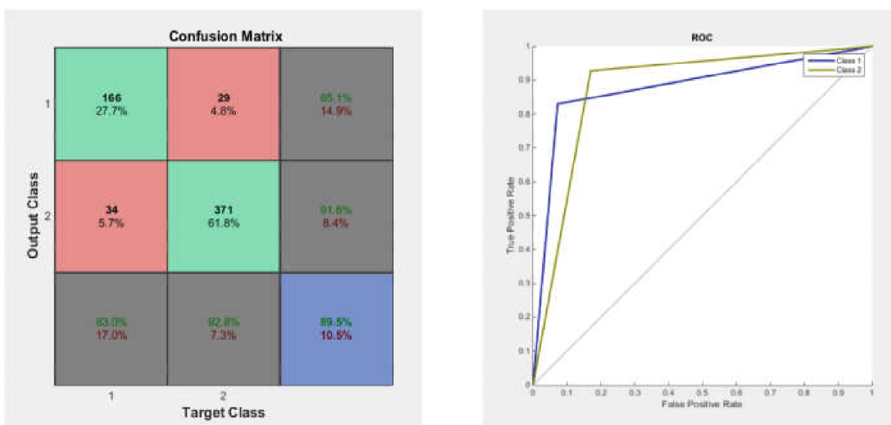
Fig. 11. The defect detection confusion matrix (a) and ROC curves (b) for the KNN model.

Defect Classification. The confusion matrix and the ROC curves for each classifier model are shown in Figs. 15, 16, 17, 18 and 19 for the models NB, KNN, ANN, SVM, and the proposed decision tree (DT). In these figures, the defect type is referred to by Class 1, and defect Type II is referred to by Class 2.



(a) (b)

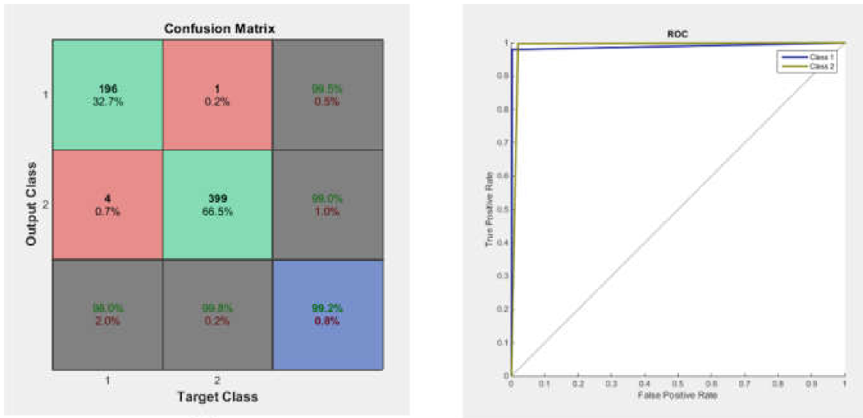
Fig. 12. The defect detection confusion matrix (a) and ROC curves (b) for the ANN model.



(a) (b)

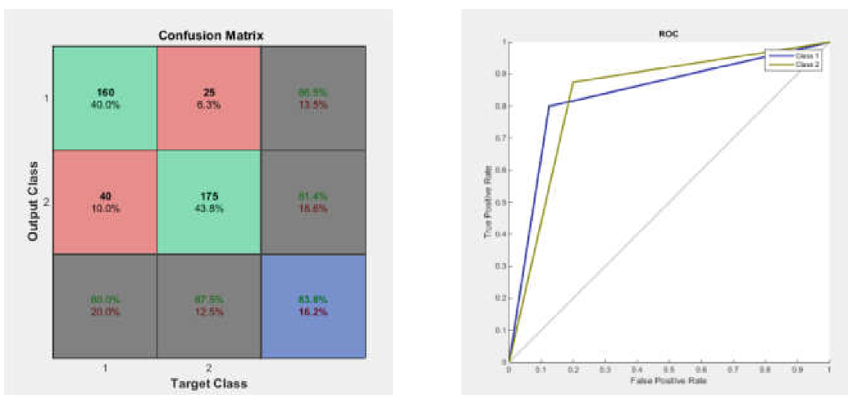
Fig. 13. The defect detection confusion matrix (a) and ROC curves (b) for the SVM model.

It should be noted from these figures that the proposed DT model outperforms all other models. It yields 99.2% accuracy for the detection and classification. Moreover, the artificial neural network model yields the worst performance at 70.2% detection accuracy and 71.4% classification accuracy. The defect detection and classification performance of all models are summarized in Table 3.



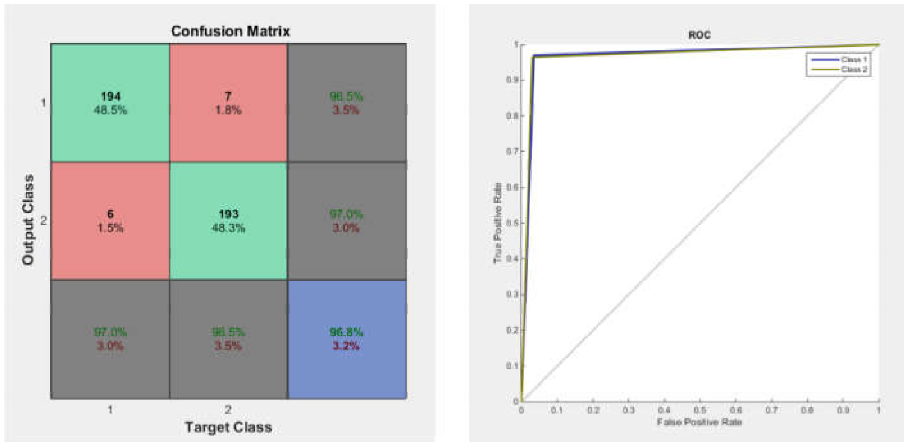
(a) (b)

Fig. 14. The defect detection confusion matrix (a) and ROC curves (b) for the DT model.



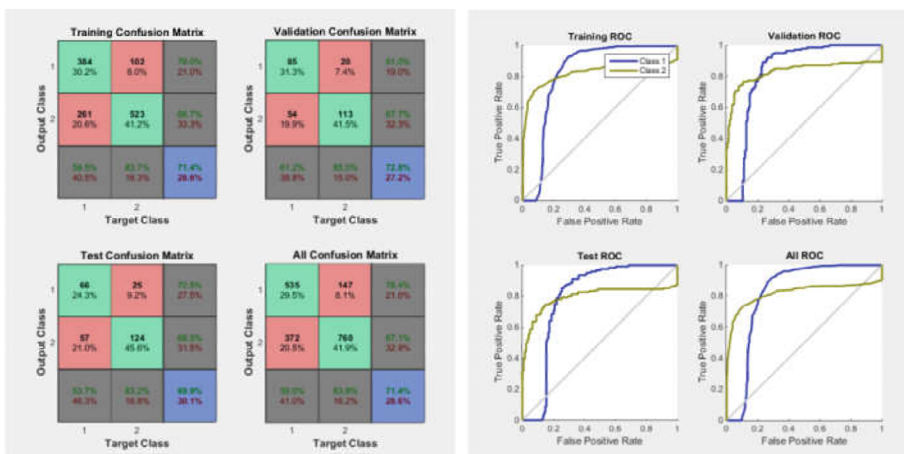
(a) (b)

Fig. 15. The defect classification confusion matrix (a) and ROC curves (b) for the NB model.



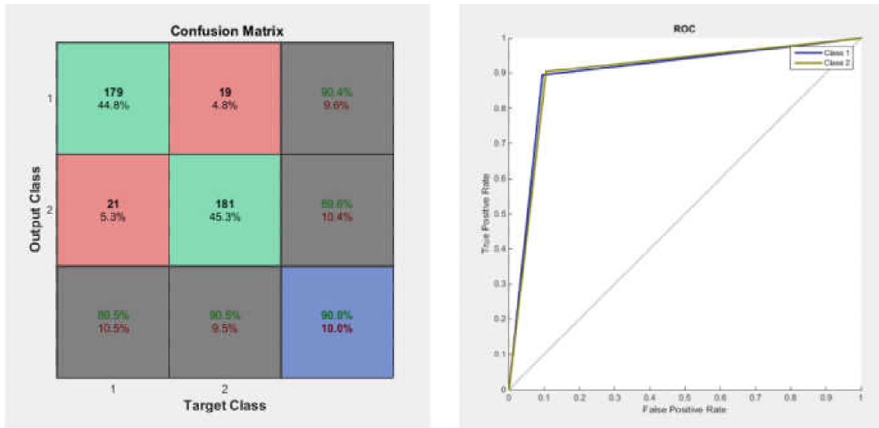
(a) (b)

Fig. 16. The confusion matrix (a) and ROC curves (b) for the KNN model.



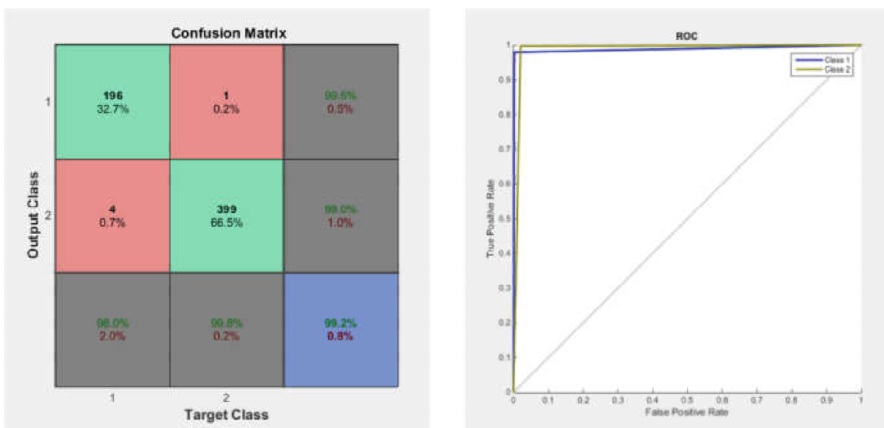
(a) (b)

Fig. 17. The defect classification confusion matrix (a) and ROC curves (b) for the ANN model.



(a) (b)

Fig. 18. The defect classification confusion matrix (a) and ROC curves (b) for the SVM model.



(a) (b)

Fig. 19. The defect classification confusion matrix (a) and ROC curves (b) for the DT model.

Table 3. Detection and classification accuracy for the NB, KNN, ANN, SVM, and DT models.

Classifier model	Defect	
	Detection	Classification
NB	87%	83.8%
KNN	98.8%	96.8%
ANN	70.2	71.4%
SVM	89.5%	90%
DT	99.2%	99.2%

5 Conclusion

The monitoring process for the oil and gas pipelines consists of two main tasks: defect detection and defect classification. The complexity and amount of the MFL monitoring data make both tasks very difficult. In this work, we proposed a decision tree-based approach as a viable monitoring tool. The new approach is evaluated using two important criteria: the receiver operating characteristics (ROC) curves and the confusion matrices. The performance of the new approach is compared with other well-known monitoring tools. Extensive experimental work has been carried out and the performance of the proposed approach along with four other well-known techniques are reported. The new approach outperforms all of them with accuracy at 99.2% for the detection and classification tasks.

Acknowledgment. This work was made possible by NPRP Grant # [5-813-1-134] from Qatar Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

References

1. Park, G.S., Park, E.S.: Improvement of the sensor system in magnetic flux leakage-type nod-destructive testing. IEEE Trans. Magn. **38**(2), 1277–1280 (2002)
2. Jiao, J., et al.: Application of ultrasonic guided waves in pipe's NDT. J. Exp. Mech. **1**, 000 (2002)
3. Jiao, J., et al.: Application of ultrasonic guided waves in pipe's NDT. J. Exp. Mech. **17**(1), 1–9 (2002)
4. Layouni, M, Tahar, S., Hamdi, M.S.: A survey on the application of neural networks in the safety assessment oil and gas pipelines. In: 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions. IEEE (2014)
5. Khodayari-Rostamabad, A., et al.: Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection. IEEE Trans. Magn. **45**(8), 3073–3084 (2009)
6. Lijian, Y., et al.: Oil-gas pipeline magnetic flux leakage testing defect reconstruction based on support vector machine. In: Second International Conference on Intelligent Computation Technology and Automation, ICICTA 2009, vol. 2. IEEE (2009)
7. Vidal-Calleja, T., et al.: Automatic detection and verification of pipeline construction features with multi-modal data. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014). IEEE (2014)
8. Song, S., Que, P.: Wavelet based noise suppression technique and its application to ultrasonic flaw detection. Ultrasonics **44**(2), 188–193 (2006)
9. Hwang, K., et al.: Characterization of gas pipeline inspection signals using wavelet basis function neural networks. NDT E Int. **33**(8), 531–545 (2000)
10. Mukhopadhyay, S., Srivastava, G.P.: Characterisation of metal loss defects from magnetic flux leakage signals with discrete wavelet transform. NDT E Int. **33**(1), 57–65 (2000)
11. Han, W., Que, P.: A modified wavelet transform domain adaptive FIR filtering algorithm for removing the SPN in the MFL data. Measurement **39**(7), 621–627 (2006)
12. Joshi, A., et al.: Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection. IEEE Trans. Magn. **42**(10), 3168–3170 (2006)

13. Qi, S., Liu, J., Jia, G.: Study of submarine pipeline corrosion based on ultrasonic detection and wavelet analysis. In: 2010 International Conference on Computer Application and System Modeling (ICCASM), vol. 12. IEEE (2010)
14. Afzal, M., Udpa, S.: Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. NDT E Int. **35**(7), 449–457 (2002)
15. Guoguang, Z., Penghui, L.: Signal processing technology of circumferential magnetic flux leakage inspection in pipeline. In: 2011 Third International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), vol. 3. IEEE (2011)
16. Kandroodi, M.R., et al.: Defect detection and width estimation in natural gas pipelines using MFL signals. In: 2013 9th Asian Control Conference (ASCC). IEEE (2013)
17. Cosham, A., Hopkins, P., Macdonald, K.A.: Best practice for the assessment of defects in pipelines—corrosion. Eng. Fail. Anal. **14**(7), 1245–1265 (2007)