



# Failure risk analysis of pipelines using data-driven machine learning algorithms

Ram K. Mazumder <sup>a</sup>, Abdullahi M. Salman <sup>b</sup>, Yue Li <sup>a,\*</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland, OH, United States

<sup>b</sup> Department of Civil and Environmental Engineering, The University of Alabama in Huntsville, United States

## ABSTRACT

Failure risk analysis of pipeline networks is essential for their effective management. Since pipeline networks are often large and complex, analyzing a large number of pipelines is often intensive and time-consuming. This study utilizes recent advances in machine learning to develop a viable alternative to computationally intensive analytical approaches to determine the failure risk of steel oil and gas pipelines. Burst failure risk of pipelines under active corrosion defects is evaluated considering the remaining strength of pipelines with corrosion pit. In the first part of the study, a comprehensive database of pipelines is generated based on information extracted from literature, and true failure pressure is predicted considering variability between predicted burst failure pressure and experimental burst test results. True burst failure pressure is estimated using various design codes in a probabilistic manner. Among the various burst failure prediction models, the DNV RP-F101 model provides the lowest variability in the failure pressure prediction. Therefore, the model is used to generate the probability of failure of pipelines at a burst limit state. Pipelines are classified based on their probability of failures, from low to severe failure risk groups. Then, eight machine learning algorithms are evaluated based on the generated dataset to identify the best failure prediction model. The performance of machine learning algorithms is evaluated on the basis of a confusion matrix and computational efficiency. It shows that XGBoost is the optimal algorithm in predicting the failure and is recommended for future analysis. The computational efficiency of the machine learning algorithm is also compared to the physics-based model. It is revealed that machine learning algorithms can perform a failure risk analysis of pipelines with greater computational efficiency than the physics-based approach. Even the slowest machine learning algorithm is about 12 times faster than the physics-based model.

## 1. Introduction

Pipelines are the most reliable and the cheapest mode of transportation systems for oil and gas. Maintaining pipeline integrity is of the utmost importance to avoid catastrophic burst failure consequences. Unfortunately, aging pipelines have experienced a high frequency of breaks in recent years [52]. The failure in aging gas and oil pipelines are mostly triggered by strength reduction due to corrosion deterioration. Corrosion has been recognized as the primary threat to pipeline performance and safety [12,25,62]. The reduction of material strength due to corrosion pit growth on the pipe wall increases the possibility of leakage and burst failures [30]. Burst failure is the most significant and often causes severe economic, environmental, and social consequences. Burst failure can be expressed by the ultimate failure pressure of a pipe at plastic collapse that illustrates the pipe's ultimate load-bearing capacity under internal pressure [25]. A number of methods have been developed in the past to estimate the burst failure pressure of pipelines subjected to corrosion defects (e.g., [7,21,33,8,46,55]). These methods mostly used semi-empirical fracture mechanics relationships where failure depends on flow stress [25]. Hence, a pipeline's failure

probability can be determined by comparing the true failure pressure subjected to corrosion deterioration and the maximum operating fluid pressure inside a pipe [20].

Accurate estimation of the failure likelihood is a challenging task as there are large uncertainties involved in the parameters associated with the burst failure prediction models. The application of physical probabilistic approach is computationally intensive given that oil and gas pipeline networks are typically large and complex, and composed of hundreds of pipeline segments. A detailed evaluation of the reliability analysis of each pipeline in a complex oil and gas transmission network is often not feasible for maintenance decision making. Utilities responsible for pipelines maintenance make considerable effort to develop cost-efficient, higher quality, and sustainable approaches. Since existing approaches to predict pipe failure or risk ratings are not computationally efficient in handling large and complex pipeline networks, a more robust technique is required [49]. Therefore there is a need to apply a new computationally efficient approach that is capable of determining the risk of each individual segment of a complex networked system without explicitly performing intensive physically-based analysis. The application of data-driven machine-learning approaches may provide a viable

\* Corresponding author.

E-mail addresses: [rkm562@case.edu](mailto:rkm562@case.edu) (R.K. Mazumder), [ams0098@uah.edu](mailto:ams0098@uah.edu) (A.M. Salman), [yxl1566@case.edu](mailto:yxl1566@case.edu) (Y. Li).

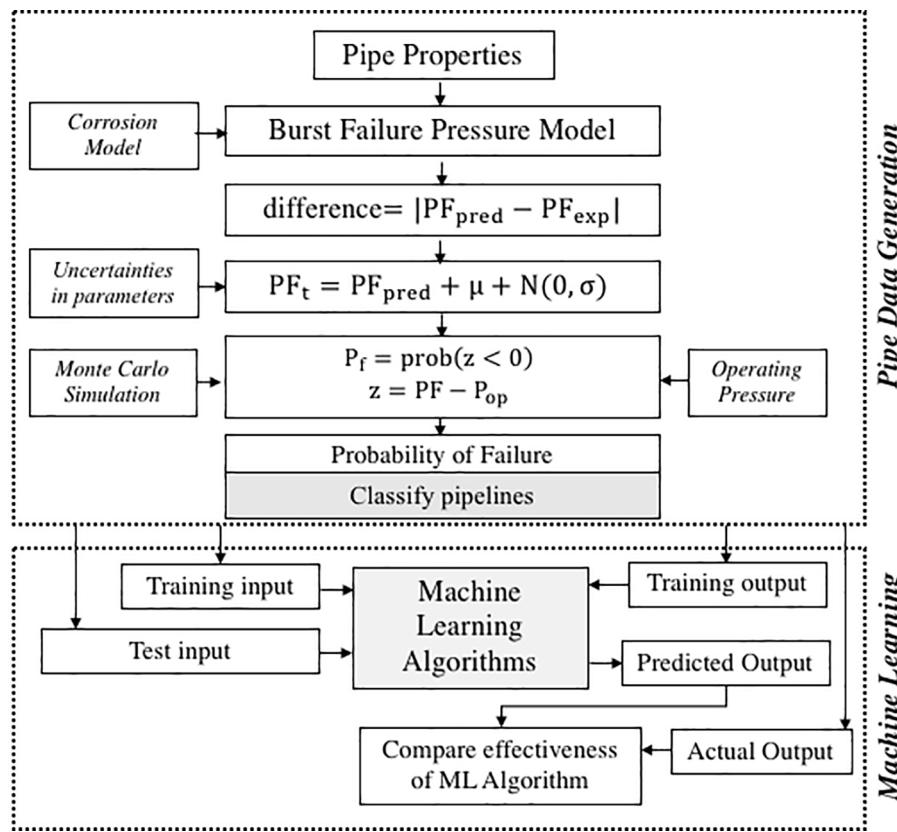


Fig. 1. Proposed Framework.

alternative to computationally demanding probabilistic physical analysis for pipeline failure risk analysis of oil and gas pipeline networks.

The application of machine learning in risk and reliability assessment has gained considerable attention from decision-makers and scientific community in recent years. The machine learning approach can be described as a semi-automated system in which computers learn from the observed data to develop an algorithm. It has been recently studied and applied in civil engineering risk analysis (e.g., [26,29,40–41,42,56,61]). Machine learning is a new dimension of research under active development. A number of algorithms, such as support vector machine, k-nearest neighbours, random forest, have emerged in the last decade [53].

Anghel [6] developed a novel type of support vector machine in a maximum manner to predict the failure probability of pipes with active corrosion defects. The author developed a classification reliability procedure that established a link between artificial intelligence and reliability methods. This approach provides a simple but promising alternative for prioritizing maintenance actions. Huang and Burton [28] utilized machine learning applications to classify the failure modes of reinforced concrete frames with infill walls. Winkler et al. [57] predicted water pipe failures using decision-tree based machine learning algorithms. Bootstrap aggregation and boosting techniques are utilized to improve the accuracy of the model. The performance of the model is evaluated using a confusion matrix. Mangalathu and Jeon [41] utilized various machine learning algorithms to identify the failure modes of circular reinforced concrete bridge columns. The authors compared the efficiency of six different machine learning algorithms for a set of experimental data. Artificial neural network provided the best efficiency among the six algorithms. Kiani et al. [32] explored the application of machine learning algorithms for the derivation of fragility curves. Siam et al. [50] used a clustering algorithm to classify masonry walls utilizing a very limited number of datasets.

Mangalathu et al. [42] applied a data-driven technique to recognize

the failure modes of concrete shear walls, where the authors utilized various boosting models in addition to other machine learning techniques. Ten input parameters are utilized to capture the geometry, material, and reinforcement characteristics in order to recognize the failure modes. The efficiency of machine learning algorithms was determined based on global accuracy, precision, and recall. This study showed a promising outcome for failure mode classification that can be further applied to other types of structures/infrastructures. Other notable machine learning applications include Qi and Zhu [48], Yu et al. [59], Jeon et al. [29], and Zhang et al. [61]. While the application of machine learning is potentially capable of solving complex problems without the necessity of explicit mechanical analysis, limited studies have been carried out to determine the feasibility of various machine learning algorithms application for pipeline burst failure risk analysis. Predicting future pipe failures may help decision-makers to take appropriate intervention measures to avoid potential consequences of burst failure.

This study explores the possibility of applying machine learning algorithms as a viable alternative to computationally intensive physical/analytical approaches. This study has three main objectives: (1) to generate a comprehensive dataset where pipeline characteristics are assumed within a realistic range and combination. Such a dataset includes geometric properties, corrosion pit dimensions, pipe age, and material characteristics of pipelines. This dataset is then used to assess the feasibility of machine learning algorithms for evaluating failure risk of pipelines, (2) to evaluate the effectiveness of machine learning algorithms in the classification of pipelines based on their failure probability, and (3) identify the contribution of each parameter to the pipeline failure probability through a sensitivity analysis.

## 2. Failure prediction model and data generation

A flowchart illustrating the proposed methodology is shown in Fig. 1.

**Table 1**  
Failure pressure models.

Sl.	Model Name	Burst Failure Pressure, FP	
1	B31G [7]	$FP = \begin{cases} \frac{1.11 \cdot 2YSt}{D} \left( \frac{1 - \frac{2d(T)}{3t}}{1 - \frac{2d(T)}{3t} M^{-1}} \right); G < 4 \\ \frac{1.11 \cdot 2YSt}{D} \left( 1 - \frac{d(T)}{t} \right); G \geq 4 \end{cases}$	$M = \sqrt{1 + 0.893 \frac{L(T)^2}{Dt}}; G = 0.893 \frac{L(T)^2}{\sqrt{Dt}}$
2	Modified B31G [33]	$FP = \frac{2.0(Ys + 68.95 \text{ MPa})t}{D} \left( \frac{1 - 0.85 \frac{d(T)}{t}}{1 - 0.85 \frac{d(T)}{t} M^{-1}} \right)$	$M = \begin{cases} \sqrt{1 + 0.6275 \frac{L(T)^2}{Dt} - 0.003375 \frac{L(T)^4}{D^2 t^2}} \text{ if } \frac{L^2}{Dt} \leq 50 \\ 0.032 \frac{L(T)^2}{Dt} + 3.3t \frac{L^2}{Dt} \leq 50 \end{cases}$
3	DNV-RP-F101 [21]	$FP = \frac{2UTS}{D-t} \left( \frac{1 - \frac{d(T)}{t}}{1 - \frac{d(T)}{t} M^{-1}} \right)$	$M = \sqrt{1 + 0.31 \frac{L(T)^2}{Dt}}$
4	Shell-92 [36]	$FP = \frac{1.8UTS}{D} \left( \frac{1 - \frac{d(T)}{t}}{1 - \frac{d(T)}{t} M^{-1}} \right)$	$M = \sqrt{1 + 0.805 \frac{L(T)^2}{Dt}}$
5	Netto et al. [46]	$FP = \frac{1.1 \cdot \sigma_y \cdot 2t}{D} \cdot \left[ 1 - 0.9435 \left( \frac{d}{t} \right)^{1.6} \left( \frac{1}{D} \right)^{0.4} \right]$	
6	Battelle [37]	$FP = \frac{2UTS_t}{D} \left( 1 - \frac{d(T)}{t} M \right)$	$M = 1 - \exp(-0.157 \frac{L(T)}{\sqrt{D(t-d(T))/2}})$

Note: FP is the failure pressure, UTS is the ultimate tensile strength, YS is yield strength, M is Folias factor, D is diameter of pipe, t is the initial thickness of pipe wall, d(T) is depth of corrosion as a function of time, L(T) is length of corrosion as a function of time, T is the time in year.

A database of burst failure tests is compiled from available literature, as shown in Appendix A1. Burst failure pressure for pipelines subjected to corrosion defects is estimated using various standards (e.g., American Society of Mechanical Engineers (ASME) B31G, Det Norske Veritas (DNV) RP-F101).

The experimental burst failure pressure obtained from the tests is compared with the predicted pressure obtained using various standards. The difference between the predicted failure pressure and the experimental failure pressure is analysed to obtain a mean error and statistical distribution of the variability in prediction. Then, the true failure pressure is defined by the sum of predicted pressure, mean error, and statistical distribution of variability in prediction with zero mean and standard deviation  $\sigma$ . The pipeline's failure probability is determined by comparing the true failure pressure and operating fluid pressure of pipe at a burst limit state. The burst failure probability of the pipeline is determined in a probabilistic manner utilizing Monte Carlo simulation. The pipeline is classified into different failure risk classes based on the failure probability of the pipeline. Utilizing this concept, a new dataset of pipelines is generated. Various machine learning algorithms are applied to the developed datasets to identify the best predictive machine learning algorithm.

A detailed description of the different parts of the framework is given in the subsections below.

## 2.1. Burst failure prediction models

Extensive research has been performed to determine the remaining strength and failure pressure of metallic pipelines subjected to corrosion deterioration. The remaining strength of pipelines is estimated using design codes approach for metal loss due to corrosion [33,36,37,8,21,46,55]. A list of commonly used failure prediction models are provided in Table 1, where mathematical equations of failure pressure are provided. However, some of these methods are extremely conservative (e.g. ASME B31G) and predicted failure pressure is much higher than actual failure pressure that may result in unnecessary renewal before pipeline's expected design life leading to significant economic consequences. Among the existing methods, ASME B31G is the most widely used assessment approach. The DNV RP-F101 method significantly improved the failure strength prediction in comparison to older approaches. More recently, researchers performed experimental and numerical analysis and found various standards use safety factors

that often lead to costly and unnecessary maintenance of pipelines [46,55].

Time-dependent failure pressure can be obtained by including time-dependent corrosion pit formation in failure pressure models, as expressed in Table 1. Burst failure pressure models may provide conservative values by ignoring the random characteristic of corrosion [39,35]. Hence, consideration of uncertainties in a corrosion model could provide a realistic scenario [3,60]. Although various protective measures, such as cathodic protection, are used for the pipelines to avoid corrosion, these measures may not always be effective. The protective coating may not last for a long time, particularly for the offshore oil industry [44]. Recent studies revealed that corrosion pit growth on the pipeline wall does not follow the linear pattern. The growth of defect depth can be modeled using power law [14,55,44] and the length of corrosion pit is modeled as a function of pipe age, as expressed below [1,13]:

$$d(T) = k(T - T_d)^\alpha \quad (1)$$

$$L(T) = L_0 + V_a(T - T_i) \quad (2)$$

where  $d(T)$  and  $L(T)$  are depth and length of corrosion pit, respectively;  $k$  and  $\alpha$  are corrosion growth coefficients;  $T_i$  is the time of the last inspection;  $V_a$  is the axial corrosion growth rate. Linear rate is attained over the pipe elapsed life ( $V_a = L_0 / \Delta T$ , where  $\Delta T$  is the pipeline elapsed time to inspection date,  $L_0$  is the initial length). Pipeline elapsed time of 15 years is assumed reasonable to guarantee that the growth rate comes to a steady state. The depth of the corrosion pit is simulated using Monte Carlo Simulation. Coefficients of corrosion are assumed to be normally distributed. The mean and coefficient of variation for  $k$  and  $\alpha$  are taken as (0.164/yr, 0.20) and (0.780, 0.15), respectively. Since the soil types are unknown, these values  $k$  and  $\alpha$  are taken for average soil type from Caleyo et al. [14]. The time required to initiate corrosion was assumed equal to 2.88 years [14,55].

## 2.2. Predicted and experimental burst failure pressure

Burst failure reliability of pipeline subjected to corrosion defect is typically performed estimating failure pressure (or remaining pipeline strength) based on various standards, as shown in Table 1. The residual between predicted failure pressure and experimental failure pressure is calculated by the following equation:

**Table 2**

Statistical distribution of random variables.

Parameters	Mean	COV	Unit	Distribution Type
Pipe diameter, D	–	–	mm	Deterministic
Wall thickness, t	–	–	mm	Deterministic
Defect depth, d	$\mu_d$	0.10	mm	Normal
Defect length, L	$\mu_L$	0.10	mm	Normal
Yield Strength (YS)	$\mu_{YS}$	0.07	MPa	Lognormal
Ultimate Tensile Strength (UTS)	$\mu_{UTS}$	0.07	MPa	Lognormal
Operating pressure, $P_{opt}$	$\mu_{P_{opt}}$	0.10	MPa	Normal

Note: COV: coefficient of variation.

$$\Delta_p = FP_{pred} - FP_{exp} \quad (3)$$

where  $\Delta_p$  is the pressure difference between predicted and experimental pressures,  $FP_{pred}$  is the predicted failure pressure estimated using various models expressed in Table 1,  $FP_{exp}$  is the experimental failure pressure obtained from the laboratory experiment.

Expected true burst failure pressure of a pipeline,  $FP_t$ , can be estimated by the sum of predicted pressure, mean error in the prediction and statistical distribution of variability in prediction [20], as

$$FP_t = FP_{pred} + \mu + N(0, \sigma^2) \quad (4)$$

where  $\mu$  is the mean difference between prediction and experimental pressures, and  $N$  represents variability of prediction of a normal distribution with zero mean and variance  $\sigma^2$ . Equation (4) is applied to determine the expected true failure pressure of pipelines. Consideration of variability between predicted and experimental results in Equation (4) reduces the conservativeness in true failure pressure prediction.

### 2.3. Burst failure limit state

The burst failure probability of pipeline is determined by comparing the true burst failure pressure and operating pressure inside a pipeline [13,55]. The pipe's failure occurs when applied internal pressure exceeds the true failure pressure of a pipeline. The limit state is defined as:

$$g(X) = FP_t - P_{opt} \quad (5)$$

where  $FP_t$  is the pipeline true burst failure pressure of the pipeline,  $P_{opt}$  is the internal fluid pressure of the pipe,  $g(X)$  is the burst limit state of the pipeline. Allowable operating pressure inside a pipeline is estimated by the following equation [16,20]:

$$P_{opt} = \frac{2t}{D} YS \times F_f \times L_f \times J_f \times T_f \quad (6)$$

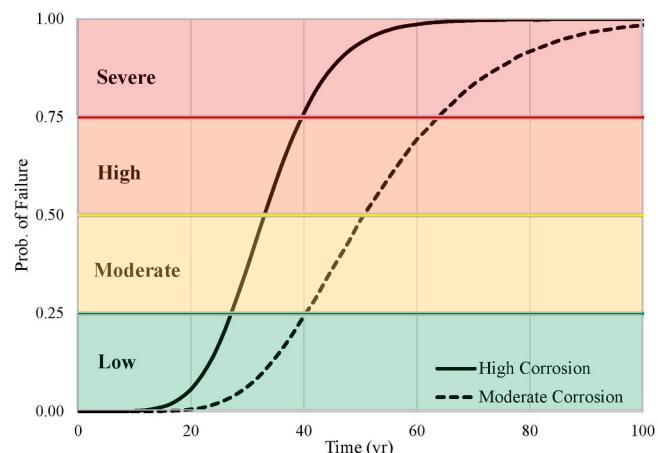
where  $F_f$  is the design safety factor (0.8);  $L_f$  is the location factor;  $J_f$  is the joint factor;  $T_f$  is the temperature factor.  $L_f$ ,  $J_f$  and  $T_f$  are taken equal to 1.0 [16]. The failure of the pipeline is defined by the failure likelihood that internal pressure exceeds the true burst strength of the pipe (positive values indicate safe operation and a negative value indicates failure in this case). The probability of failure of a pipeline with corrosion defect is defined using Monte Carlo simulation, as below;

$$P_f = \frac{\sum g(x) \leq 0}{N_m} \quad (7)$$

where  $N_m$  is the number of Monte Carlo data points.

The probability of failure is determined using a physical probabilistic approach where time-dependent true failure pressure is estimated by incorporating time-dependent corrosion model. The statistical distribution of random variables for Equations (5)-(7) is taken from Calyeo et al. [13] and Wang and Zarghamee [55] and shown in Table 2.

A large number of pipeline dataset is derived based on the properties obtained from previous research. The failure probability of the pipelines is modeled by comparing the failure pressure and operating pressure, as

**Fig. 2.** Probability of failure levels.

shown by the limit state in Equation (5). Based on the predicted failure probability, each pipeline is classified into low (0–0.25), moderate (0.25–0.5), high (0.5–0.75), and severe (0.75–1.0) failure risk classes, as shown in Fig. 2. This classification is performed similar to the risk matrix, where the probability of pipeline failure is classified from low to severe failure risk groups.

## 3. Machine learning algorithms

The objective of using machine learning algorithms is to classify each pipe into failure risk class in a complex and large oil and gas distribution network to support asset management decisions. Eight machine learning algorithms are applied to classify the pipeline failure likelihood at the burst limit state. Pipelines burst failure probability is classified into four failure risk classes depending on the attributes of parameters. The various machine learning algorithms are discussed briefly below.

### 3.1. K-Nearest neighbours

K-nearest neighbors (KNN) is the most basic and useful algorithms used in the industry for classification problems [9]. KNN stores available data and classifies new input based on similarity among the pre-existing data set. In this method, for a new observation  $x$ , KNN tries to match the nearest  $K$  number of values in the stored data that are close to  $x$ . This approach is suitable when little or no information about the data is available. The conditional probability that a new data ( $x$ ) belongs to class  $k$ ,  $P_k(x)$ , can be estimated as below [41]:

$$P_k(x) = Pr(Y = k | X = x) = \frac{1}{k} \sum_{i \in N_k} I(y_i = k) \quad (8)$$

where  $N_k$  is the number of data points.

### 3.2. Decision tree

A decision tree (DT) is a supervised learning algorithm for classifying data. The DT is a predictive approach that classifies an item based on past observation. The algorithm is a treelike structure upside down where each internal node represents a test on an attribute, each branch illustrates the outcome of the test, and terminal nodes hold class levels. In this approach, a tree is generated by splitting the dataset into a number of non-overlapping subsets. Splitting process is repeated for each subset in a recursive pattern. The advantage of using the DT algorithm is that it is easy to interpret the results to non-technical users, and little effort is required compared to other algorithms. This is regarded as one of the most powerful and popular tools for classifying datasets [53,42]. The dataset is split into 's' distinct and overlapping

Actual Data								
Predicted Data	True Positive	False Positive	True Positive + True Negative					
	False Negative	True Negative	True Positive + True Negative + False Positive + False Negative					
			Recall = $\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$					
			Precision = $\frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$					

**Fig. 3.** Confusion matrix.

regions. The binary splitting is performed based on the Gini Index (GI) expressed as below [10]:

$$GI = \sum_{k=1}^l \widehat{P}_{mk} (1 - \widehat{P}_{mk}) \quad (9)$$

where GI represents total variance across the k-classes,  $\widehat{P}_{mk}$  represents the portion of training observation in m-th region from the k-th class.

### 3.3. Random forest

A random forest (RF) is an ensemble learning method of classifications. This method is an easy to use supervised machine learning approach. In the RF model, an ensemble decision tree is usually built with bragging [11]. The idea of using the bragging method is to use a combination of learning models to increase overall accuracy of the results. Instead of using all the features used in the DT, RF uses random features dragged from the input features. This algorithm has three main features: a) generating bootstrap samples from the training data, b) developing a decision tree from each bootstrap using best split, and c) predicting the output of a new data utilizing decision trees developed from bootstrap [41]. The random forest regression predictor is expressed as below [53]:

$$\widehat{P}_{rf}(x) = \frac{1}{N_t} \sum_{k=1}^{N_t} f_{N_k}(x) \quad (10)$$

where  $\widehat{P}_{rf}$  is the predicted outcome from a RF from a total number of  $N_t$  trees;  $f_{N_k}$  denotes the predictor of an individual tree with an input vector  $x$ .

### 3.4. Naïve Bayes

The naïve Bayes (NB) is a machine learning classification algorithm developed based on Bayes theorem. In this approach, the Bayesian theory is applied with the ‘naïve’ assumption of conditional intercedence among input variables given the value of the class variable [47]. It assumes the effect of an input parameter is independent of other input values in the same class. The classification function is expressed as below [41]:

$$P(x) = P(Y = f|X = x) = \frac{\pi_f f_f(x)}{\sum_{l=1}^k \pi_l f_l(x)} \quad (11)$$

where  $\pi_f$  is the prior probability of randomly chosen observation and  $f_f$  is the probability density function from the f-th class.

### 3.5. Boosting algorithms

Boosting algorithms, such as AdaBoost, XGBoost, LightBoost (LGBM), and CatBoost, are used to improve the performance of a model by incorporating a set of weak classifiers from a strong classifier. The weak classifier is included in such a way that it improves the performance of the algorithm significantly. In AdaBoost, observations are weighted equally, and then higher weight is assigned for incorrect

**Table 3**  
Summary test results and failure pressure prediction statistics.

	B31G	Modified B31G	DNV	Battelle	Shell	Netto	Actual
Mean	10.1	13.6	14.7	14.2	9.6	9.4	15.5
Stdev	5.3	4.9	6.4	6.2	4.4	6.1	5.6
Max	23.5	22.8	25.9	25.2	19.1	21.2	28.0
Min	2.6	4.6	4.9	4.5	2.5	0.6	8.0
%Error							
Mean	35.9	16.8	11.3	13.5	39.9	44.1	–
Stdev	18.9	12.2	9.8	9.1	10.2	21.5	–

Note: Stdev: standard deviation.

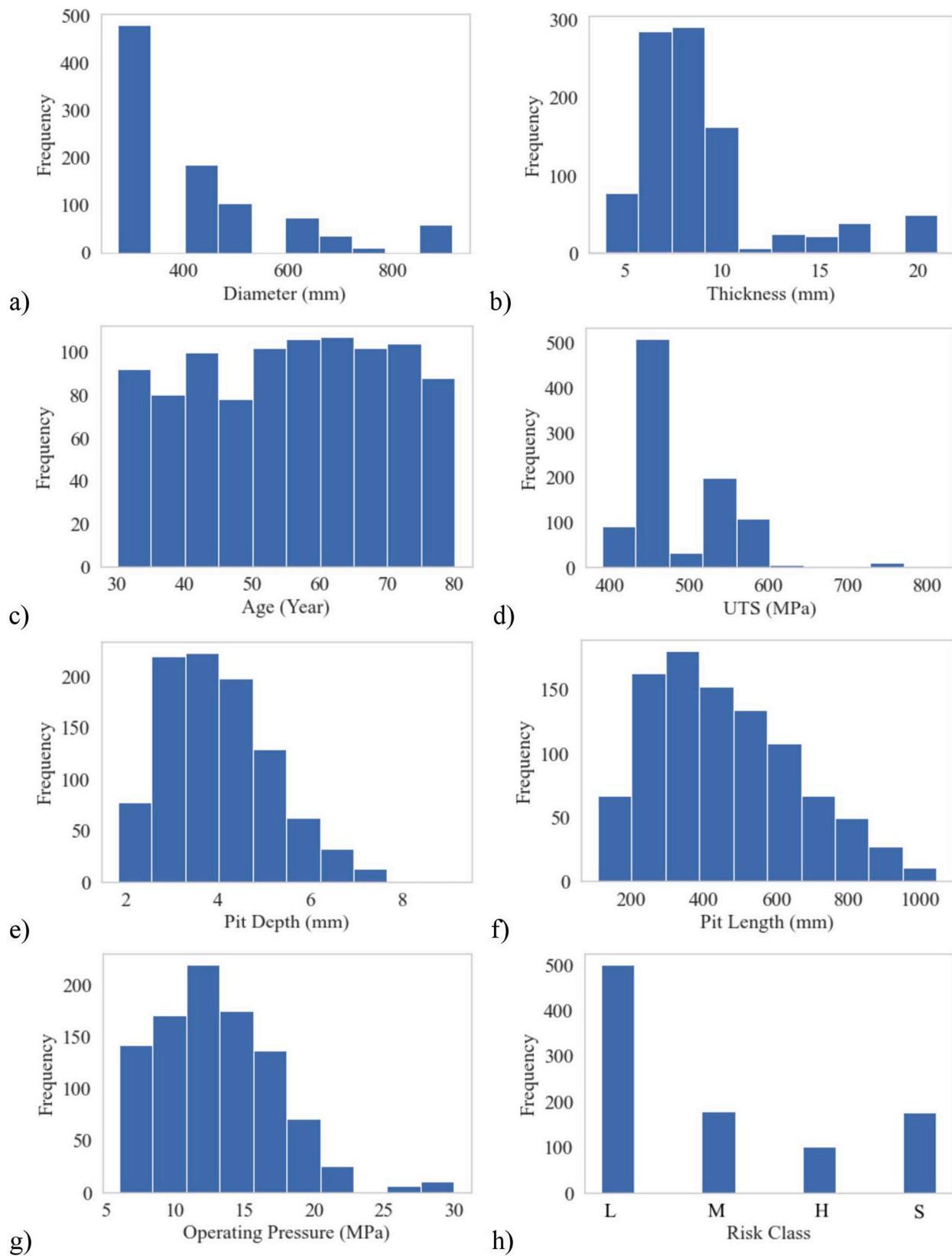
classifiers than the weight assigned for a weak classifier [58]. The other boosting algorithms are used as a gradient boosting where decision trees are used as base weak learner, and iteratively update the algorithm to obtain targeted outcomes [24,42]. XGBoost tries to reduce the misclassification error in the previous model [18]. Since this model uses sequential training steps, it is relatively slower to implement. LightGBM uses a decision tree algorithm where the model implication happens leafs-wise [31]. The CatBoost handles categorical features by taking the input of features names [22,42]. A detail of these boosting algorithms can be found elsewhere [58,18,31,22].

The accuracy of various machine learning algorithms is determined based on the confusion matrix. Previous studies also determined the accuracy of machine learning algorithms using the confusion matrix [49,41,42]. In a confusion matrix, the accuracy of an algorithm is expressed by the total percentage of correct prediction. However, sometimes an unbalanced dataset may result in misleading the prediction accuracy. Hence, the other two measures of the confusion matrix, *recall*, and *precision* can be very useful for determining the performance of algorithms. Recall denotes the percentage of right prediction for ‘true positive’ classes, whereas Precision denotes the percentage of right prediction for ‘true-negative’ classes. A generic confusion matrix and corresponding mathematical expressions are presented in Fig. 3 [38,49].

Since the objective of the current study is to determine machine learning algorithms that are computationally efficient and can be used as a viable alternative to the physics-based approach, the optimal algorithm is identified based on the confusion matrix and the competitive efficiency to execute the analysis.

## 4. Case studies

A set of 92 experimental test results for oil and gas pipelines with corrosion defect is compiled from previous studies and presented in Appendix A1. Freire et al. [23] performed 14 full-scale laboratory tests for steel pipelines with corrosion defects to understand the internal pressure failure mechanism. Mok et al. [45] performed a burst behavior of 5 steel pipelines with long external corrosion. Similarly, Chen et al. [17], Mannucci and Harris [43], Kim et al. [34], and Shuai et al. [51] performed burst failure tests for pipelines with external corrosion defects. Cronin and Pick [20] performed a set of experimental tests and developed a database for the steel pipelines removed from the operation due to corrosion defects on pipeline walls. These experimental test results are compared with various standards codes. For the pipelines



**Fig. 4.** a) Diameter; b) thickness; c) age; d) ultimate yield strength; e) depth of corrosion; f) length of corrosion; g) operating pressure; and h) failure risk class.

documented in [Appendix A1](#), burst failure pressures are predicted utilizing the six different burst failure pressure prediction models shown in [Table 1](#). Experimental burst failure pressures are compared with the predicted results to determine the variability in prediction. The

summary of test results and errors (%) are illustrated in [Table 3](#). The error percentage is expressed as:

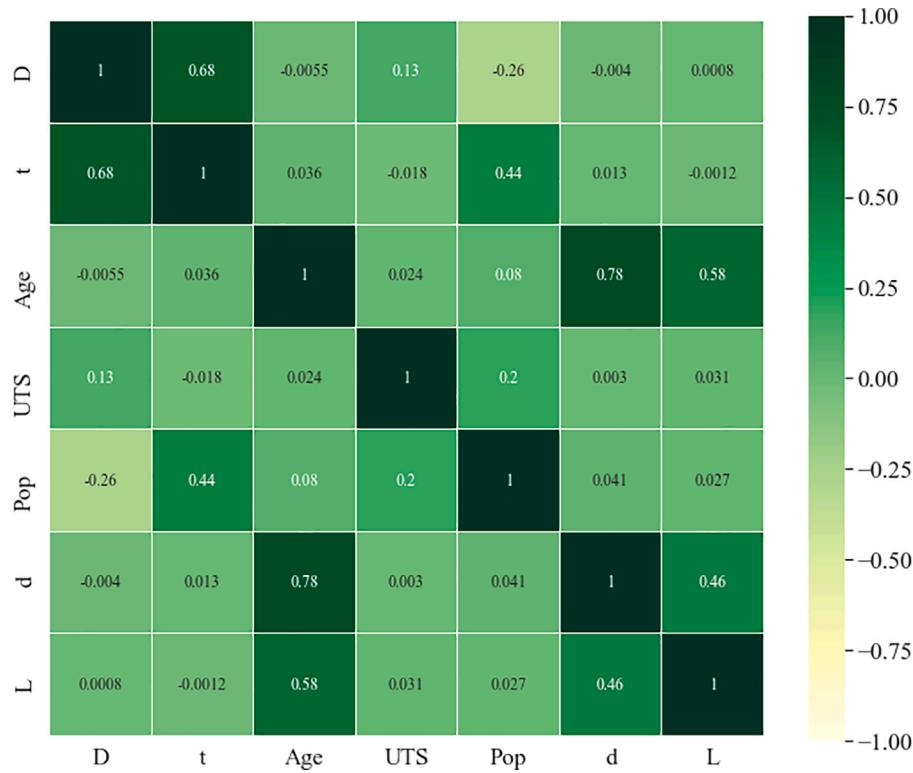


Fig. 5. Correlation matrix.

$$\%Error = \frac{|FP_{predicted} - FP_{exp}|}{FP_{exp}} \times 100 \quad (12)$$

**Table 3** shows that DNV-RP-F101 provides the best prediction as it has the lowest level of %error in the prediction, whereas B31G and Netto et al. [46] models provide the conservative results in the prediction. Literature also supports that DNV-RP-F101 model can be applied to steel pipelines under various conditions [19,27,4]. Hence, a new database for true burst failure pressure is simulated based on burst failure pressure estimated using DNV-RP-F101 standard, mean predicted error, and normal distribution of variability in prediction, using Eq. (5).

A set of 959 pipelines data is randomly generated within the realistic/possible ranges of pipeline characteristics (e.g., diameter, thickness, ultimate tensile strength). All these pipelines are assumed to be made of steel. Only the parameters required to estimate the burst failure pressure are generated. Fig. 4 shows a summary of pipelines data generated in this study. Constraints for different parameters are assumed based on previous studies. The range of diameter was chosen to be 274–914 mm, and the range of thickness was chosen as 5–20 mm, as shown in Fig. 4(a) and 4(b), respectively. The age of the pipelines is randomly assigned between 0 and 100 years, as shown in Fig. 4(c). The ultimate tensile strengths of the pipelines mostly lie between 450 and 550 MPa, as shown in Fig. 4(d).

The size of the corrosion pit depends largely on the pipe age and corrosion growth rate. In the field, the corrosion pit's size can be estimated using a smart measuring device (e.g., smart pig). Depth and length of corrosion pit are predicted using Equations (1) & (2), respectively. The depth and length of corrosion gradually increased over time. Length of corrosion defect ( $L_0$ ) at the inspection was assumed to be equal to 50–200 mm [13]. Fig. 4(e) and 4(f) show the summary of depth and length of corrosion pit, respectively. The operating internal fluid pressure was estimated based on the CSA [16] guideline expressed by Equation (6), as shown in Fig. 4(g). The probability of failure of each pipeline is estimated using a probabilistic approach by comparing the burst failure strength with the internal operating pressure, as expressed

by the burst limit state in Equation (5). Monte Carlo simulation is performed to execute the burst failure limit state and generate the probability of failure value. Statistical distributions of random variables in the simulation are shown in **Table 2**. Then, pipelines are classified into various risk classes, as expressed in **Fig. 2**. **Fig. 4(h)** shows a summary of pipeline failure risk classes obtained from the simulation.

A correlation matrix is developed to learn about the relationship between various input pairs for failure probability classification, as shown in **Fig. 5**. A high positive correlation coefficient value of a particular cell represents a strong relationship between the input parameters corresponding to the cell. It can be seen from the correlation matrix that some parameters are strongly correlated, and others are weakly correlated. For instance, the corrosion pit's depth and length of corrosion pit are strongly correlated with the pipe age. This is because the corrosion pit's size is estimated as a function of pipe age, using Equations (1) & (2). On the other hand, it can be said that no correlation exists between pipe age and diameter.

Once pipelines data is generated and classified into various risk classes, as shown in **Fig. 4(h)**, eight machine learning algorithms described in the methodology are applied to identify the best failure prediction algorithm. The first four algorithms (i.e., KNN, DT, RF, and NB) have been used extensively in recent research. The use of boosting algorithms is still limited [42]. In each machine learning approach, 70% of data is used to train and develop the machine learning model, while the rest of the 30% is used to validate the models. Training and testing data are randomly selected. 5-fold cross-validation is also performed to estimate the model accuracy in order to reduce biases in data splitting for training and testing.

The performance of each machine learning model for both training and test set is presented using confusion matrixes as shown in **Fig. 6**. A general confusion matrix represents four prediction terminologies: True Positive, True Negative, False Positive, and False Negative, as shown in **Fig. 3**. True positive and true negative represent a state where data is accurately predicted. On the other hand, false positive and false negative represent ‘Type I error’ and ‘Type II error’ in prediction, respectively

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	50.7%	1.6%	0.0%	0.0%	97.0%
	M	3.9%	13.4%	0.4%	0.1%	75.0%
	H	0.4%	3.7%	5.1%	0.4%	52.0%
	S	0.1%	0.6%	1.6%	17.7%	88.0%
Precision	92.0%	69.0%	71.0%	97.0%	80.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	52.3%	0.0%	0.0%	0.0%	100.0%
	M	0.0%	17.9%	0.0%	0.0%	100.0%
	H	0.0%	0.0%	9.7%	0.0%	100.0%
	S	0.0%	0.0%	0.0%	20.1%	100.0%
Precision	100.0%	100.0%	100.0%	100.0%	100.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	52.3%	0.0%	0.0%	0.0%	100.0%
	M	0.0%	17.9%	0.0%	0.0%	100.0%
	H	0.0%	0.0%	9.7%	0.0%	100.0%
	S	0.0%	0.0%	0.0%	20.1%	100.0%
Precision	100.0%	100.0%	100.0%	100.0%	100.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	48.0%	4.2%	0.0%	0.1%	92.0%
	M	5.2%	11.8%	0.9%	0.0%	66.0%
	H	0.4%	3.3%	3.9%	2.1%	40.0%
	S	0.0%	1.6%	1.6%	16.8%	84.0%
Precision	89.0%	56.0%	60.0%	88.0%	80.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	52.3%	0.0%	0.0%	0.0%	100.0%
	M	0.0%	17.9%	0.0%	0.0%	100.0%
	H	0.0%	0.0%	9.7%	0.0%	100.0%
	S	0.0%	0.0%	0.0%	20.1%	100.0%
Precision	100.0%	100.0%	100.0%	100.0%	100.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	52.3%	0.0%	0.0%	0.0%	100.0%
	M	0.0%	17.9%	0.0%	0.0%	100.0%
	H	0.0%	0.0%	9.7%	0.0%	100.0%
	S	0.0%	0.0%	0.0%	20.1%	100.0%
Precision	100.0%	100.0%	100.0%	100.0%	100.0%	

		Predicted Class				Recall
Observed Class	L	M	H	S		
	L	52.3%	0.0%	0.0%	0.0%	100.0%
	M	0.0%	17.9%	0.0%	0.0%	100.0%
	H	0.0%	0.0%	9.7%	0.0%	100.0%
	S	0.0%	0.0%	0.0%	20.1%	100.0%
Precision	100.0%	100.0%	100.0%	100.0%	100.0%	

Fig. 6. Confusion matrices for training sets.

[5]. In this classification problem, diagonal cells in the confusion matrix accurately predicted the targeted data. Other cells, except diagonal ones in the confusion matrix, are not accurately predicted [38]. Fig. 6(a)-(h) show the confusion matrixes developed based on the training dataset. The fifth row and fifth column of the confusion matrix represent the precision and recall values, respectively. The accuracy of the model is represented by the last cell (5:5) of the matrix. Fig. 6 shows that KNN, DT, RF, NB, AdaBoost, XGBoost, LGBoost, and CatBoost models have accuracy of 80%, 100%, 100%, 80%, 79%, 100%, 100% and 100% for the training dataset, respectively.

Fig. 7 represents the confusion matrixes for the test sets. Accuracies of KNN, DT, RF, NB, AdaBoost, XGBoost, LGBoost, and CatBoost for the test sets are 75%, 77%, 82%, 74%, 74%, 85%, 84% and 86%, respectively. Additionally, since the database generated in the current approach is still not large enough in compare to a data typically presence in a real utility network, cross-fold resampling procedure is applied to determine the efficiency of machine learning models. Cross-validation is used to avoid the selection bias of dataset and overfitting problems [15].

After performing a 5-fold cross-validation, mean accuracy of KNN, DT, RF, NB, AdaBoost, XGBoost, LGBoost, and CatBoost models are found to be 77%, 80%, 85%, 78%, 70%, 84%, 84% and 78%, respectively. Among these algorithms, RF provides the best accuracy in prediction.

Since the objective of this study is to determine computationally efficient algorithms as a viable alternative to the physics-based approach, the computational efficiency of each algorithm is determined. The current analysis is performed using a Jupyter notebook on a personal computer with 4 GB RAM and 1.6 GHz Core i5 processor. Computation time requires to perform KNN, DT, RF, NB, AdaBoost, XGBoost, LGBoost, and CatBoost analyses are 1, 0.3, 5.9, 0.4, 1.9, 2.1, 4.6, and 39 secs, respectively. Although RF provides the best accuracy over other algorithms, it is not the fastest algorithm. DT is the fastest algorithm in this case study. Among the three best-performing algorithms, the computational times for RF and LGBoost are about 2.8 and 2.2 times slower than XGBoost. Hence, XGBoost can be applied as the optimal algorithm as an alternative to a physics-based approach. While even the slowest algorithm takes about 39 secs, the physics-based

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	20.3%	1.9%	0.0%	0.0%	91.0%
	M	3.3%	4.9%	0.6%	0.0%	56.0%
	H	0.6%	1.9%	1.9%	1.0%	35.0%
	S	0.0%	0.3%	0.9%	5.2%	81.0%
Precision	84.0%	54.0%	57.0%	83.0%	75.0%	

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	20.3%	1.9%	0.0%	0.0%	91.0%
	M	2.5%	4.9%	1.2%	0.1%	56.0%
	H	0.0%	1.5%	2.5%	1.5%	46.0%
	S	0.1%	0.0%	0.9%	5.4%	84.0%
Precision	88.0%	59.0%	55.0%	77.0%	77.0%	

(a) K-Nearest Neighbours

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	21.3%	0.9%	0.0%	0.0%	96.0%
	M	2.2%	6.3%	0.3%	0.0%	71.0%
	H	0.0%	1.9%	1.8%	1.8%	32.0%
	S	0.0%	0.1%	0.3%	6.0%	93.0%
Precision	91.0%	68.0%	75.0%	77.0%	82.0%	

(b) Decision Tree

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	20.6%	1.6%	0.0%	0.0%	93.0%
	M	3.6%	4.8%	0.3%	0.1%	54.0%
	H	0.6%	2.2%	1.2%	1.5%	22.0%
	S	0.0%	0.3%	0.7%	5.4%	84.0%
Precision	83.0%	53.0%	53.0%	77.0%	74.0%	

(d) Naïve Bayes

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	21.2%	1.0%	0.0%	0.0%	95.0%
	M	1.2%	7.2%	0.4%	0.0%	81.0%
	H	0.0%	1.9%	2.2%	1.3%	41.0%
	S	0.0%	0.0%	0.4%	6.0%	93.0%
Precision	95.0%	71.0%	71.0%	82.0%	85.0%	

(e) AdaBoost

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	20.7%	1.5%	0.0%	0.0%	93.0%
	M	1.2%	7.0%	0.6%	0.0%	80.0%
	H	0.0%	1.8%	2.5%	1.2%	46.0%
	S	0.0%	0.1%	0.3%	6.0%	93.0%
Precision	95.0%	67.0%	74.0%	83.0%	84.0%	

(f) XGBoost

		Predicted Class				Recall
		L	M	H	S	
Observed Class	L	20.9%	1.3%	0.0%	0.0%	94.0%
	M	1.0%	7.3%	0.4%	0.0%	83.0%
	H	0.0%	1.9%	2.7%	0.9%	49.0%
	S	0.0%	0.0%	0.4%	6.0%	93.0%
Precision	95.0%	69.0%	75.0%	87.0%	86.0%	

(g) LGBoost

(h) CatBoost

Fig. 7. Confusion matrices for test sets.

approach (using the DNV-RP-F101 model) takes about 470 secs to evaluate the failure probability of pipelines, which is about 12 times more computationally demanding. The optimal algorithm (XGBoost) is more than 220 times faster than the physics-based approach. This comparison represents how a machine learning algorithm can be a better alternative for the future failure risk analysis of pipelines. Boosting algorithms such as XGBoost and LGBoost performed well and faster relative to other two algorithms and can be suggested for future analysis. The CatBoost is proven to be the slowest algorithm. Although RF has the highest accuracy among the first four widely used algorithms, it takes a relatively long training time. On the other hand, DT is recommended for future analysis as it is faster to train with relatively good prediction accuracy.

## 5. Conclusions

Failure risk analysis of pipelines is very important for effectively managing the extensive and aging pipe networks in the oil and gas

industries. The application of physics-based reliability assessment for large and complex oil and gas networks requires huge computational costs. This study investigated the feasibility of using machine learning algorithms as an alternative to the physically-based approach, either deterministic or probabilistic, for predicting the failure of oil and gas pipelines. This study showed that Machine learning algorithms exhibited promising performance in pipeline failure prediction.

A set of 959 pipelines data is randomly generated within the realistic/possible ranges of pipeline characteristics. Burst failure probability of pipelines was investigated by comparing expected true burst failure pressure and operating pressure. Then pipelines are classified into various failure risk classes depending on the probability failure. Eight machine learning algorithms are applied, and their performance was evaluated using a confusion matrix. Among these algorithms, the random forest provides the best accuracy in failure risk prediction. However, the computational time required for the random forest is relatively higher than other best performing boosting algorithms XGBoost and LGBoost. The XGBoost is the optimal machine learning

**Table A1**

Burst Failure Pressure Summary (experimental and predicted).

Sl.	D mm	t mm	L mm	d mm	YS MPa	UTS MPa	Exp. MPa	B31G MPa	M.B31G MPa	DNV MPa	Battelle MPa	Shell MPa	Netto MPa	Pop MPa	Ref.
1	459	8.1	40	5.4	601	684	23.0	20.2	20.6	21.9	20.6	15.0	17.0	17.0	Freire et al. [23]
2	459	8	40	3.8	589	731	24.0	19.5	21.3	24.6	23.3	16.4	17.6	16.4	
3	324	9.8	256	7.1	452	542	14.0	8.4	14.3	12.7	12.5	8.2	8.9	21.9	
4	324	9.7	306	6.8	452	542	14.0	9.0	14.7	12.8	12.2	8.5	7.7	21.7	
5	324	9.7	350	6.9	452	542	14.0	8.7	14.2	12.0	11.1	8.0	6.2	21.7	
6	324	9.7	395	6.9	452	542	13.0	8.7	14.1	11.7	10.6	7.9	5.0	21.7	
7	324	9.9	433	7.3	452	542	12.0	8.1	13.5	10.8	9.6	7.3	3.6	22.1	
8	324	9.7	467	7	452	542	12.0	8.4	13.6	11.0	9.7	7.5	3.1	21.7	
9	324	9.8	489	7	452	542	12.0	8.7	13.9	11.3	10.0	7.7	2.8	21.9	
10	324	9.8	500	7	452	542	12.0	8.7	13.8	11.3	10.0	7.7	2.5	21.9	
11	324	9.7	528	7.1	452	542	11.0	8.1	13.1	10.4	9.1	7.1	1.5	21.7	
12	508	14.6	500	10.4	478	600	15.0	8.8	14.4	12.9	12.1	8.2	7.2	22.0	
13	508	14.3	500	10.3	478	600	13.0	8.4	13.9	12.3	11.5	7.8	6.9	21.5	
14	508	14.8	500	9.7	478	600	16.0	10.7	16.2	15.4	14.6	9.9	8.4	22.3	
15	508	6.6	381	2.6	540	610	11.0	9.4	11.3	10.9	10.6	8.2	8.1	11.2	Mok et al. [45]
16	508	6.4	900	3.4	540	610	8.0	7.1	8.8	7.8	7.3	6.0	2.8	10.9	
17	508	6.4	900	2.2	540	610	12.0	9.9	11.1	10.6	10.2	8.2	5.7	10.9	
18	508	6.4	1000	3.2	540	610	8.0	7.6	9.1	8.2	7.7	6.3	2.8	10.9	
19	508	6.7	1016	2.7	540	610	12.0	9.4	10.8	10.2	9.6	7.9	4.4	11.4	
20	762	17.5	50	8.8	495	565	28.0	23.5	17.0	25.8	24.3	19.1	19.9	18.2	Kim et al. [34]
21	762	17.5	100	8.8	495	565	24.0	19.9	17.0	24.1	22.8	16.9	18.2	18.2	
22	762	17.5	200	8.8	495	565	22.0	13.6	17.0	20.7	20.5	14.0	16.0	18.2	
23	762	17.5	300	8.8	495	565	20.0	9.9	16.9	18.5	18.7	12.7	14.4	18.2	
24	762	17.5	600	8.8	495	565	17.0	12.5	16.5	15.8	15.4	11.4	10.9	18.2	
25	762	17.5	900	8.8	495	565	15.0	12.5	16.1	14.9	14.0	11.0	8.4	18.2	
26	762	17.5	200	4.2	474	557	24.0	13.0	21.1	24.1	23.4	17.1	18.5	17.4	
27	762	17.5	200	8.9	474	557	22.0	13.0	16.2	20.3	20.1	13.3	15.3	17.4	
28	762	17.5	200	13.1	474	557	17.0	13.1	11.2	14.3	15.3	8.3	12.9	17.4	
29	762	17.5	100	8.4	474	557	24.0	19.1	16.8	24.0	22.7	16.4	17.7	17.4	
30	762	17.5	300	8.5	474	557	20.0	9.5	16.6	18.6	18.7	12.5	14.0	17.4	
31	762	17.5	200	8.4	474	557	23.0	13.0	16.8	20.8	20.5	13.8	15.6	17.4	
32	762	17.5	200	9	474	557	23.0	13.0	16.1	20.2	20.0	13.2	15.2	17.4	
33	426	7	160	2.7	240	390	11.0	3.0	7.5	10.0	10.0	5.1	5.6	6.3	Chen et al. [17]
34	426	7	150	3.8	240	390	10.0	3.2	6.3	8.5	8.7	4.1	5.0	6.3	
35	426	7	150	5.2	240	390	8.0	3.2	4.6	5.6	6.2	2.5	4.2	6.3	
36	529	9	350	4.7	285	415	9.0	5.1	7.5	8.3	8.2	4.7	4.9	7.8	
37	529	9	160	4.7	285	415	16.0	4.6	7.7	10.1	10.2	5.4	6.5	7.8	
38	529	9	150	5.3	285	415	14.0	4.8	7.0	9.5	9.7	4.9	6.3	7.8	
39	720	8	180	4.3	425	535	10.0	4.4	6.9	8.2	8.4	5.1	6.6	7.6	
40	720	8	320	4.4	425	535	9.0	2.6	6.6	6.9	7.0	4.5	5.5	7.6	
41	720	8	180	6.2	425	535	8.0	4.4	4.6	5.1	5.7	2.9	5.6	7.6	
42	305	6.4	26	5	351	543	15.0	14.4	14.9	19.8	18.4	9.4	11.4	11.8	Shuai et al. [51]
43	305	6.4	33	4.3	382	570	16.0	14.7	16.0	21.1	19.9	10.7	12.4	12.8	
44	305	6.4	37	4.6	351	463	14.0	12.9	14.0	16.0	15.3	8.8	10.9	11.8	
45	324	6	19	3.6	382	570	16.0	14.6	15.8	20.7	19.3	11.6	12.2	11.3	
46	324	10.3	243	5.2	380	514	23.0	6.7	18.3	20.8	20.6	12.4	11.9	19.3	
47	324	10.3	243	5.2	380	514	22.0	6.7	18.3	20.8	20.6	12.4	11.9	19.3	
48	508	6.6	381	2.6	443	599	11.0	7.7	9.5	10.7	10.4	6.7	6.6	9.2	
49	508	6.4	900	3.4	430	573	8.0	5.6	7.2	7.3	6.8	4.7	2.3	8.7	
50	508	6.4	1000	3.2	435	573	8.0	6.1	7.5	7.7	7.2	5.1	2.2	8.8	
51	508	6.7	1016	2.7	430	601	12.0	7.5	8.9	10.0	9.5	6.3	3.5	9.1	
52	324	9.8	256	7	423	590	14.0	8.1	13.8	14.3	14.1	7.9	8.5	20.5	
53	324	9.7	350	6.9	423	590	14.0	8.1	13.4	13.1	12.1	7.5	5.8	20.3	
54	324	9.9	433	7.1	423	590	12.0	8.1	13.3	12.6	11.3	7.3	3.8	20.7	
55	324	9.7	528	7.1	423	590	11.0	7.5	12.4	11.3	9.9	6.7	1.4	20.3	
56	914	16.4	150	9	739	813	21.0	19.8	18.0	24.5	23.8	17.1	20.0	21.2	Mannucci and Harris [43],
57	914	16.4	450	6	739	813	24.0	8.5	21.8	22.4	22.3	16.9	18.0	21.2	
58	273	4.7	48	2.6	351	454	14.0	8.2	11.3	12.8	12.5	7.5	9.0	9.7	Cronin and Pick [20]
59	273	4.9	102	2.2	351	454	15.0	4.7	10.4	12.1	12.2	7.5	8.5	10.1	
60	274	4.9	46	1.6	351	454	15.0	8.6	13.5	15.2	14.7	9.7	10.6	10.0	
61	274	5	124	2.2	351	454	13.0	4.1	10.7	11.9	12.0	7.5	8.2	10.2	
62	274	5	38	2.7	351	454	15.0	10.1	12.9	14.6	14.0	8.8	10.0	10.2	
63	529	9	350	4.7	285	415	9.0	5.1	7.5	8.3	8.2	4.7	4.9	7.8	
64	324	8.5	0	0	356	469	25.0	20.7	22.3	25.3	24.6	16.8	20.7	14.9	
65	323	8.6	64	2.2	356	469	24.0	13.9	18.9	24.2	23.3	15.4	16.5	15.2	
66	324	8.6	0	0	356	469	24.0	21.0	22.6	25.6	24.9	17.0	21.0	15.1	
67	324	8.5	0	0	356	469	25.0	20.7	22.3	25.3	24.6	16.8	20.7	14.9	
68	323	8.6	61	2.7	356	469	25.0	14.3	18.0	23.9	22.9	15.0	16.0	15.2	
69	322	8.3	0	0	356	469	22.0	20.4	21.9	24.8	24.2	16.5	20.4	14.7	
70	324	8.7	0	0	356	469	24.0	21.2	22.8	25.9	25.2	17.2	21.2	15.3	
71	324	8.4	0	0	356	469	23.0	20.5	22.0	25.0	24.3	16.6	20.5	14.8	

(continued on next page)

**Table A1 (continued)**

Sl.	D mm	t mm	L mm	d mm	YS MPa	UTS MPa	Exp. MPa	B31G MPa	M.B31G MPa	DNV MPa	Battelle MPa	Shell MPa	Netto MPa	Pop MPa	Ref.
72	324	8.6	127	2.7	356	469	22.0	8.5	17.9	21.6	21.2	13.4	14.2	15.1	
73	323	8.5	51	2.2	356	469	22.0	15.4	18.7	24.3	23.3	15.5	16.6	15.0	
74	324	8.6	0	0	356	469	25.0	21.0	22.6	25.6	24.9	17.0	21.0	15.1	
75	864	9.6	213	3.6	400	508	11.0	4.1	7.8	9.2	9.3	6.0	6.9	7.1	
76	864	9.5	185	3	400	508	11.0	4.6	8.2	9.8	9.7	6.4	7.3	7.0	
77	273	8.3	241	4	409	481	21.0	14.3	19.1	18.6	18.1	12.9	11.6	19.9	
78	273	5.3	0	0	389	502	17.0	16.8	17.8	19.9	19.5	13.6	16.8	12.1	
79	613	6.4	1433	3.6	403	535	8.0	4.1	5.3	5.2	4.9	3.4	0.6	6.7	
80	612	6.4	1372	2.6	403	535	10.0	5.6	6.6	6.9	6.7	4.6	2.3	6.7	
81	507	5.7	132	3	462	587	11.0	4.5	7.6	9.2	9.4	5.7	7.2	8.3	
82	505	5.7	462	3.3	462	587	8.0	4.9	6.7	6.4	6.0	4.3	4.0	8.3	
83	508	5.7	620	3.8	462	587	9.0	3.8	5.6	4.9	4.5	3.3	2.3	8.3	
84	508	5.6	597	3.4	462	587	8.0	4.4	6.1	5.7	5.2	3.8	2.9	8.1	
85	508	5.6	170	2.5	462	587	12.0	3.5	8.1	9.3	9.4	6.0	7.1	8.1	
86	324	9.8	256	7.1	452	542	14.0	8.4	14.3	12.7	12.5	8.2	8.9	21.9	
87	324	9.7	395	6.9	452	542	13.0	8.7	14.1	11.7	10.6	7.9	5.0	21.7	
88	324	9.9	433	7.3	452	542	12.0	8.1	13.5	10.8	9.6	7.3	3.6	22.1	
89	324	9.7	467	7	452	542	12.0	8.4	13.6	11.0	9.7	7.5	3.1	21.7	
90	324	9.8	489	7	452	542	12.0	8.7	13.9	11.3	10.0	7.7	2.8	21.9	
91	324	9.8	500	7	452	542	12.0	8.7	13.8	11.3	10.0	7.7	2.5	21.9	
92	324	9.7	528	7.1	452	542	11.0	8.1	13.1	10.4	9.1	7.1	1.5	21.7	

\*D: Diameter; t: thickness; L: length of corrosion defect; d: depth of corrosion defect; YS: Yield Strength; UTS: Ultimate Tensile Strength; Popt: Operating pressure.

algorithm that can be applied to determine the failure risk of pipelines. Also, analyses revealed that machine learning algorithms' computational efficiency is significantly faster than the physics-based model. Hence, machine learning algorithms can be used as a viable alternative to computationally demanding analytical methods.

The current study has some limitations that need to be considered in future research. The burst failure models only account for external corrosion, and the effect of internal corrosion was not considered in this study. Due to insufficient data available on real-time burst failure pressures, existing literature relies heavily on experimental burst test results while evaluating existing burst failure models [2,4,25,27]. The current study also relied on the full-scale burst test results published in the literature where each experimental test has some constraints. However, regardless of the difference between the experimental and actual data, the current study hypothesizes that machine learning algorithms can still be applied to predict pipelines' failure prediction. In the future, the current application can be further updated with actual failure data.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This research was supported, in part, by the National Science Foundation (NSF) Critical Resilient Interdependent Infrastructure Systems and Processes (CRISP) under Grant No. NSF- 1638320. This support is thankfully acknowledged. The authors also acknowledge Cleveland Water Department for their support of this project. However, the writers take sole responsibility for the views expressed in this paper, which may not represent the position of the NSF or their respective institutions.

#### Appendix A1

#### References

- [1] Ahammed M. Prediction of remaining strength of corroded pressurised pipelines. *Int J Press Vessels Pip* 1997;71(3):213–7.
- [2] Al-Owaisi S, Becker AA, Sun W, Al-Shabibi A, Al-Maharbi M, Pervez T, et al. An experimental investigation of the effect of defect shape and orientation on the burst pressure of pressurised pipes. *Eng Fail Anal* 2018;93:200–13.
- [3] Aljaroudi A, Khan F, Akinturk A, Haddara M, Thodi P. Risk-based assessment of offshore crude oil pipelines and condition-monitoring systems. *J Pipeline Eng* 2016;15(1).
- [4] Amaya-Gómez R, Sánchez-Silva M, Bastidas-Arteaga E, Schoefs F, Munoz F. Reliability assessments of corroded pipelines based on internal pressure—A review. *Eng Fail Anal* 2019;98:190–214.
- [5] Ang AHS, Tang WH. Probability concepts in engineering planning and design: Emphasis on application to civil and environmental engineering. Wiley; 2007.
- [6] Anghel CI. Risk assessment for pipelines with active defects based on artificial intelligence methods. *Int J Press Vessels Pip* 2009;86(7):403–11.
- [7] ASME B31 Committee. ASME B31G-2009: manual for determining the remaining strength of corroded pipelines. Am Soc Mech Eng 2009.
- [8] Batt AD, Fu B, Kirkwood MG, Vu D. New methods for determining the remaining strength of corroded pipelines. In: Proceedings of the international conference on offshore mechanics and arctic engineering. American Society of Mechanical Engineers; 1997. p. 221–8.
- [9] Beyer K, Goldstein J, Ramakrishnan R, Shaft U. When is “nearest neighbor” meaningful?. In: International conference on database theory. Berlin, Heidelberg: Springer; 1999. p. 217–35.
- [10] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC Press; 1984.
- [11] Breiman L. Random forests. *Machine Learn* 2001;45(1):5–32.
- [12] Boxall, J. B., O'Hagan, A., Pooladsaz, S., Saul, A. J., & Unwin, D. M. (2007). Estimation of burst rates in water distribution mains. In Proceedings of the Institution of Civil Engineers-Water Management (Vol. 160, No. 2, pp. 73–82). Thomas Telford Ltd.
- [13] Caleylo F, Gonzalez JL, Hallen JM. A study on the reliability assessment methodology for pipelines with active corrosion defects. *Int J Press Vessels Pip* 2002;79(1):77–86.
- [14] Caleylo F, Velázquez JC, Valor A, Hallen JM. Probability distribution of pitting corrosion depth and rate in underground pipelines: A Monte Carlo study. *Corros Sci* 2009;51(9):1925–34.
- [15] Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Machine Learn Res* 2010;11:2079–107.
- [16] CSA (Canadian Standards Association) (1999). CSA Z662-99, Oil and Gas Pipeline Systems. Ontario, Canada.
- [17] Chen JM, Meng HM, Li YY. Pipeline prescription analysis after corrosion and explosive test. *Oil Gas Storage Transp* 1998;17(3):28–30.
- [18] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785–794).
- [19] Cosham A, Hopkins P, Macdonald KA. Best practice for the assessment of defects in pipelines—Corrosion. *Eng Fail Anal* 2007;14(7):1245–65.
- [20] Cronin DS, Pich RJ. Experimental database for corroded pipe: evaluation of RSTRENG and B31G. 2000 3rd international pipeline conference. American society of mechanical engineers digital collection. 2000.
- [21] DNV (Det Norske Veritas) (1999). Corroded Pipelines: DNV Recommended Practice RP-F101, 1999. Det Norske Veritas.

- [22] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.
- [23] Freire JLF, Vieira RD, Castro JTP, Benjamin AC. Part 3: Burst tests of pipeline with extensive longitudinal metal loss. *Exp Tech* 2006;30(6):60–5.
- [24] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. New York: Springer Series in Statistics; 2001. Vol. 1, No. 10.
- [25] Gao J, Yang P, Li X, Zhou J, Liu J. Analytical prediction of failure pressure for pipeline with long corrosion defect. *Ocean Eng* 2019;191:106497.
- [26] Ghosh J, Padgett JE, Dueñas-Osorio L. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probab Eng Mech* 2013;34:189–99.
- [27] Hasan S, Khan F, Kenny S. Probability assessment of burst limit state due to internal corrosion. *Int J Press Vessels Pip* 2012;89:48–58.
- [28] Huang H, Burton HV. Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. *J Build Eng* 2019;25:100767.
- [29] Jeon JS, Shafeezadeh A, DesRoches R. Statistical models for shear strength of RC beam-column joints using machine-learning techniques. *Earthquake Eng Struct Dyn* 2014;43(14):2075–95.
- [30] Jin WL, Zhang EY, Shao JW, Liu DH. Cause analysis and countermeasure for submarine pipeline failure. *Bull Sci Technol* 2004;6.
- [31] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W. Lightgbm: A highly efficient gradient boosting decision tree. In: Advances in neural information processing systems; 2017. p. 3146–54.
- [32] Kiani J, Camp C, Pezeshk S. On the application of machine learning techniques to derive seismic fragility curves. *Comput Struct* 2019;218:108–22.
- [33] Kiefner JF, Vieth PH. Evaluating pipe-1. new method corrects criterion for evaluating corroded pipe. *Oil Gas J* 1990;88(32).
- [34] Kim YP, Baek JH, Kim US, Go YT. The evaluation of burst pressure for corroded pipeline by full scale burst test. *Trans Korean Soc Mech Eng A* 2004;26(1):203–10.
- [35] Kim, D. W., Mohd, M. H., Lee, B. J., Kim, D. K., Seo, J. K., Kim, B. J., & Paik, J. K. (2013). Investigation on the burst strength capacity of aging subsea gas pipeline. In International Conference on Offshore Mechanics and Arctic Engineering (Vol. 55362, p. V04AT04A027). American Society of Mechanical Engineers.
- [36] Klever FJ, Stewart G, van der Valk CA. New developments in burst strength predictions for locally corroded pipelines (No. CONF-950695-). New York, NY (United States): American Society of Mechanical Engineers; 1995.
- [37] Leis BN, Stephens DR. An alternative approach to assess the integrity of corroded line pipe-part I: current status. The seventh international offshore and polar engineering conference. International Society of Offshore and Polar Engineers; 1997.
- [38] Luque A, Carrasco A, Martín A, de las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recogn* 2019;91:216–31.
- [39] Ma B, Shuai J, Xu X. A study on new edition assessment criteria for the remaining strength of corroded pipeline. In: ICPTT 2011: Sustainable solutions for water, sewer, gas, and oil pipelines; 2011. p. 63–72.
- [40] Mangalathu S, Jeon JS. Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. *Eng Struct* 2018;160:85–94.
- [41] Mangalathu S, Jeon JS. Machine learning-based failure mode recognition of circular reinforced concrete bridge columns: Comparative study. *J Struct Eng* 2019;145(10):04019104.
- [42] Mangalathu S, Hwang SH, Jeon JS. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng Struct* 2020;219:110927.
- [43] Mannucci, G, Harris, D. Fracture properties of API X 100 gas pipeline steels. Directorate-General for Research; 2002.
- [44] Melchers RE. Progression of pitting corrosion and structural reliability of welded steel pipelines. *Oil Gas Pipelines* 2015:327–42.
- [45] Mok DHB, Pick RJ, Glover AG, Hoff R. Bursting of line pipe with long external corrosion. *Int J Press Vessels Pip* 1991;46(2):195–216.
- [46] Netto TA, Ferraz US, Estefen SF. The effect of corrosion defects on the burst pressure of pipelines. *J Constr Steel Res* 2005;61(8):1185–204.
- [47] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Machine Learn Res* 2011;12:2825–30.
- [48] Qi F, Zhu AX. Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps. *Comput Geosci* 2011;37(9):1425–36.
- [49] Robles-Velasco A, Cortés P, Muñozuri J, Onieva L. Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliab Eng Syst Saf* 2020;196:106754.
- [50] Siam, A., Ezzeldin, M., & El-Dakhakhni, W. (2019, December). Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls. In Structures (Vol. 22, pp. 252–265). Elsevier.
- [51] Shuai Y, Shuai J, Liu CY. Research on the reliability methods of corroded pipeline. *Petrol Sci Bull* 2017;2(2):288–97.
- [52] Teixeira AP, Soares CG, Netto TA, Estefen SF. Reliability of pipelines with corrosion defects. *Int J Press Vessels Pip* 2008;85(4):228–37.
- [53] Villarin MC, Rodriguez-Galiano VF. Machine learning for modeling water demand. *J Water Resour Plann Manage* 2019;145(5):04019017.
- [55] Wang N, Zarghami MS. Evaluating fitness-for-service of corroded metal pipelines: structural reliability bases. *J Pipeline Syst Eng Pract* 2014;5(1):04013012.
- [56] Wang Z, Pedroni N, Zentner I, Zio E. Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment. *Eng Struct* 2018;162:213–25.
- [57] Winkler D, Haltmeier M, Kleidörfer M, Rauch W, Tscheikner-Grati F. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct Infrastruct Eng* 2018;14(10):1402–11.
- [58] Ying C, Qi-Guang M, Jia-Chen L, Lin G. Advance and prospects of AdaBoost algorithm. *Acta Autom Sin* 2013;39(6):745–58.
- [59] Yu L, Porwal A, Holden EJ, Dentith MC. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput Geosci* 2012;45:229–39.
- [60] Zakikhani K, Nasiri F, Zayed T. A review of failure prediction models for oil and gas pipelines. *J Pipeline Syst Eng Pract* 2020;11(1):03119001.
- [61] Zhang Y, Burton HV, Sun H, Shokrabbadi M. A machine learning framework for assessing post-earthquake structural safety. *Struct Saf* 2018;72:1–16.
- [62] Mazumder Ram K, Salman AM, Li Y, Yu X. Reliability analysis of water distribution systems using physical probabilistic pipe failure method. *J Water Res Plan Manage* 2019;145(2):04018097. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001034](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001034). In this issue.