ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/345643171

Failure Risk Analysis of Pipelines using Data-Driven Machine Learning Algorithms

Article *in* Structural Safety · November 2020

DOI: 10.1016/j.strusafe.2020.102047

CITATIONS
103

READS
2,027

3 authors, including:

Abdullahi M. Salman
University of Alabama in Huntsville
57 PUBLICATIONS **1,189** CITATIONS
SEE PROFILE

Yue Li
Case Western Reserve University
175 PUBLICATIONS **4,585** CITATIONS
SEE PROFILE

All content following this page was uploaded by Ram Krishna Mazumder on 09 November 2020.

The user has requested enhancement of the downloaded file.

Failure Risk Analysis of Pipelines using Data-Driven Machine Learning Algorithms

Ram K. Mazumder[1], Abdullahi M. Salman[2] and Yue Li[3*]

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.

Uncorrected Proof

[1] Research Assistant, Department of Civil Engineering, Case Western Reserve University, Cleveland, Ohio, rxm562@case.edu
[2] Assistant Professor, Department of Civil & Environmental Engineering, The University of Alabama in Huntsville, ams0098@uah.edu
[3] Professor, Department of Civil Engineering, Case Western Reserve University, Cleveland, Ohio, yxl1566@case.edu
* Corresponding Author.

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 (Teixeira et al. 2008). 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 (Boxall et al. 2007; Gao et al. 2019). The reduction of material strength due to corrosion pit growth on the pipe wall increases the possibility of leakage and burst failures (Jin et al. 2004). 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 (Gao et al. 2019). A number of methods have been developed in the past to estimate the burst failure pressure of pipelines subjected to corrosion defects (e.g., ASME B31G 2009; DNV 1999; Kiefner and Vieth 1990; Batte et al. 1997; Netto et al. 2005; Wang and Zarghamee 2014). These methods mostly used semi-empirical fracture mechanics relationships where failure depends on flow stress (Gao et al. 2019). 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 (Cronin and Pick 2000).

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

Uncorrected Proof

55 cost-efficient, higher quality, and sustainable approaches. Since existing approaches to predict
56 pipe failure or risk ratings are not computationally efficient in handling large and complex
57 pipeline networks, a more robust technique is required (Robles-Velasco et al. 2020). Therefore
58 there is a need to apply a new computationally efficient approach that is capable of determining
59 the risk of each individual segment of a complex networked system without explicitly
60 performing intensive physically-based analysis. The application of data-driven machine-
61 learning approaches may provide a viable alternative to computationally demanding
62 probabilistic physical analysis for pipeline failure risk analysis of oil and gas pipeline networks.
63 The application of machine learning in risk and reliability assessment has gained
64 considerable attention from decision-makers and scientific community in recent years. The
65 machine learning approach can be described as a semi-automated system in which computers
66 learn from the observed data to develop an algorithm. It has been recently studied and applied
67 in civil engineering risk analysis (e.g., Ghosh et al. 2013; Jeon et al. 2014; Mangalathu and
68 Jeon 2018, 2019; Mangalathu et al. 2020; Wang et al. 2018; Zhang et al. 2018). Machine
69 learning is a new dimension of research under active development. A number of algorithms,
70 such as support vector machine, k-nearest neighbours, random forest, have emerged in the last
71 decade (Villarin et al. 2019).
72 Anghel (2009) developed a novel type of support vector machine in a maximum manner to
73 predict the failure probability of pipes with active corrosion defects. The author developed a
74 classification reliability procedure that established a link between artificial intelligence and
75 reliability methods. This approach provides a simple but promising alternative for prioritizing
76 maintenance actions. Huang and Burton (2019) utilized machine learning applications to
77 classify the failure modes of reinforced concrete frames with infill walls. Winkler et al. (2018)
78 predicted water pipe failures using decision-tree based machine learning algorithms. Bootstrap
79 aggregation and boosting techniques are utilized to improve the accuracy of the model. The

Uncorrected Proof

performance of the model is evaluated using a confusion matrix. Mangalathu and Jeon (2019) 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. (2019) explored the application of machine learning algorithms for the derivation of fragility curves. Siam et al. (2019) used a clustering algorithm to classify masonry walls utilizing a very limited number of datasets.

Mangalathu et al. (2020) 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 (2011), Yu et al. (2012), Jeon et al. (2014), and Zhang et al. (2018). 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

Uncorrected Proof

105 then used to assess the feasibility of machine learning algorithms for evaluating failure risk of
106 pipelines, (2) to evaluate the effectiveness of machine learning algorithms in the classification
107 of pipelines based on their failure probability, and (3) identify the contribution of each
108 parameter to the pipeline failure probability through a sensitivity analysis.
109
110 **2. Failure Prediction Model and Data Generation**
111 A flowchart illustrating the proposed methodology is shown in Figure 1. A database of
112 burst failure tests is compiled from available literature, as shown in Appendix A1. Burst failure
113 pressure for pipelines subjected to corrosion defects is estimated using various standards (e.g.,
114 American Society of Mechanical Engineers (ASME) B31G, Det Norske Veritas (DNV) RP-
115 F101). The experimental burst failure pressure obtained from the tests is compared with the
116 predicted pressure obtained using various standards. The difference between the predicted
117 failure pressure and the experimental failure pressure is analysed to obtain a mean error and
118 statistical distribution of the variability in prediction. Then, the true failure pressure is defined
119 by the sum of predicted pressure, mean error, and statistical distribution of variability in
120 prediction with zero mean and standard deviation σ. The pipeline's failure probability is
121 determined by comparing the true failure pressure and operating fluid pressure of pipe at a
122 burst limit state. The burst failure probability of the pipeline is determined in a probabilistic
123 manner utilizing Monte Carlo simulation. The pipeline is classified into different failure risk
124 classes based on the failure probability of the pipeline. Utilizing this concept, a new dataset of
125 pipelines is generated. Various machine learning algorithms are applied to the developed
126 datasets to identify the best predictive machine learning algorithm.
127 A detailed description of the different parts of the framework is given in the subsections
128 below.

Uncorrected Proof

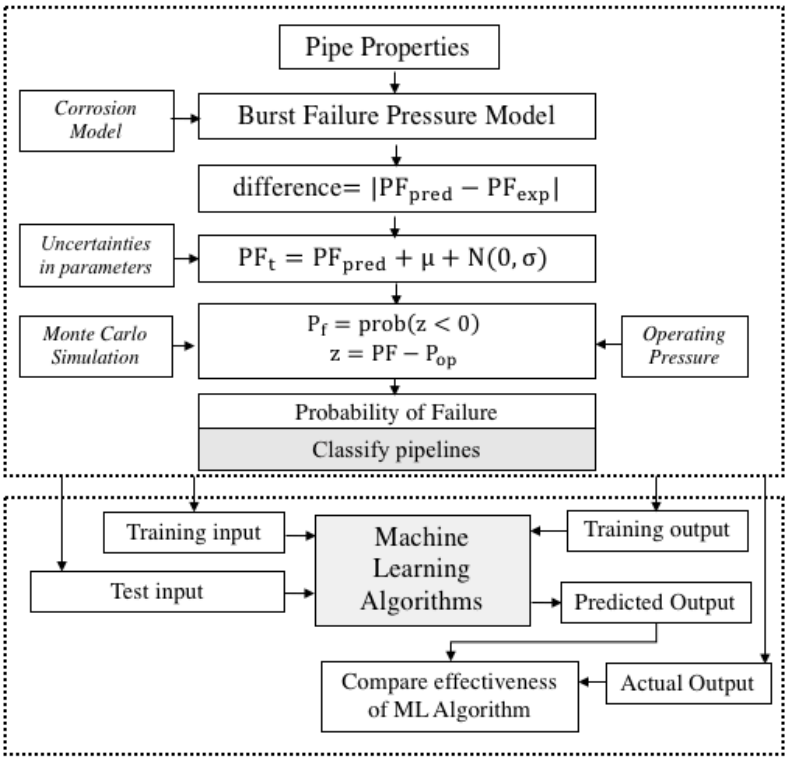

Figure 1: Proposed Framework

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 (Kiefner and Vieth 1990; Klever and Stewart 1995; Leis and Stephens 1997; Batte et al. 1999; DNV 1999; Netto et al. 2005; Wang and Zarghamee 2014). 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 model in comparison to older approaches. More recently, researchers performed experimental and

Uncorrected pdf

numerical analysis and found various standards use safety factors that often lead to costly and unnecessary maintenance of pipelines (Netto et al. 2005; Wang and Zarghamee 2014).

Table 1: Failure pressure models

Sl.	Model Name	Burst Failure Pressure, FP
1	B31G *(ASME-B31G 1991)*	$FP = \begin{cases} \frac{1.11 \cdot 2YSt}{D}\left(\frac{1-\frac{2d(T)}{3t}}{1-\frac{2d(T)}{3t}M^{-1}}\right); G < 4 \\ \frac{1.11 \cdot 2YSt}{D}\left(1-\frac{d(T)}{t}\right); G \geq 4 \end{cases}$ $M = \sqrt{1+0.893\frac{L(T)^2}{Dt}}$; $G = 0.893\frac{L(T)^2}{\sqrt{Dt}}$
2	Modified B31G *(Kiefner and Vieth 1990)*	$FP = \frac{2.0\,(YS + 68.95\ MPa)t}{D}\left(\frac{1-0.85\frac{d(T)}{t}}{1-0.85\frac{d(T)}{t}M^{-1}}\right)$ $M = \begin{cases} \sqrt{1+0.6275\frac{L(T)^2}{Dt}-0.003375\frac{L(T)^4}{D^2t^2}} & \text{if} \frac{L^2}{Dt} \leq 50 \\ 0.032\frac{L(T)^2}{Dt}+3.3 & \text{if } \frac{L^2}{Dt} \leq 50 \end{cases}$
3	DNV-RP-F101 *(DNV 1999)*	$FP = \frac{2\ UTS}{D-t}\left(\frac{1-\frac{d(T)}{t}}{1-\frac{d(T)}{t}M^{-1}}\right)$ $M = \sqrt{1+0.31\frac{L(T)^2}{Dt}}$
4	Shell-92 *(Klever and Stewart 1995)*	$FP = \frac{1.8\ UTS}{D}\left(\frac{1-\frac{d(T)}{t}}{1-\frac{d(T)}{t}M^{-1}}\right)$ $M = \sqrt{1+0.805\frac{L(T)^2}{Dt}}$
5	Netto et al. (2005)	$FP = \frac{1.1 \cdot \sigma_y \cdot 2t}{D} \cdot \left[1-0.9435\left(\frac{d}{t}\right)^{1.6}\left(\frac{l}{D}\right)^{0.4}\right]$
6	Battelle *(Leis and Stephens 1997)*	$FP = \frac{2UTSt}{D}\left(1-\frac{d(T)}{t}M\right)$ $M = 1-\exp\left(-0.157\frac{L(T)}{\sqrt{D(t-d(T))/2}}\right)$

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.

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 (Ma et al. 2011; Kim et al. 2013). Hence, consideration of uncertainties in a corrosion model could provide a realistic scenario (Aljaroudi et al. 2014; Zakikhani et al. 2020). 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 (Melchers 2015). 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 (Caleyo et al. 2009; Wang and Zarghamee 2014;

Uncorrected Proof

161 Melchers 2015) and the length of corrosion pit is modeled as a function of pipe age, as
162 expressed below (Ahammed 1997; Caleyo et al. 2002):

163 $$d(T) = k(T - T_d)^{\alpha} \quad (1)$$

164 $$L(T) = L_0 + V_a(T - T_i) \quad (2)$$

165 where d(T) and L(T) are depth and length of corrosion pit, respectively; k and α are
166 corrosion growth coefficients; T_i is the time of the last inspection; V_a is the axial corrosion
167 growth rate. Linear rate is attained over the pipe elapsed life ($V_a = L_0/\Delta T$, where ΔT is the
168 pipeline elapsed time to inspection date, L_0 is the initial length). Pipeline elapsed time of 15
169 years is assumed reasonable to guarantee that the growth rate comes to a steady state. The depth
170 of the corrosion pit is simulated using Monte Carlo Simulation. Coefficients of corrosion are
171 assumed to be normally distributed. The mean and coefficient of variation for κ and α are taken
172 as (0.164/yr, 0.20) and (0.780, 0.15), respectively. Since the soil types are unknown, these
173 values κ and α are taken for average soil type from Caleyo et al. (2009). The time required to
174 initiate corrosion was assumed equal to 2.88 years (Caleyo et al. 2009; Wang and Zarghamee
175 2014).

176 2.2 Predicted and Experimental Burst Failure Pressure

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

181 $$\Delta_p = FP_{pred} - FP_{exp} \quad (3)$$

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

Uncorrected Proof

185 Expected true burst failure pressure of a pipeline, FP_t, can be estimated by the sum of
186 predicted pressure, mean error in the prediction and statistical distribution of variability in
187 prediction (Cronin and Pick 2000), as

188 $$FP_t = FP_{pred} + \mu + N(0, \sigma^2) \quad (4)$$

189 where μ is the mean difference between prediction and experimental pressures, and N
190 represents variability of prediction of a normal distribution with zero mean and variance σ^2.
191 Equation (4) is applied to determine the expected true failure pressure of pipelines.
192 Consideration of variability between predicted and experimental results in Equation (4) reduces
193 the conservativeness in true failure pressure prediction.

194 2.3 Burst Failure Limit State

195 The burst failure probability of pipeline is determined by comparing the true burst failure
196 pressure and operating pressure inside a pipeline (Caleyo et al., 2002; Wang and Zarghamee,
197 2013). The pipe's failure occurs when applied internal pressure exceeds the true failure pressure
198 of a pipeline. The limit state is defined as:

199 $$g(X) = FP_t - P_{opt} \quad (5)$$

200 where FP_t is the pipeline true burst failure pressure on the pipeline, P_{opt} is the internal fluid
201 pressure of the pipe, $g(X)$ is the limit state of the pipeline. Allowable operating pressure inside
202 a pipeline is estimated by the following equation (CSA 1999; Cronin and Pick, 2000):

203 $$P_{opt} = \frac{2t}{D} YS \times F \times L \times J \times T \quad (6)$$

204 where F is the design safety factor (0.8); L is the location factor; J is the joint factor; T is
205 the temperature factor equal to 1.0 (CSA 1999). The failure of the pipeline is defined by the
206 failure likelihood that internal pressure exceeds the true burst strength of the pipe (positive
207 values indicate safe operation and a negative value indicates failure in this case). The

Uncorrected Proof

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 Caleyo et al. (2002) and Wang and Zarghamee (2014) and shown in Table 2.

Table 2: Statistical distribution of random variables

Parameters	**Mean**	**COV**	**Unit**	**Distribution Type**
Pipe diameter, D	-	-	mm	Deterministic
Wall thickness, t	-$_t$	-	mm	Deterministic
Defect depth, d	μ_d	0.10	mm	Normal
Defect length, L	μ_L	0.10	mm	Normal
Yield Strength (YS)	μ_{YS}	0.07	MPa	Lognormal
Ultimate Tensile Strength (UTS)	μ_{UTS}	0.07	MPa	Lognormal
Operating pressure, P_{opt}	μ_{Popt}	0.10	MPa	Normal

Note: COV: Coefficient of Variation

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 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 Figure 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.

Uncorrected Proof

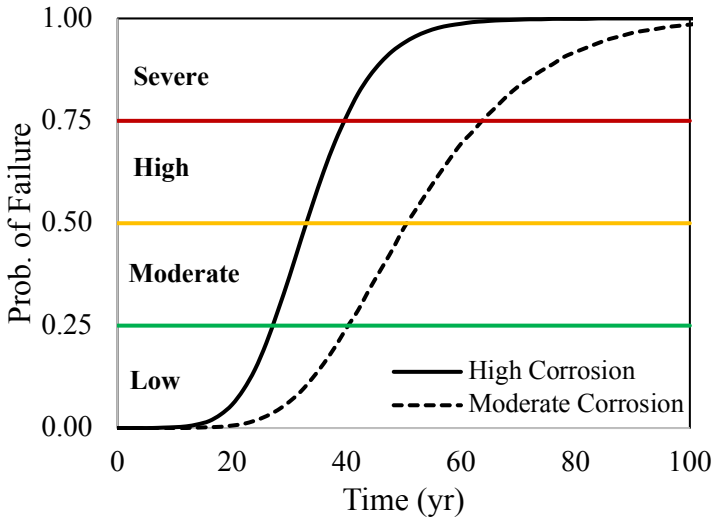

Figure 2: Probability of failure levels

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) are the most basic and useful algorithms used in the industry for classification problems (Beyer et al. 1999). 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 can be estimated below, $P_k(x)$ (Mangalathu and Jeon 2019):

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

where N_k is the number of data points.

Uncorrected Proof

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 (Villarin et al. 2019; Mangalathu et al. 2020). The dataset is split into 's' distinct and overlapping regions. The binary splitting is performed based on the Gini Index (GI) expressed as below (Breiman et al. 1984):

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

where GI represents total variance across the k-classes, $\hat{P}_{mk}$ 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 (Breiman 2001). 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 (Mangalathu and Jeon 2019). The random forest regression predictor is expressed as below (Villarin et al. 2019):

Uncorrected Proof

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

where $\hat{P}_{rf}$ 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 (Pedregosa et al. 2011). 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 (Mangalathu and Jeon 2019):

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

where π_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 (LGBoost), 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 classifiers than the weight assigned for a weak classifier. 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 (Friedman et al. 2001; Mangalathu et al. 2000). XGBoost tries to reduce the misclassification error in the previous model. Since this model uses sequential training steps, it is relatively slower to implement. LightGBM uses a decision tree algorithm where the model implication

Uncorrected Proof

happens leafs-wise. The CatBoost handles categorical features by taking the input of features names (Mangalathu et al. 2020). A detail of these boosting algorithms can be found elsewhere (Ying et al. 2013; Chen et al. 2016; Ke et al. 2017; Dorogush et al. 2018).

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 (Robles-Velasco et al. 2020; Mangalathu and Jeon 2019; Mangalathu et al. 2020). 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 Figure 3 (Luque et al. 2019; Robles-Velasco et al. 2020).

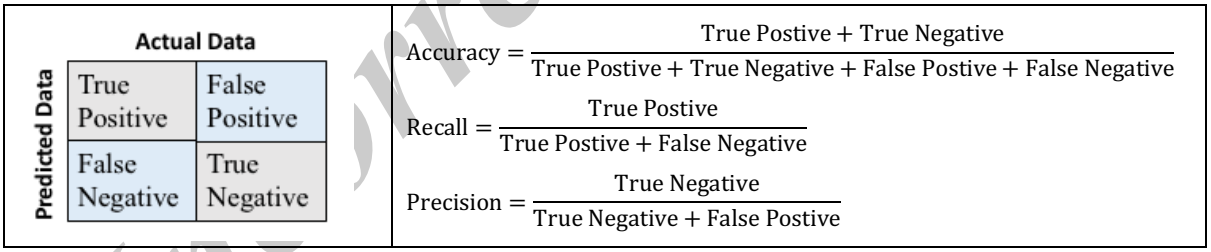

Figure 3: Confusion matrix

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 competitional efficiency to execute the analysis.

Uncorrected Proof

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. (2006) performed 14 full-scale laboratory tests for steel pipelines with corrosion defects to understand the internal pressure failure mechanism. Mok et al. (1991) performed a burst behavior of 5 steel pipelines with long external corrosion. Similarly, Chen et al. (1998), Mannucci and Harris (2002), Kim et al. (2004), and Shuai et al. (2017) performed burst failure tests for pipelines with external corrosion defects. Cronin and Pick (2000) 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 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:

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

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. (2005) 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 (Cosham et al. 2007; Hasan et al. 2012; Amaya-Gómez et al. 2019). 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 Equation (5).

Uncorrected Proof

337 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	-

338 Note: Stdev: Standard Deviation

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

348 The size of the corrosion pit depends largely on the pipe age and corrosion growth rate. In
349 the field, the corrosion pit's size can be estimated using a smart measuring device (e.g., smart
350 pig). Depth and length of corrosion pit are predicted using Equations (1) & (2), respectively.
351 The depth and length of corrosion gradually increased over time. Length of corrosion defect
352 (L0) at the inspection was assumed to be equal to 50-200 mm (Calyeo et al. 2002). Figures 4(e)
353 and 4(f) show the summary of depth and length of corrosion pit, respectively. The operating
354 internal fluid pressure was estimated based on the CSA (1999) guideline expressed by Equation
355 (6), as shown in Figure 4(g). The probability of failure of each pipeline is estimated using a
356 probabilistic approach by comparing the burst failure strength with the internal operating
357 pressure, as expressed by the burst limit state in Equation (5). Monte Carlo simulation is

Uncorrected Proof

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 Figure 2. Figure 4(h) shows a summary of pipeline failure risk classes obtained from the simulation.

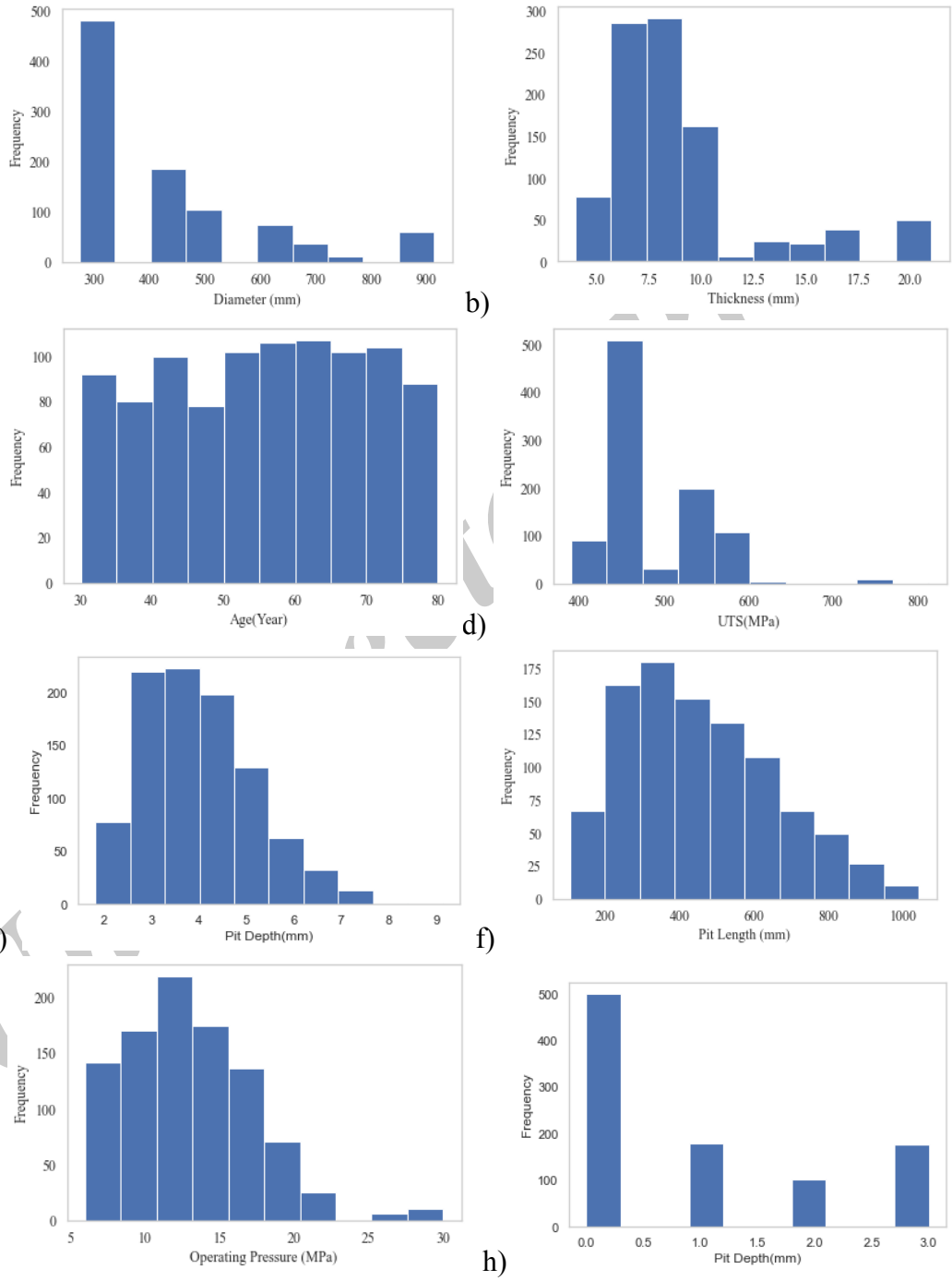

Figure 4: a) Diameter; b) thickness; c) age; d) ultimate yield strength; e) depth of corrosion; f) length of corrosion; g) operating pressure; and h) failure risk class

A correlation matrix is developed to learn about the relationship between various input pairs for failure probability classification, as shown in Figure 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.

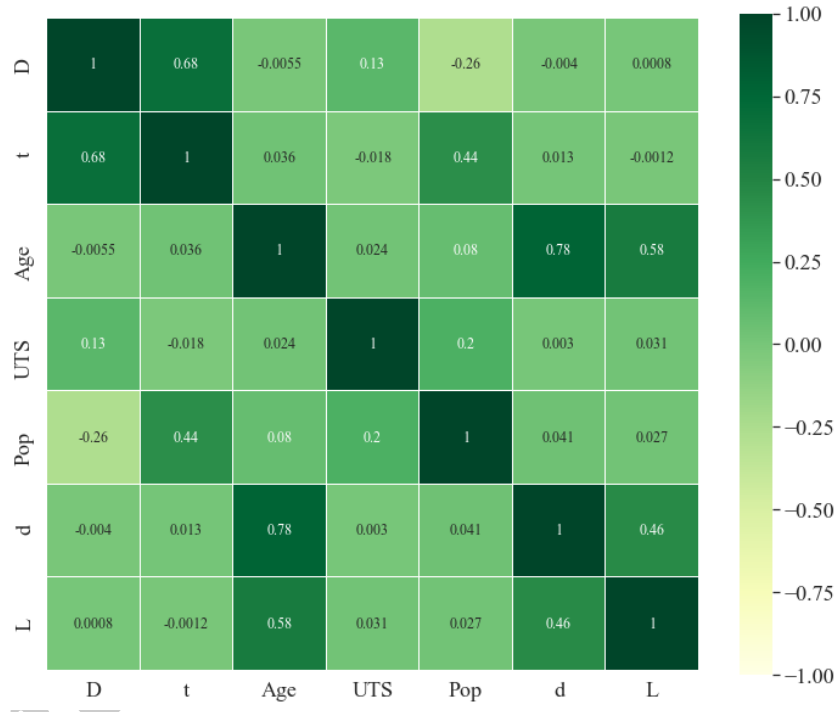

Figure 5: Correlation matrix

Once pipelines data is generated and classified into various risk classes, as shown in Figure 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 (Mangalathu et al., 2020). 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.

Under Proof

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 Figure 6. A general confusion matrix represents four prediction terminologies: True Positive, True Negative, False Positive, and False Negative, as shown in Figure 3. True positive and true negative represents 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 (Ang and Tang 2007). 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 (Luque et al. 2019). Figures 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. Figure 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.

Uncorrected Proof

(a): K-Nearest Neighbours

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(b): Decision Tree

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(c): Random Forest

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(d): Naïve Bayes

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(e): Ada Boost

Observed Class \ Predicted Class	L	M	H	S	Recall
L	42.6%	9.5%	0.0%	0.1%	81.0%
M	0.3%	16.2%	1.3%	0.0%	91.0%
H	0.0%	1.0%	8.3%	0.3%	86.0%
S	0.1%	0.0%	8.6%	11.3%	56.0%
Precision	99.0%	61.0%	46.0%	96.0%	79.0%

(f): XGBoost

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(g): LGBoost

Observed Class \ Predicted Class	L	M	H	S	Recall
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%

(h): CatBoost

Figure 6: Confusion matrices for training sets

Figure 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

Uncorrected Proof

problems (Cawley and Talbot 2010). 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.

		Predicted Class				
		L	M	H	S	Recall
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%

(a): K-Nearest Neighbours

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		88.0%	59.0%	55.0%	77.0%	77.0%

(b): Decision Tree

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		91.0%	68.0 %	75.0 %	77.0 %	82.0%

(c): Random Forest

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		83.0%	53.0%	53.0%	77.0%	74.0%

(d): Naïve Bayes

		Predicted Class				
		L	M	H	S	Recall
Observed Class	L	17.6%	4.6%	0.0%	0.0%	79.0%
	M	0.1%	7.7%	0.9%	0.0%	88.0%
	H	0.0%	1.2%	4.0%	0.3%	73.0%
	S	0.0%	0.0%	3.9%	2.5%	40.0%
Precisio n		99.0%	57.0%	46.0%	89.0%	74.0%

(e): Ada Boost

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		95.0%	71.0%	71.0%	82.0%	85.0%

(f): XGBoost

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		95.0%	67.0 %	74.0 %	83.0 %	84.0%

(g): LGBoost

		Predicted Class				
		L	M	H	S	Recall
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%
Precisio n		95.0%	69.0%	75.0%	87.0%	86.0%

(h): CatBoost

Figure 7: Confusion matrices for test sets

Uncorrected Proof

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, Ada Boost, 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 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

Uncorrected Proof

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 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 (Gao et al. 2011; Hasan et al. 2012; Al-Owaisi et al. 2018; Amaya-Gómez et al. 2019). 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.

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.

References

1. Ahammed, M. (1997). Prediction of remaining strength of corroded pressurised pipelines. International journal of pressure vessels and piping, 71(3), 213-217.
2. Al-Owaisi, S., Becker, A. A., Sun, W., Al-Shabibi, A., Al-Maharbi, M., Pervez, T., & Al-Salmi, H. (2018). An experimental investigation of the effect of defect shape and orientation on the burst pressure of pressurised pipes. *Engineering Failure Analysis*, *93*, 200-213
3. Aljaroudi, A., Khan, F., Akinturk, A., Haddara, M., & Thodi, P. (2016). Risk-based assessment of offshore crude oil pipelines and condition-monitoring systems. *Journal of Pipeline Engineering*, *15*(1).
4. Amaya-Gómez, R., Sánchez-Silva, M., Bastidas-Arteaga, E., Schoefs, F., & Munoz, F. (2019). Reliability assessments of corroded pipelines based on internal pressure–A review. *Engineering Failure Analysis*, *98*, 190-214.
5. Ang, A. H. S., & Tang, W. H. (2007). Probability concepts in engineering planning and design: Emphasis on application to civil and environmental engineering. Wiley.
6. Anghel, C. I. (2009). Risk assessment for pipelines with active defects based on artificial intelligence methods. International journal of pressure vessels and piping, 86(7), 403-411.
7. ASME B31 Committee. (2009). ASME B31G-2009: manual for determining the remaining strength of corroded pipelines. In Am Soc Mech Eng.
8. Batte, A. D., Fu, B., Kirkwood, M. G., & Vu, D. (1997). New methods for determining the remaining strength of corroded pipelines. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering (pp. 221-228). American Society of Mechanical Engineers.
9. Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999, January). When is "nearest neighbor" meaningful?. In International conference on database theory (pp. 217-235). Springer, Berlin, Heidelberg.
10. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
11. Breiman, L. (2001). Random forests. Machine learning, 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.

Author Accepted Proof

13. Caleyo, F., Gonzalez, J. L., & Hallen, J. M. (2002). A study on the reliability assessment methodology for pipelines with active corrosion defects. International journal of pressure vessels and piping, 79(1), 77-86.
14. Caleyo, F., Velázquez, J. C., Valor, A., & Hallen, J. M. (2009). Probability distribution of pitting corrosion depth and rate in underground pipelines: A Monte Carlo study. *Corrosion Science*, *51*(9), 1925-1934.
15. Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, *11*, 2079-2107.
16. CSA (Canadian Standards Association) (1999). CSA Z662-99, Oil and Gas Pipeline Systems. Ontario, Canada.
17. Chen, J. M., Meng, H. M., & Li, Y. Y. (1998). Pipeline prescription analysis after corrosion and explosive test. Oil Gas Storage Transp, 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, K. A. (2007). Best practice for the assessment of defects in pipelines–Corrosion. *Engineering Failure Analysis*, *14*(7), 1245-1265.
20. Cronin, D. S., & Pick, R. J. (2000). Experimental database for corroded pipe: evaluation of RSTRENG and B31G. In 2000 3rd International Pipeline Conference. American Society of Mechanical Engineers Digital Collection.
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, J. L. F., Vieira, R. D., Castro, J. T. P., & Benjamin, A. C. (2006). Part 3: Burst tests of pipeline with extensive longitudinal metal loss. Experimental Techniques, 30(6), 60-65.
24. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
25. Gao, J., Yang, P., Li, X., Zhou, J., & Liu, J. (2019). Analytical prediction of failure pressure for pipeline with long corrosion defect. Ocean Engineering, 191, 106497.
26. Ghosh, J., Padgett, J. E., & Dueñas-Osorio, L. (2013). Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. Probabilistic Engineering Mechanics, 34, 189-199.
27. Hasan, S., Khan, F., & Kenny, S. (2012). Probability assessment of burst limit state due to internal corrosion. *International Journal of pressure vessels and piping*, *89*, 48-58.
28. Huang, H., & Burton, H. V. (2019). Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. Journal of Building Engineering, 25, 100767.
29. Jeon, J. S., Shafieezadeh, A., & DesRoches, R. (2014). Statistical models for shear strength of RC beam-column joints using machine-learning techniques. Earthquake engineering & structural dynamics, 43(14), 2075-2095.
30. Jin, W. L., Zhang, E. Y., Shao, J. W., & Liu, D. H. (2004). Cause analysis and countermeasure for submarine pipeline failure. Bulletin of Science and Technology, 6.

Uncorrected Proof

31. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (pp. 3146-3154).
32. Kiani, J., Camp, C., & Pezeshk, S. (2019). On the application of machine learning techniques to derive seismic fragility curves. Computers & Structures, 218, 108-122.
33. Kiefner, J. F., & Vieth, P. H. (1990). Evaluating pipe--1. new method corrects criterion for evaluating corroded pipe. Oil and Gas Journal, 88(32).
34. Kim, Y. P., Baek, J. H., Kim, U. S., & Go, Y. T. (2004). The evaluation of burst pressure for corroded pipeline by full scale burst test. Transactions of the Korean Society of Mechanical Engineers A, 26(1), 203-210.
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, F. J., Stewart, G., & van der Valk, C. A. (1995). New developments in burst strength predictions for locally corroded pipelines (No. CONF-950695-). American Society of Mechanical Engineers, New York, NY (United States).
37. Leis, B. N., & Stephens, D. R. (1997). An alternative approach to assess the integrity of corroded line pipe-part I: current status. In The Seventh International Offshore and Polar Engineering Conference. International Society of Offshore and Polar Engineers.
38. Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216-231.
39. Ma, B., Shuai, J., & Xu, X. (2011). 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* (pp. 63-72).
40. Mangalathu, S., & Jeon, J. S. (2018). Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. Engineering Structures, 160, 85-94.
41. Mangalathu, S., & Jeon, J. S. (2019). Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: Comparative study. Journal of Structural Engineering, 145(10), 04019104.
42. Mangalathu, S., Hwang, S. H., & Jeon, J. S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures, 219, 110927.
43. Mannucci, G., & Harris, D. (2002). Fracture properties of API X 100 gas pipeline steels. Directorate-General for Research.
44. Melchers, R. E. (2015). Progression of pitting corrosion and structural reliability of welded steel pipelines. *Oil and Gas Pipelines*, 327-342.
45. Mok, D. H. B., Pick, R. J., Glover, A. G., & Hoff, R. (1991). Bursting of line pipe with long external corrosion. International journal of pressure vessels and piping, 46(2), 195-216.
46. Netto, T. A., Ferraz, U. S., & Estefen, S. F. (2005). The effect of corrosion defects on the burst pressure of pipelines. Journal of constructional steel research, 61(8), 1185-1204.

Under Review Preprint

47. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.
48. Qi, F., & Zhu, A. X. (2011). Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps. Computers & Geosciences, 37(9), 1425-1436.
49. Robles-Velasco, A., Cortés, P., Muñuzuri, J., & Onieva, L. (2020). Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering & System Safety*, *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, C. Y. (2017). Research on the reliability methods of corroded pipeline. Petroleum Science Bulletin, 2(2), 288-297.
52. 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.
53. Villarin, M. C., & Rodriguez-Galiano, V. F. (2019). Machine learning for modeling water demand. Journal of Water Resources Planning and Management, 145(5), 04019017.
54. Wang, Y. Q., Wang, W. B., & Feng, Q. S. (2009). Remaining strength assessment for corroded pipelines. Corrosion and Protection-Nanchang, 29(1), 28.
55. Wang, N., & Zarghamee, M. S. (2014). Evaluating fitness-for-service of corroded metal pipelines: structural reliability bases. Journal of Pipeline Systems Engineering and Practice, 5(1), 04013012.
56. Wang, Z., Pedroni, N., Zentner, I., & Zio, E. (2018). Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment. Engineering Structures, 162, 213-225.
57. Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. Structure and Infrastructure Engineering, 14(10), 1402-1411.
58. Ying, C., Qi-Guang, M., Jia-Chen, L., & Lin, G. (2013). Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica*, *39*(6), 745-758.
59. Yu, L., Porwal, A., Holden, E. J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. Computers & Geosciences, 45, 229-239.
60. Zakikhani, K., Nasiri, F., & Zayed, T. (2020). A Review of Failure Prediction Models for Oil and Gas Pipelines. *Journal of Pipeline Systems Engineering and Practice*, *11*(1), 03119001.
61. Zhang, Y., Burton, H. V., Sun, H., & Shokrabadi, M. (2018). A machine learning framework for assessing post-earthquake structural safety. Structural safety, 72, 1-16.

Accepted Proof

621

Table A1: Burst Failure Pressure (experimental and predicted)

Sl.	D mm	t mm	L mm	d mm	YS MPa	UTS MPa	Actual 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	
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	Freire et al. (2006)
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	
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	Mok et al. 1991
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	
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	Kim et al. (2004)
33	426	7	160	2.7	240	390	11.0	3.0	7.5	10.0	10.0	5.1	5.6	6.3	
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	Chen et al. (1998)
42	305	6.4	26	5	351	543	15.0	14.4	14.9	19.8	18.4	9.4	11.4	11.8	
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	Shuai et al. (2017)
56	914	16.4	150	9	739	813	21.0	19.8	18.0	24.5	23.8	17.1	20.0	21.2	
57	914	16.4	450	6	739	813	24.0	8.5	21.8	22.4	22.3	16.9	18.0	21.2	MH

Uncorrected Proof

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	
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	
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	

Cronin and Pick (2000)

*MH: Mannucci and Harris (2002); D: Diameter; t: thickness; L: length of corrosion defect; d: depth of corrosion defect; YS: Yield Strength; UTS: Ultimate Tensile Strength; Popt: Operating pressure

Uncorrected Proof

View publication stats