ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B: Mechanical Engineering
Online journal at:
asmedigitalcollection.asme.org/risk

Synthetic Data Generation Using Generative Adversarial Network for Burst Failure Risk Analysis of Oil and Gas Pipelines

Ram Krishna Mazumder[1]
Arcadis U.S. Inc.,
222 S. Main Street,
Akron, OH 44308
e-mails: RamKrishna.Mazumder@arcadis.com;
rxm562@case.edu

Gourav Modanwal[1]
Emory University,
201 Dowman Dr.,
Atlanta, GA 30322
e-mail: gourav.modanwal@emory.edu

Yue Li
Leonard Case Jr. Professor in Engineering,
Department of Civil and Environmental Engineering,
Case Western Reserve University,
Cleveland, OH 44106
e-mail: yxl1566@case.edu

Despite the pipeline network being the safest mode of oil and gas transportation systems, the pipeline failure rate has increased significantly over the last decade, particularly for aging pipelines. Predicting failure risk and prioritizing the riskiest asset from a large set of pipelines is one of the demanding tasks for the utilities. Machine learning (ML) application in pipeline failure risk prediction has recently shown promising results. However, due to safety and security concerns, obtaining sufficient operation and failure data to train ML models accurately is a significant challenge. This study employed a Generative Adversarial Network (GAN) based framework to generate synthetic pipeline data (D_{Syn}) using a subset (70%) of experimental burst test results data (D_{Exp}) compiled from the literature to overcome the limitation of accessing operational data. The proposed framework was tested on (1) real data, and (2) combined real and generated synthetic data. The burst failure risk of corroded oil and gas pipelines was determined using probabilistic approaches, and pipelines were classified into two classes depending on their probability of failure: (1) low failure risk (P_f: 0–0.5) and (2) high failure risk (P_f: >0.5). Two random forest (RF) models (M_{Exp} and M_{Comb}) were trained using a subset of 70% of actual experimental pipeline data, (D_{Exp}) and a combination of 70% of actual experimental and 100% of synthetic data, respectively. These models were validated on the remaining subset (30%) of experimental test data. The validation results reveal that adding synthetic data can further improve the performance of the ML models. The area under the ROC Curve was found to be 0.96 and 0.99 for real model ($M_{Exp)}$ and combined model (M_{Comb}) data, respectively. The combined model with improved performance can be used in strategic oil and gas pipeline resilience improvement planning, which sets long-term critical decisions regarding maintenance and potential replacement of pipes. [DOI: 10.1115/1.4062741]

1 Introduction

Community resilience relies on the smooth functionality and rapid recovery of civil infrastructure systems, such as water distribution systems, electrical power systems, and oil and gas transportation networks, road networks, after any disruptive event [1–2]. For oil and gas transportation, although pipeline systems are considered the safest mode of oil and gas transportation, the rate of failure of oil and gas pipelines has increased significantly in recent years [3–4]. The oil and gas transportation industry in the U.S. faces billions of dollars in losses due to aging pipeline failures. About 10,000 failures and 6.1 billons dollars in losses have been reported since 2002 [3,5]. More than half of American oil and gas pipelines are installed before 1960, and urban oil and gas systems are among the oldest [6]. Corrosion defects are the primary cause of the failure of pipelines, followed by natural hazards and third-party activities [7,8]. Corrosion growth on pipe walls weakens a pipeline's integrity, leading to initiate burst failure in pipelines. The increasing failure rate highlights the research need for improved and computationally efficient failure prediction models of aging oil and gas pipelines to avoid imminent failure and minimize consequences on society. The ASCE infrastructure report graded oil and gas transportation pipelines as a 'C-' in 2021, identifying large investments in oil and gas infrastructure to improve the condition of existing pipelines [6]. Given a large number of poor pipelines, oil and gas transportation utilities often struggle to adapt failure prediction models to maintain and prioritize their large number of assets due to budget and resource constraints.

A major challenge in adapting a failure risk assessment is to obtain an actual/real operational and failure dataset, which is often not readily accessible due to public safety and security concerns [4,9,10]. Many past studies failed to gain its acceptability as these models developed based on limited experimental and historical data since they are not comprehensive and location-specific [2,11,12]. In the absence of actual pipeline data, researchers rely on information from literature, numerical data, generating data from combining attributes, and synthetically generated data to model pipeline failure risk [2,4]. Although actual data is preferable for model generation, computer-simulated synthetic data can be used when actual data are not accessible or available [13]. However, such synthetic data

[1] R. K. Mazumder and G. Modanwal contributed equally to this paper.

Manuscript received February 9, 2023; final manuscript received June 7, 2023; published online July 14, 2023. Assoc. Editor: Cao Wang.

should be generated in a way such that as possible as actual characteristics [4]. Recent advancements in Machine Learning (ML) models have gained popularity in generating synthetic data in various fields [14–20]. Therefore, this study aims to generate synthetic pipeline data using the ML technique.

While actual and/or synthetic data is a prerequisite for burst failure risk analysis of oil and gas pipelines, the failure risk prediction model needs to be capable of predicting failure with acceptable accuracy and must be efficient in analyzing large pipeline data with minimal computational cost. Models developed in failure risk analysis of pipelines can be broadly classified into two categories, including qualitative models [21–26] and quantitative models [7,27–32]. Qualitative models typically assign the likelihood of failure or risk to pipelines based on individual or institutional perception. On the other hand, physics-based quantitative models estimate the remaining strength of the pipeline considering the effect of corrosion and compare it with external loading to determine the likelihood of failure of the pipeline. The remaining strength of pipelines is estimated as the failure pressure of pipe following various standards and guidelines, such as DNV-RP-F101, SHELL92, ASME B31G, modified ASME B31G, RSTRENG, CPS, SAFE, among others [7,27,33–37]. While earlier studies estimated the failure pressure of pipelines using these models through a deterministic approach, recent studies expanded these models using probabilistic analysis considering uncertainty associated with random variables [7,29,38]. The burst failure probability of a pipe was assessed by comparing burst failure pressure with operating pressure [7]. Although these physics-based probabilistic models can predict the burst failure risk with reasonable accuracy, the application of these models suffers from their high computational costs. In light of developing efficient failure risk prediction models, the feasibility of ML and other intelligent techniques has been investigated recently as an alternative to traditional failure prediction models [2,38–43]. These studies showed promising results in failure risk prediction compared to conventional failure risk models. Hence, in addition to generating synthetic data, this study trained the ML model using synthetic data to predict the burst failure risk of pipelines.

This study develops an ML-based framework for generating synthetic data as an alternative to actual data to overcome challenges associated with pipeline data accessibility. We first employed an ML-based generative adversarial network (GAN) model to generate synthetic pipeline data as an alternative to actual pipeline data. Then, the viability of synthetic data is assessed by training random forest (RF) model on synthetically generated data. The objective of this study is two manifolds: (1) generate synthetic pipeline data similar to actual characteristics of pipelines and (2) investigate the feasibility of generated synthetic data in failure risk prediction of pipelines.

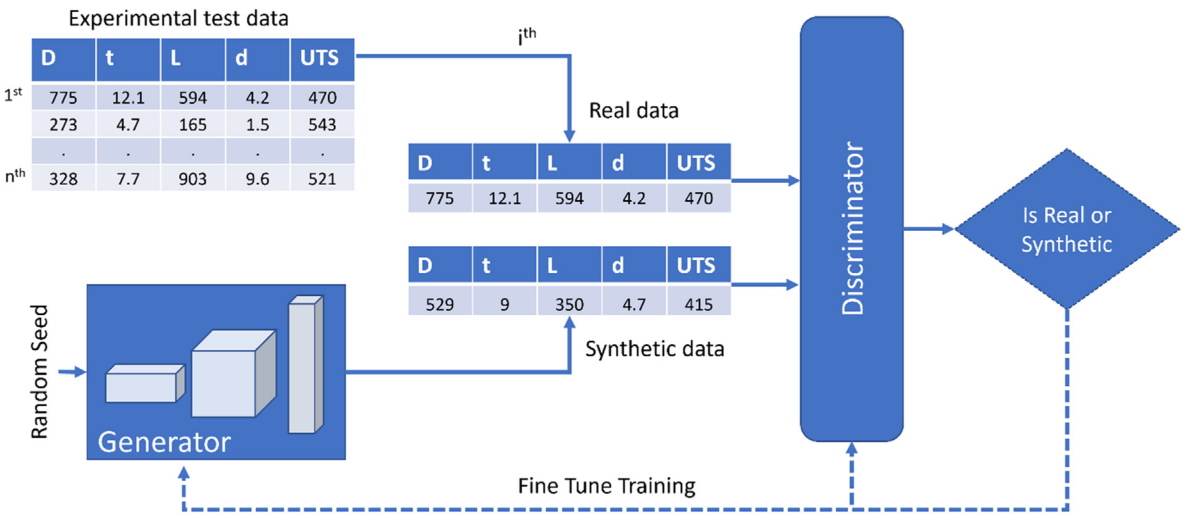
2 Background

2.1 Synthetic Data Modeling. Pipeline failure risk assessment requires actual/real data, including geometric properties of pipe, corrosion size, pipe age, material strength, etc., to predict the likelihood of failure under various conditions. Pipelines' operational and experimental data are strictly regulated due to the safety and security of the pipeline's assets and populations, which is often inaccessible and not readily available [2,4]. The objective of data synthesis is to generate new data so that new datasets do not contain any of the real data points. Synthetically generated data has privacy advantages, allowing users to analyze similar actual data without accessing the actual data [44]. As an alternative to actual/real pipeline data, synthetic data can be used for predicting failure risk analysis. When the data size is small and inadequate, synthetic data can be used alongside actual/real datasets. The utilization of synthetic data may raise a question in developing failure risk models and their applicability in failure risk analysis. Therefore, a validation test is required to ensure the applicability of data-driven models with accessible actual data [4].

Synthetic data can be generated in three ways: (i) perturbation of actual data, (ii) combining attributes from the real data, and (iii) generating samples from the statistical distribution. Various approaches are used to generate synthesized data, including Bayesian networks, Gaussian copulas, and normalization. However, these methods have multiple drawbacks related to data sizing and complexity. The use of ML-based Generative Adversarial Networks (GANs) has shown promising results in generating synthetic data, starting with images (e.g., StyleGAN) and recently extending to tabular data. GAN can overcome challenges faced by other statistical models and anonymization tools. While most of the models are trained by log-likelihood, GAN does not require explicit data analysis and is capable of deep representation of data without extensively annotating actual data [15,16]. This approach is so powerful that it can create synthetic data by learning characteristics of actual data that did not exist before but looked real [45]. Goodfellow et al. [46] originally introduced GANs architecture containing two neural networks: (1) a generator—that learns to generate instances from example, new data and (2) a discriminator—that evaluates data as real or synthetically generated. The two networks are trained in an adversarial manner until the generator model produces plausible examples of training data that trick the discriminator model half of the time. Beyond this point, GANs generator network became capable of producing synthetic data that never existed but looks real. Xu and Veeramachaneni [47] evaluated the performance of GANs on three datasets where the GAN model outperformed conventional statistical generative models. This study employed GAN to generate synthetic pipeline data learning from actual pipeline characteristics. Figure 1 shows the GAN architecture used to generate synthetic pipeline failure data.

2.2 Failure Risk Analysis Models. External corrosion on the pipeline wall is the most common cause of pipeline burst failure [7,29,48]. The remaining strength of a pipe of corroded steel pipelines can be estimated using codes and standards guidelines [28,29,4950]. As a result of corrosion growth on pipeline wall overtimes, the failure likelihood increases with pipeline ages. Therefore, the failure likelihood at any given age of the pipeline can be classified into deterministic risk classes to support asset management tasks [38]. After generating synthetic data, pipelines are classified into various failure risk classes using physics-based models and then apply the best predictive ML algorithm to evaluate the applicability of synthetic data. Hence, a comprehensive review is performed on physics-based and ML-based pipeline failure risk models.

2.2.1 Physics-Based Models. Researchers use various standards and guidelines to quantify the failure pressure of pipelines (e.g., [27,33,34,51]). These methods mostly used semi-empirical fracture mechanics relationships where failure depends on flow stress [52]. Burst failure of a pipeline occurs when the remaining strength of a pipeline significantly reduces due to corrosion or another form of deterioration, and internal operating pressure exceeds the remaining strength of pipe material [52–54]. The effect of corrosion on burst failure pressure can be accounted for by modeling time-variant corrosion growth on pipeline walls. As a result of corrosion deterioration, pipeline thickness changes over time. Hence, time-variant failure pressure is modeled by incorporating the time-variant thickness changes of pipeline in specified burst failure pressure estimation models [7,27,33–35]. The burst failure probability of pipelines is determined using a limit state function that compares burst failure pressure (i.e., remaining strength of the pipeline) and maximum operating fluid pressure [7,29,49,55]. Among existing standards, ASME B31G and DNV-RP-F101 are widely accepted failure pressure estimation models, regardless of their simplification and conservative estimation [38]. Physics-based predictive models can provide better accuracy in failure risk prediction over statistical models [10,56,57] and are very useful in supporting maintenance decision-making. However, the application of these models is time-consuming, computationally demanding, and requires a complete dataset with



	D	t	L	d	UTS
1st	775	12.1	594	4.2	470
	273	4.7	165	1.5	543
	.	.	.	.	.
nth	328	7.7	903	9.6	521

D	t	L	d	UTS
775	12.1	594	4.2	470

D	t	L	d	UTS
529	9	350	4.7	415

Fig. 1 GAN training pipeline used to generate synthetic pipeline failure data

desired attributes, which is often not readily available, making it difficult for utilities to adapt to their asset management plan [10,38].

2.2.2 Machine Learning Models. Machine learning is a semi-automated system in which computers learn and construct computational relationships from the observed data to predict the new observation. To overcome this limitation related to physics-based models, ML applications in failure risk analysis have gained considerable attention in recent years due to their higher efficacy and eliminating the manual burden required for computation. Machine learning models have been extensively investigated in recent years analyzing failure risk of pipelines. Researchers successfully trained various ML models for failure risk analysis of pipelines, including oil and gas pipelines failure risk analysis [38], maintenance decision-making, predicting failure pressure [2], condition assessment [58], water mains failures prediction [59]. Pipeline standards provide equations to determine the remaining strength (i.e., failure pressure) of pipelines considering corrosion pit on the external surface of pipelines, which is traditionally used to predict the failure of pipelines. Several algorithms have emerged in the last decade, such as support vector machines, k-nearest neighbors, and random forests [60]. However, ML application in failure risk analysis is limited to their application with analytically/numerically produced datasets rather than experimental datasets. The infrastructure dataset is not readily available and requires significant effort to process. Due to the security of utility and infrastructure networks, this dataset is often inaccessible. In case of unavailability of the dataset, researchers explored how limited datasets can be utilized and synthesized realistically to develop risk assessment strategies [38,52].

Researchers applied data-driven techniques to determine the efficiency of failure-predictive algorithms for structural performance assessment. For instance, Mangalathu et al. [42] applied a data-driven technique to recognize the failure modes of concrete shear walls. A comprehensive review of pipelines can be found in Rachman et al. [4]. Past studies mainly focused on amazing performance of individual pipes under corrosion deteriorations [2]. Wang and Li [61] used a data-driven approach for pipeline risk analysis based on the cluster model. Utilizing a small set of experimental data, Mazumder et al. [38] investigated the feasibility of eight ML models where RF was the most efficient for failure risk prediction. Among various ML Algorithms used for failure risk analysis of pipelines, including ANN, SVM, KNN, DT, and boosting, RF models have demonstrated higher accuracy and efficiency in predicting failure risk compared to other algorithms, such as decision tree [38,58,62–66]. The RF models are more accurate and less likely to have overfitting problems.

Overall, ML has shown promising results in pipeline failure risk analysis, which can be applied to estimate the likelihood and consequence of failure to prioritize pipelines. While traditional physics-based models from standards and guidelines can predict the burst failure risk of pipelines, performing extensive simulations and high computational cost is necessary for large oil and gas networks composed of hundreds of pipelines. In contract, ML can solve complex problems without explicit mechanical analysis. Researchers have shown that ML models can be efficiently applied as an alternative to physics-based models (e.g., Ref. [38]). By synchronizing operational data with ML models, it will be possible to predict failure risk in real-time ML models, it will be possible to predict failure risk in real-time. However, only a few studies investigate the feasibility of various ML algorithms in the failure risk analysis of pipelines.

3 Methodology

Figure 2 illustrates a graphical overview of the framework developed in this research. The framework contains three modules: (1) the GAN model to generate synthetic data by learning from actual/real pipeline training datasets, (2) failure risk classification of the pipeline using burst failure limit, and (3) failure risk validation applying the RF model.

First, the GAN model (Gaussian Copula) is trained using 70% of experimental burst test data (D_{Exp}) in the data generation module. A set of pipeline attributes, including diameter (D), thickness (t), corrosion pit depth (d), corrosion pit length (L), and ultimate tensile strength (UTS) are used to train the GAN model to generate synthetic data. The data generator of the GAN model learns from a real dataset characteristic to generate synthetic data until the discriminator of the GAN model cannot differentiate between synthetic and real data. Next, the quality of the synthetic dataset is compared with the real dataset using a statistical approach. The second module performs burst failure analysis using burst failure limit state for both real and synthetic data. The burst failure probability for each pipeline was estimated using probabilistic approaches, as discussed later in this paper. The failure probabilities are then classified into two categories: (1) low failure likelihood and (2) high failure likelihood. In the third module, once pipelines are classified into failure risk classes, two approaches are followed to train the Random Forest model to evaluate the efficiency and feasibility of GAN-induced synthetic data for pipeline failure risk analysis. The first RF model was trained only using 70% of real experimental test data (D_{Exp}), and the second RF model was trained using a combination of real and synthetic data (i.e., 70% of real

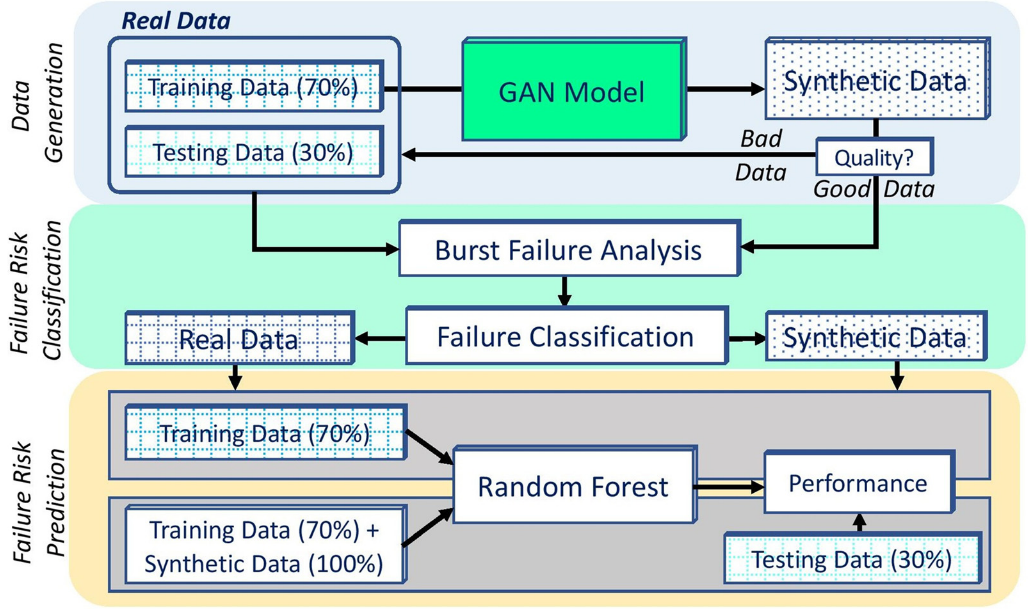


Fig. 2 ML-based failure risk analysis framework

experimental test data (D_{Exp}) and all synthetic data (D_{Syn})). The performance of both RF models was tested against 30% of real experimental test data.

3.1 Burst Failure Analysis. The burst failure probability of a pipeline is estimated using the burst failure limit state by comparing the pipeline burst failure pressure and internal pressure [7,29]. Burst failure pressure is defined as the ultimate failure pressure of a pipe at plastic collapse that represents the pipe's ultimate load-bearing capacity under internal pressure [28,52]. Failure pressure of a pipe can be estimated using various standards [27,28,51,67,68]. Time-variant burst failure pressure is estimated by accounting for pipeline thickness loss due to corrosion deterioration over time. The burst failure limit state is defined as

$$f(x) = \mathrm{FP}_T - P_{\mathrm{opt}} \tag{1}$$

where FP_t is the burst failure pressure of pipeline at time T, P_{opt} is the operating pressure of the pipe, and $f(x)$ is the limit state of the pipeline.

Among existing models to estimate the remaining strength of the pipeline, the DNV-RP-101F model is widely used to estimate the pipeline burst failure pressure. Recent studies compared burst failure prediction models and have recommended the DNV-RP-F101 model as one of the acceptable models to estimate burst failure pressure [2,38,52]. Using the DNV-RP-F101 model, the pipeline burst failure pressure is estimated as

$$\mathrm{FP}_T = \frac{2\,UTS}{D-t}\left(\frac{1-\dfrac{d(T)}{t}}{1-\dfrac{d(T)}{t}M^{-1}}\right); \quad M = \sqrt{1+0.31\frac{L(T)^2}{Dt}} \tag{2}$$

where UTS is the ultimate tensile strength, M is the Folias factor, D is the diameter, t is the initial thickness, $d(T)$ is depth of corrosion pit at time T, $L(T)$ is length of corrosion at time T, and T is the time in years.

The operating pressure of pipelines is typically regulated by standards and guidelines. To simulate burst failure probability, internal operating pressure, P_{opt}, for pipelines is determined using guidelines provided by the Canadian Standards Association [36,69].

$$P_{\mathrm{opt}} = \frac{2t}{D} YS \cdot F \cdot L \cdot J \cdot T^{\circ} \tag{3}$$

where F is the design safety factor (≈0.8), L is the location factor, J is the joint factor, T° is the temperature factor (≈1.0), and YS is the yield strength of pipeline [69].

The burst failure probability is computed using probabilistic analysis. The burst failure occurs when the internal operating pressure exceeds the failure pressure of the pipe (i.e., a negative value indicates failure, and a positive indicates nonfailure). A Monte Carlo Simulation is carried out to account for the uncertainty associated with random variables. The probability of failure of a pipeline is defined as

$$P_f = \sum \frac{f(x) \leq 0}{N_m} \tag{4}$$

where N_m is the number of Monte Carlo Simulation data points. Table 1 shows the stochastic models of random variables used [29,49].

Once the failure probability of a pipeline is determined, all pipelines are classified into failure risk classes depending on their failure probabilities. Two failure risk classes are defined based on the probability of failure: (1) low failure risk ($\mathrm{P_f} \leq 0.5$) and (2) high failure risk ($\mathrm{P_f} > 0.5$). Figure 3 shows an example of failure risk classification where failure probabilities greater than 0.5 belong to the high failure likelihood class and vice versa.

3.2 Random Forest. Random Forest model is employed in this study to evaluate the efficiency of the dataset's viability and the ML

Table 1 Statistical distribution of random variables

Parameters	Mean	COV	Unit	Distribution type
Pipe diameter, D	—	—	mm	Deterministic
Wall thickness, t	—	—	mm	Deterministic
Defect depth, d	μ_{d}	0.10	mm	Normal
Defect length, L	μ_{L}	0.10	mm	Normal
Ultimate tensile strength (UTS)	μ_{UTS}	0.07	MPa	Lognormal
Operating pressure, P_{opt}	μ_{Popt}	0.10	MPa	Normal

COV: Coefficient of variation.

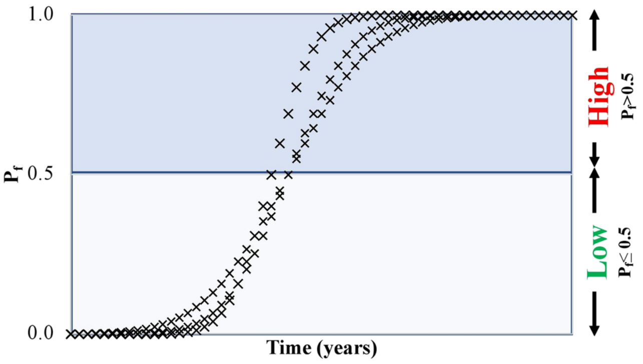


Fig. 3 Probability of failure levels

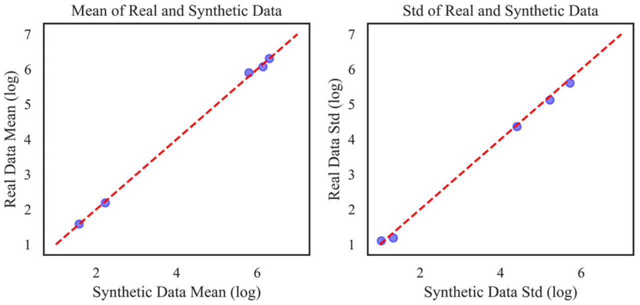


Fig. 4 Absolute log means and standard deviations of real and GAN-generated synthetic data. All values are log-transformed.

model's performance in failure risk analysis. ML has shown promising efficiency in pipeline failure risk analysis, particularly for analyzing large assets typical in oil or gas transportation networks [2,4,38]. An RF model is an ensemble learning method of classifications consisting of several individual decision trees. This RF uses two powerful supervised ML techniques: bagging and random feature selection [70]. An ensemble decision tree is usually built with bragging, whereas bragging is a combination of learning models to increase the accuracy of the results. The RF analyzes random features from the input features rather than using all features. The RF includes three main features, includes: (1) generating bootstrap samples from the training data, (2) developing a decision tree from each bootstrap using best split, and (3) predicting the output of new data utilizing decision trees developed from bootstrap [41]. The RF regression predictor is defined as [60]:

$$\hat{P}_{\mathrm{rf}}(x) = \frac{1}{N_t}\sum_{1}^{N_t} f_{N_k}(x) \qquad (5)$$

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

4 Result and Discussion

4.1 Synthetic Data Generation. A set of 92 experimental burst failure tests ($\mathrm{D_{Exp}}$) of steel pipelines subject to pitting corrosion was used in this study, originally assembled by Mazumder et al. [38]. This dataset contains five pipeline parameters, including diameter (D), thickness (t), corrosion pit depth (d), corrosion pit length (L), and ultimate tensile strength (UTS) of materials. Experimental data was divided into two sets: (1) a training set and (2) a testing set. The training set contains 70% of experimental burst test data that were used to train the GAN model to generate synthetic data. The test set contains 30% of experimental data that was separated to validate the ML-based accuracy in failure risk prediction, as described in Fig. 2.

The GAN model generated 100 synthetic pipeline data (D_{Syn}) where each pipe data contains D, t, d, L, and UTS. Synthetically generated data were examined to understand the differences and similarities between real and synthetic data. We first compared the means and standard deviations of real and synthetic data using a log–log plot; the expectation is that the plotted values follow the diagonal, suggesting real and synthetic data have comparable means and standard deviations, as shown in Fig. 4. These descriptive property checks provide good confidence in whether the data generator was able to capture the basic properties of materials.

Many synthetic data generators suffer from a partial collapse in the data generation process as such datasets are skewed significantly and contain many duplicates [44]. To avoid such a large amount of duplication, GAN models generate attributes closer to real data without duplicating a large proportion of the data. Hence, Cumulative Sum plots were generated to compare each parameter's real and synthetic data on top of each other. As shown in Fig. 5, a very high-quality matching was observed between real and synthetic data. The GAN model well captured the similarity for four parameters: thickness (t), corrosion length (L), corrosion depth (d), and UTS. Whereas three types of dimeters D (i.e., 324, 508, and 762 mm) dominate in a real dataset, the GAN model generates diameter sizes uniformly distributed over the entire range (i.e., minimum to maximum real pipe diameter). Overall, Fig. 5 represents that the GAN model generates a synthetic dataset closer to the real characteristics.

To further evaluate the performance of the GAN model and the characteristics of synthetic data, Principal Component Analysis (PCA) was performed, enabling the evaluation of a dataset's characteristics by comparing them using two dimensions, as shown in Fig. 6. The first two components of PCA analysis show well-distributed synthetic data and ranges within a realistic bound of actual/real data. The magnitude of generated data is much larger than the actual/real data, which covers a wide range of the dataset.

Figure 7 compares histograms and probability density functions of synthetic and actual/real data for all five parameters. Similar patterns are observed for pipeline real (i.e., 70% of $\mathrm{D_{Exp}} = 64$) and synthetic datasets (i.e., $\mathrm{D_{Syn}} = 100$). The probability density functions shape looks similar. A Kolmogorov–Smirnov (K-S) test was used to quantitatively compares the distributions of all the numerical columns of the generated synthetic GAN data with experimental data. Inverted K-S statistic ranges from 0 to 1.0, wherein a high value suggests a good fit. The average inverted K-S D statistic across all the features was 0.85, suggesting that it is more likely it is that the two samples were drawn from the same distribution. The Mean Correlation between synthetic and real data across columns was 0.9704. The qualitative comparison in Fig. 7 agrees with a good similarity between the datasets.

The GAN model was trained with 64 burst failure test data to generate 100 synthetic data. Synthetic data were examined in detail to understand the differences and similarities between independent holdout experimental data (i.e., 28 burst failure test data). This study was executed with a personal Nvidia GeForce RTX 3080 Ti computer with 16 GB memory. The model training took less than half an hour of computation on the GPU, and the final trained model took ∼1 min to generate 100 synthetic data. Given the characteristics of the experimental pipeline burst test data investigated in this study, GAN showed higher efficiency in synthetic data generation, and the RF model showed a greater accuracy in failure risk prediction on synthetic data generated from GAN model. A similar study on synthetic data generation on a different experimental dataset may require additional data depending on their characteristics. While SCADA produces a large set of various operational data, the current failure risk prediction requires only the internal operational pressure of a pipeline, which would potentially reduce the database processing efforts and computational cost. Pipeline operational database are typically stored as a database table format (e.g., Open Database Connectivity). ML algorithms can be applied in the production stage to process operational data and analyze the failure risk of an individual pipeline.

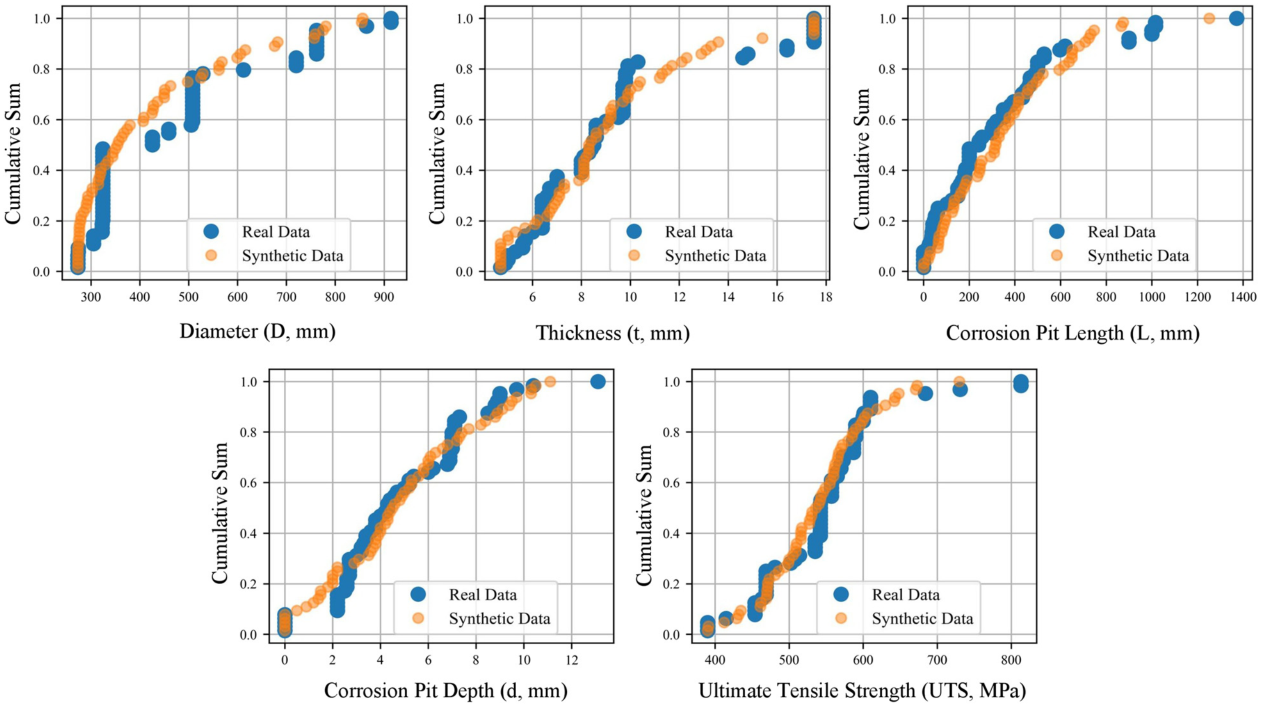


Fig. 5 Cumulative Sum plots of pipeline parameters

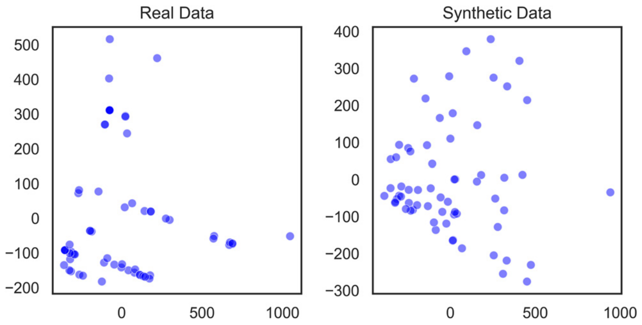


Fig. 6 First two components of principal component analysis in real and GAN-generated synthetic data

4.2 Failure Risk Prediction. Previous research extensively investigated various machine learning algorithms in the failure risk of oil and gas pipelines, in which the RF model showed promising outcomes in prediction with greater accuracy and efficiency. As discussed earlier, this study utilized a dataset from Mazumder et al. [38] to generate synthetic data. Mazumder et al. [38] tested eight ML algorithms on the same experimental dataset used in this study, where RF was found as the best predictive model. Hence, this study also trained RF models to evaluate the usefulness of synthetic data in failure risk prediction. The RF models (M_{Exp} and M_{Comb}) were trained with two sets of pipeline failure data: (1) 70% of real data (D_{Exp}), and (2) 70% of real data (D_{Exp}), along with all the synthetic data (D_{Syn}).

The performance of a classification problem was evaluated by Receiver Operating Characteristic (ROC) and Area Under the ROC Curve (AUC) curves. The ROC evaluates the performance of classification models at various classification thresholds. ROC is a probability curve, and AUC represents the degree or measure of separability. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is the probability that the model ranks a random positive example more highly than a random negative example. The first classifier model (M_{Exp}) achieved an AUC of 0.96, 95% CI [0.90–1.00], and the second classifier model (M_{Comb}) achieved an improved AUC of 0.99, 95% CI [0.98–1.00] on the hold-out set (i.e., 30% of $D_{\mathrm{Exp}} = 28$).

The performance measurement for machine learning classification is further evaluated using a confusion matrix, wherein each column in a confusion matrix represents an actual class, while each row represents a predicted class. The accuracy of an algorithm is expressed by the total percentage of correct prediction [41,71]. To overcome the possibility of misleading prediction accuracy due to unbalanced data, additional measures like Recall, Precision, F1-Score, Sensitivity, and Specificity, were also used. The higher these metrics are, the better the model performance is. Figure 8 shows a generic form of the confusion matrix and its performance measures [72]. Table 2 provides these key classification metrics on the independent validation set (i.e., 30% of testing data). Both RF models (M_{Exp} and M_{Comb}) performed well in the burst failure risk analysis of pipelines. M_{Exp} achieved an accuracy of 86% with a sensitivity value of 0.75 and a specificity value of 0.94 while M_{Comb} achieved an improved accuracy of 96% with an improved sensitivity value of 1.00 and a specificity value of 0.94 in failure risk prediction. The sensitivity metric evaluates the model's ability to predict true positives (i.e., predicted as failure when the actual pipe is failed). M_{Comb} had a higher F1 score (harmonic mean of precision and recall) than M_{Exp} suggesting superiority in classification performance in predicting failure risk (Fig. 9).

The results show that the RF model, with the incorporation of synthetic data, can predict burst failure risk classifications with higher accuracy and can be applied in failure risk predictions to support asset management.

5 Conclusions

Failure risk analysis of pipelines is crucial for effective asset management, as well as understanding the resilience of aging oil and gas pipeline systems. Predicting pipe failures may help decision-makers to identify the riskiest segments and take appropriate intervention measures to avoid potential consequences in developing resilient oil and gas pipeline systems. A major challenge in

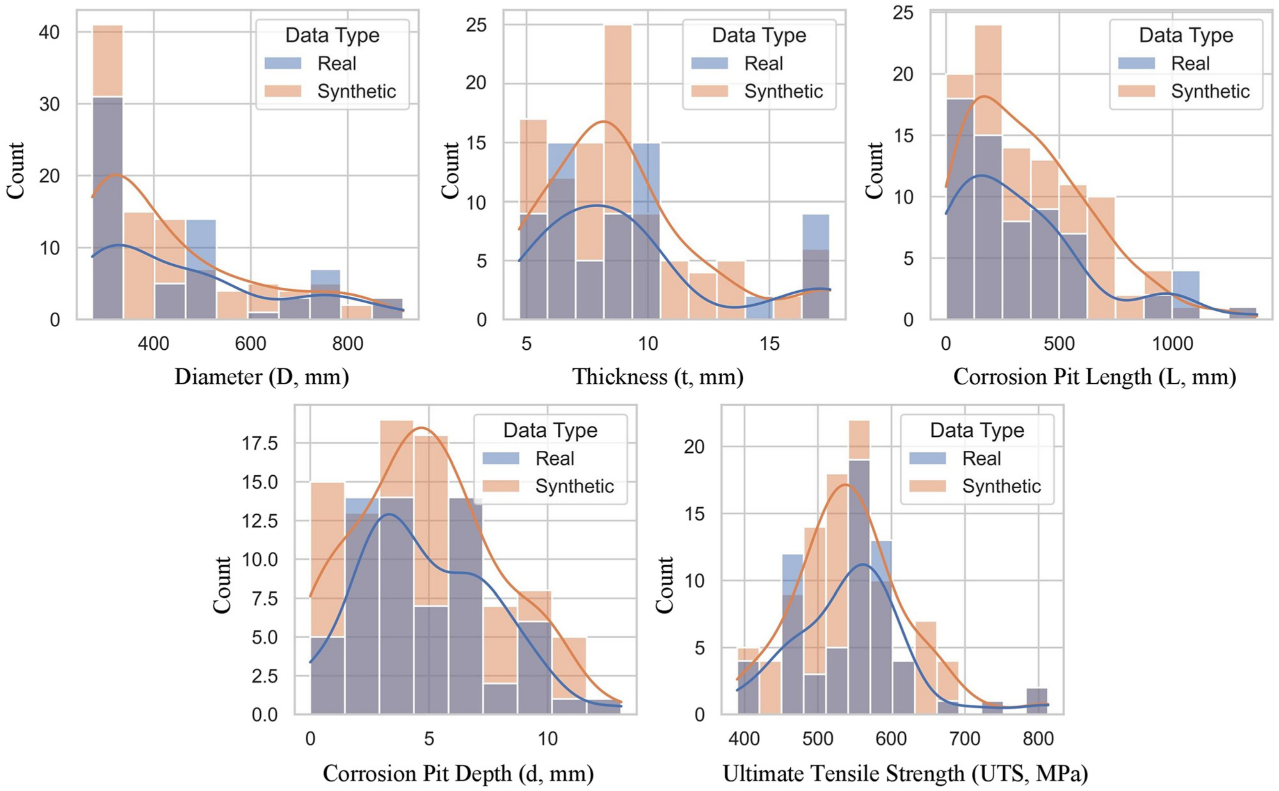


Fig. 7 Histograms and probability density functions of real and synthetic data

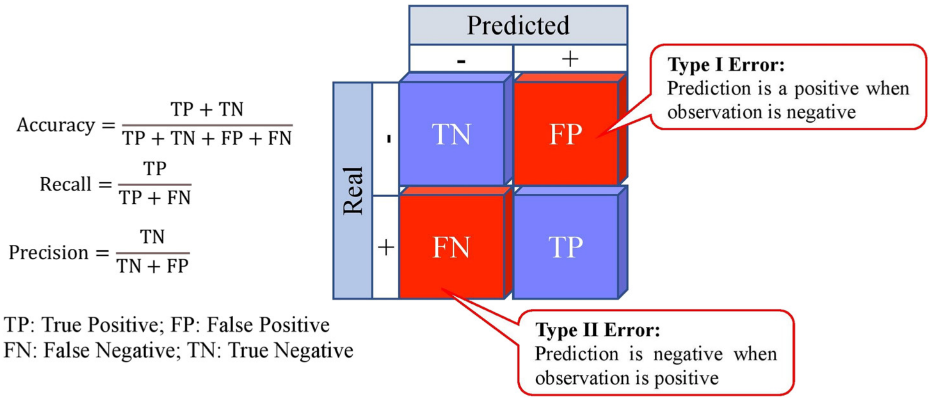

$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FP}}$
TP: True Positive; FP: False Positive
FN: False Negative; TN: True Negative
Predicted
-
+
Real
-
+
TN
FP
FN
TP
Type I Error:
Prediction is a positive when observation is negative
Type II Error:
Prediction is negative when observation is positive

Fig. 8 Confusion matrix

Table 2 Key classification metrics

Model and dataset	Accuracy	AUC	Recall	Prec.	F1	SENS	SPEC
M_{exp}: 70% of real (D_{Exp})	0.86	0.96	0.75	0.90	0.82	0.75	0.94
M_{comb}: 70% of real (D_{Exp}) + 100% Synthetic (D_{Syn})	0.96	0.99	1.00	0.92	0.96	1.00	0.94

studying the failure risk of pipelines is limited accessibility to real/actual operational pipeline data due to security and safety concerns. To overcome actual data sharing risks, this study employed an ML-based framework to generate synthetic pipeline data similar to real/actual pipeline characteristics. This study provides two major contributions to pipeline failure risk analysis literature by (1) developing an ML-based framework to generate synthetic data as an alternative to actual pipeline data and (2) assessing the feasibility of synthetic data by training an ML model in predicting burst failure likelihood. A GAN is used to generate synthetic data by learning from actual/real experimental data, which can be used as an alternative to real data. The characteristics of a synthetic dataset have been tested against real data, which has shown promising results and can be used for pipeline failure risk analysis. Synthetic data do not require privacy controls and allow users to study and share without concern about public safety. We further trained RF models to investigate the efficiency of ML models on real and synthetic data for predicting the failure of oil and gas pipelines. Our results show that with the incorporation of synthetic data, the RF model can predict burst failure risk classifications with improved

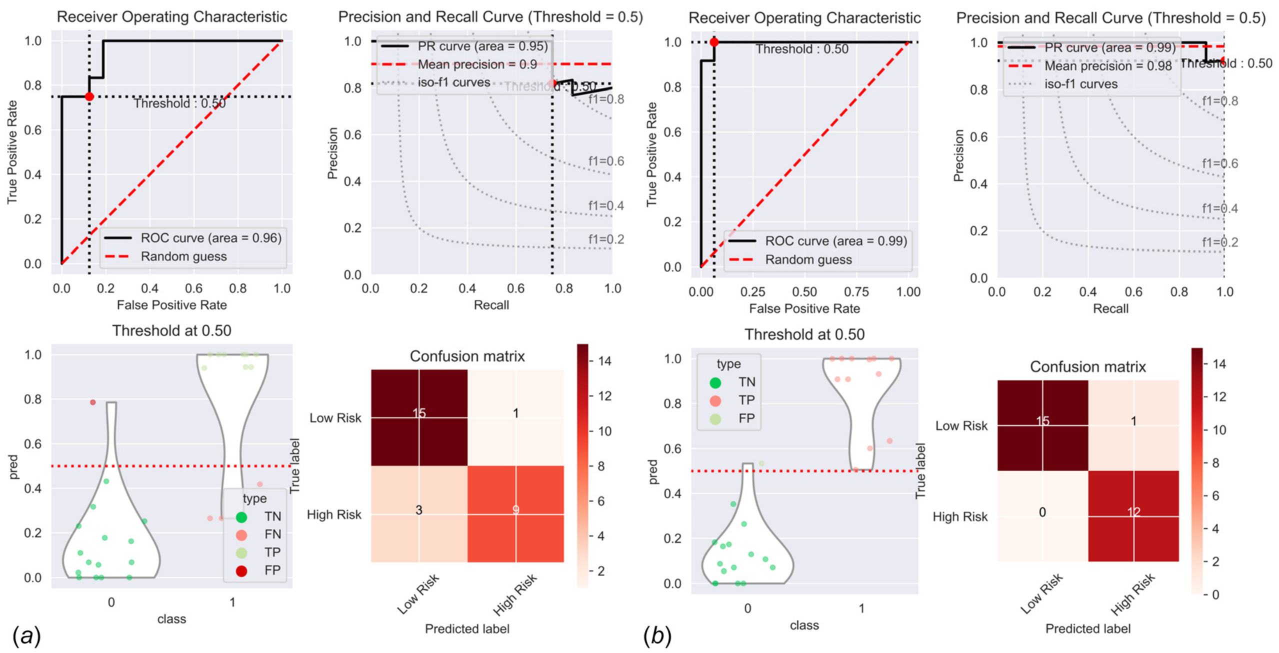


Fig. 9 Performance metrics for binary classifier models demonstrating ROC curves, precision-recall curve, class score distribution, and confusion matrices on independent validation set (i.e., 30% of D_{Exp} = 28) for (a) model trained on 70% of real data (M_{Exp}) and (b) model trained on 70% of real data along with synthetic data (M_{Comb})

AUC (0.99 versus 0.96) and accuracy (96% versus 86%) than real data only. Hence, the GAN model can be utilized for generating synthetic data and training an improved ML model for pipeline failure risk analysis. The amount of data needed for training the GAN model depends on the complexity of the problem being solved. Although the GAN application in this study showed greater efficiency in generating synthetic data based on 92 experimental burst failure tests, additional experimental failure test data may be required for a different dataset depending on its characteristics. In such a situation, the numerical simulation may be performed to generate additional data. The GAN model's applicability was based on experimental burst failure test data from pipelines with single corrosion defects. However, additional experimental or numerical simulation data on burst failure analysis of pipelines with different types of corrosion defects are required, which is out of the scope of this paper. Future studies may explore the feasibility of applying the GAN model to pipelines with various types of defects.

The RF model is very effective as it requires minimal computational efforts to support preventive maintenance tasks for asset management and improving the resilience of oil and gas pipelines. Although the framework is illustrated for oil and gas pipelines, the proposed framework can be applied to generate synthetic data for similar civil infrastructure systems. While this study was focused on burst failure analysis of oil and gas pipelines resulting from active corrosion, these pipelines also experienced failure due to geological hazards (e.g., earthquakes, landslides) and third-party events. Future research should investigate the potential applicability of ML and GAN models in failure risk analysis of pipelines subjected to these types of hazards.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Lin, P., and Wang, N., 2016, "Building Portfolio Fragility Functions to Support Scalable Community Resilience Assessment," Sustainable Resilient Infrastruct., **1**(3–4), pp. 108–122.

[2] Wang, X., Mazumder, R. K., Salarieh, B., Salman, A. M., Shafieezadeh, A., and Li, Y., 2022, "Machine Learning for Risk and Resilience Assessment in Structural Engineering: Progress and Future Trends," J. Struct. Eng., **148**(8), p. 03122003.

[3] Zakikhani, K., Nasiri, F., and Zayed, T., 2020, "A Review of Failure Prediction Models for Oil and Gas Pipelines," J. Pipeline Syst. Eng. Pract., **11**(1), p. 03119001.

[4] Rachman, A., Zhang, T., and Ratnayake, R. C., 2021, "Applications of Machine Learning in Pipeline Integrity Management: A State-of-the-Art Review," Int. J. Pressure Vessels Piping, **193**, p. 104471.

[5] DOT, 2020, "Distribution, Transmission & Gathering, LNG, and Liquid Accident and Incident Data," U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration, Washington, DC.

[6] ASCE 2021, "ASCE 2021 Infrastructure Report Card," ASCE, Reston, VA, accessed July 1, 2023, http://www.infrastructurereportcard.org/

[7] Teixeira, A. P., Soares, C. G., Netto, T. A., and Estefen, S. F., 2008, "Reliability of Pipelines With Corrosion Defects," Int. J. Pressure Vessels Piping, **85**(4), pp. 228–237.

[8] Zhu, X. K., 2021, "A Comparative Study of Burst Failure Models for Assessing Remaining Strength of Corroded Pipelines," J. Pipeline Sci. Eng., **1**(1), pp. 36–50.

[9] Ferraz, I. M. N., Garcia, A. C., and Bernardini, F. V. C., 2008, "Artificial Neural Networks Ensemble Used for Pipeline Leak Detection Systems," International Pipeline Conference, Vol. 48579, Calgary, AB, Canada, Sept. 29–Oct. 3, pp. 739–747.

[10] Mazumder, R. K., Salman, A. M., Li, Y., and Yu, X., 2018, "Performance Evaluation of Water Distribution Systems and Asset Management," J. Infrastruct. Syst., **24**(3), p. 03118001.

[11] Bertolini, M., and Bevilacqua, M., 2006, "Oil Pipeline Spill Cause Analysis: A Classification Tree Approach," J. Qual. Maint. Eng., **12**(2), pp. 186–198.

[12] Zhou, Q., Wu, W., Liu, D., Li, K., and Qiao, Q., 2016, "Estimation of Corrosion Failure Likelihood of Oil and Gas Pipeline Based on Fuzzy Logic Approach," Eng. Failure Anal., **70**, pp. 48–55.

[13] Xu, Q., Zhang, L., and Liang, W., 2013, "Acoustic Detection Technology for Gas Pipeline Leakage," Process Safety Environ. Prot., **91**(4), pp. 253–261.

[14] Ayala-Rivera, V., Portillo-Dominguez, A. O., Murphy, L., and Thorpe, C., 2016, "COCOA: A Synthetic Data Generator for Testing Anonymization Techniques," Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2016, Dubrovnik, Croatia, Sept. 14–16, Proceedings, Springer International Publishing.

[15] Metz, L., Poole, B., Pfau, D., and Sohl-Dickstein, J., 2016, "Unrolled Generative Adversarial Networks," International Conference on Learning Representations (ICLR) 2017, Vancouver, BC, Canada, Apr. 24–26.

[16] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A., 2018, "Generative Adversarial Networks: An Overview," IEEE Signal Process. Mag., **35**(1), pp. 53–65.

[17] Berkson, E. E., VanCor, J. D., Esposito, S., Chern, G., and Pritt, M., 2019, "Synthetic Data Generation to Mitigate the Low/No-Shot Problem in Machine Learning," 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, Oct. 15–17, pp. 1–7.

[18] Shah, S., Gandhi, D., and Kothari, J., 2020, "Machine Learning Based Synthetic Data Generation Using Iterative Regression Analysis," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, Nov. 5–7, pp. 1093–1100.

[19] Dewi, C., Chen, R. C., Liu, Y. T., and Tai, S. K., 2022, "Synthetic Data Generation Using DCGAN for Improved Traffic Sign Recognition," Neural Comput. Appl., **34**(24), pp. 21465–21480.

[20] Gujar, S., Shah, T., Honawale, D., Bhosale, V., Khan, F., Verma, D., and Ranjan, R., 2022, "GenEthos: A Synthetic Data Generation System With Bias Detection and Mitigation," International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), Kochi, India, June 23–25, pp. 1–6.

[21] Dziubiński, M., Frątczak, M., and Markowski, A. S., 2006, "Aspects of Risk Analysis Associated With Major Failures of Fuel Pipelines," J. Loss Prev. Process Ind., **19**(5), pp. 399–408.

[22] Yuhua, D., and Datao, Y., 2005, "Estimation of Failure Probability of Oil and Gas Transmission Pipelines by Fuzzy Fault Tree Analysis," J. Loss Prev. Process Ind., **18**(2), pp. 83–88.

[23] Shahriar, A., Sadiq, R., and Tesfamariam, S., 2012, "Risk Analysis for Oil & Gas Pipelines: A Sustainability Assessment Approach Using Fuzzy Based Bow-Tie Analysis," J. Loss Prev. Process Ind., **25**(3), pp. 505–523.

[24] Jamshidi, A., Yazdani-Chamzini, A., Yakhchali, S. H., and Khaleghi, S., 2013, "Developing a New Fuzzy Inference System for Pipeline Risk Assessment," J. Loss Prev. Process Ind., **26**(1), pp. 197–208.

[25] Lu, L., Liang, W., Zhang, L., Zhang, H., Lu, Z., and Shan, J., 2015, "A Comprehensive Risk Evaluation Method for Natural Gas Pipelines by Combining a Risk Matrix With a Bow-Tie Model," J. Natural Gas Sci. Eng., **25**, pp. 124–133.

[26] Guo, Y., Meng, X., Wang, D., Meng, T., Liu, S., and He, R., 2016, "Comprehensive Risk Evaluation of Long-Distance Oil and Gas Transportation Pipelines Using a Fuzzy Petri Net Model," J. Natural Gas Sci. Eng., **33**, pp. 18–29.

[27] Kiefner, J. F., and Vieth, P. H., 1990, "Evaluating Pipe–1. New Method Corrects Criterion for Evaluating Corroded Pipe," Oil Gas J., **88**(32), pp. 56–59.

[28] Netto, T. A., Ferraz, U. S., and Estefen, S. F., 2005, "The Effect of Corrosion Defects on the Burst Pressure of Pipelines," J. Constr. Steel Res., **61**(8), pp. 1185–1204.

[29] Wang, N., and Zarghamee, M. S., 2014, "Evaluating Fitness-for-Service of Corroded Metal Pipelines: Structural Reliability Bases," J. Pipeline Syst. Eng. Pract., **5**(1), p. 04013012.

[30] Oliveira, N., Bisaggio, H., and Netto, T., 2016, "Probabilistic Analysis of the Collapse Pressure of Corroded Pipelines," ASME Paper No. OMAE2016-54299.

[31] Ossai, C. I., Boswell, B., and Davies, I. J., 2015, "Estimation of Internal Pit Depth Growth and Reliability of Aged Oil and Gas Pipelines—a Monte Carlo Simulation Approach," Corrosion, **71**(8), pp. 977–991.

[32] Dundulis, G., Žutautaitė, I., Janulionis, R., Ušpuras, E., Rimkevičius, S., and Eid, M., 2016, "Integrated Failure Probability Estimation Based on Structural Integrity Analysis and Failure Data: Natural Gas Pipeline Case," Reliab. Eng. Syst. Saf., **156**, pp. 195–202.

[33] Det Norske Veritas, 1999, "Corroded Pipelines: DNV Recommended Practice RP-F101," Det Norske Veritas, Høvik, Norway.

[34] ASME B31G, 2009, "ASME B31G-2009: Manual for Determining the Remaining Strength of Corroded Pipelines," ASME, New York.

[35] Ritchie, D., and Last, S., 1995, "Burst Criteria of Corroded Pipelines-Defect Acceptance Criteria," Proceedings of the EPRG/PRC 10th Biennial Joint Technical Meeting Online Pipe Research, Cambridge, UK, Apr. 18, pp. 1–11.

[36] Cronin, D. S., and Pick, R. J., 2000, "Experimental Database for Corroded Pipe: Evaluation of RSTRENG and B31G," ASME Paper No. IPC2000-190.

[37] Wang, W., Smith, M. Q., Popelar, C. H., and Maple, J. A., 1998, "A New Rupture Prediction Model for Corroded Pipelines Under Combined Loadings," ASME Paper No. IPC1998-2064.

[38] Mazumder, R. K., Salman, A. M., and Li, Y., 2021, "Failure Risk Analysis of Pipelines Using Data-Driven Machine Learning Algorithms," Struct. Safety, **89**, p. 102047.

[39] Ghosh, J., Padgett, J. E., and Dueñas-Osorio, L., 2013, "Surrogate Modeling and Failure Surface Visualization for Efficient Seismic Vulnerability Assessment of Highway Bridges," Probab. Eng. Mech., **34**, pp. 189–199.

[40] Jeon, J. S., Shafieezadeh, A., and DesRoches, R., 2014, "Statistical Models for Shear Strength of RC Beam-Column Joints Using Machine-Learning Techniques," Earthquake Eng. Struct. Dyn., **43**(14), pp. 2075–2095.

[41] Mangalathu, S., and Jeon, J., S., 2019, "Machine Learning–Based Failure Mode Recognition of Circular Reinforced Concrete Bridge Columns: Comparative Study," J. Struct. Eng., **145**(10), p. 04019104.

[42] Mangalathu, S., Hwang, S. H., and Jeon, J. S., 2020, "Failure Mode and Effects Analysis of RC Members Based on Machine-Learning-Based SHapley Additive exPlanations (SHAP) Approach," Eng. Struct., **219**, p. 110927.

[43] Mazumder, R. K., 2020, "Risk-Based Asset Management Framework for Water Distribution Systems," Doctoral dissertation, Case Western Reserve University, Cleveland, OH.

[44] Brenninkmeijer, B., de Vries, A., Marchiori, E., and Hille, Y., 2019, "On the Generation and Evaluation of Tabular Data Using GANs," Doctoral dissertation, Radboud University, Nijmegen, The Netherlands.

[45] Xue, A., 2021, "End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, Jan. 3–8, pp. 3863–3871.

[46] Goodfellow, I. J., Shlens, J., and Szegedy, C., 2014, "Explaining and Harnessing Adversarial Examples," International Conference on Learning Representations (ICLR) 2015, San Diego, CA, May 7–9.

[47] Xu, L., and Veeramachaneni, K., 2018, "Synthesizing Tabular Data Using Generative Adversarial Networks," preprint arXiv:1811.11264.

[48] Cosham, A., and Hopkins, P., 2004, "The Effect of Dents in Pipelines—Guidance in the Pipeline Defect Assessment Manual," Int. J. Pressure Vessels Piping, **81**(2), pp. 127–139.

[49] Caleyo, F., Gonzalez, J. L., and Hallen, J. M., 2002, "A Study on the Reliability Assessment Methodology for Pipelines With Active Corrosion Defects," Int. J. Pressure Vessels Piping, **79**(1), pp. 77–86.

[50] Mazumder, R. K., Salman, A. M., Li, Y., and Yu, X., 2019, "Reliability Assessment of Corroded Water Distribution Networks," Pipelines 2019: Condition Assessment, Construction, and Rehabilitation, American Society of Civil Engineers, Reston, VA, pp. 343–353.

[51] Batte, A. D., Fu, B., Kirkwood, M. G., and Vu, D., 1997, "New Methods for Determining the Remaining Strength of Corroded Pipelines," ASME paper No. IPC2002-27147.

[52] Gao, J., Yang, P., Li, X., Zhou, J., and Liu, J., 2019, "Analytical Prediction of Failure Pressure for Pipeline With Long Corrosion Defect," Ocean Eng., **191**, p. 106497.

[53] Jin, W. L., Zhang, E. Y., Shao, J. W., and Liu, D. H., 2004, "Cause Analysis and Countermeasure for Submarine Pipeline Failure," Bull. Sci. Technol., **20**, pp. 529–533.

[54] Boxall, J. B., O'Hagan, A., Pooladsaz, S., Saul, A. J., and Unwin, D. M., 2007, "Estimation of Burst Rates in Water Distribution Mains," Proc. Inst. Civil Eng.-Water Manage., **160**(2), pp. 73–82.

[55] Amaya-Gómez, R., Sánchez-Silva, M., Bastidas-Arteaga, E., Schoefs, F., and Muñoz, F., 2019, "Reliability Assessments of Corroded Pipelines Based on Internal Pressure–a Review," Eng. Failure Anal., **98**, pp. 190–214.

[56] Rajani, B., and Kleiner, Y., 2001, "Comprehensive Review of Structural Deterioration of Water Mains: Physically Based Models," Urban Water, **3**(3), pp. 151–164.

[57] Kleiner, Y., and Rajani, B., 2001, "Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models," Urban Water, **3**(3), pp. 131–150.

[58] El-Abbasy, M. S., Senouci, A., Zayed, T., Mirahadi, F., and Parvizsedghy, L., 2014, "Artificial Neural Network Models for Predicting Condition of Offshore Oil and Gas Pipelines," Autom. Construction, **45**, pp. 50–65.

[59] Fan, X., Wang, X., Zhang, X., and Yu, P. A. X. B., 2022, "Machine Learning Based Water Pipe Failure Prediction: The Effects of Engineering, Geology, Climate and Socio-Economic Factors," Reliab. Eng. Syst. Safety, **219**, p. 108185.

[60] Villarin, M. C., and Rodriguez-Galiano, V. F., 2019, "Machine Learning for Modeling Water Demand," J. Water Resour. Plann. Manage., **145**(5), p. 04019017.

[61] Wang, Z., and Li, S., 2020, "Data-Driven Risk Assessment on Urban Pipeline Network Based on a Cluster Model," Reliab. Eng. Syst. Saf., **196**, p. 106781.

[62] Ouadah, A., 2018, "Pipeline Defects Risk Assessment Using Machine Learning and Analytical Hierarchy Process," 2018 International Conference on Applied Smart Systems (ICASS), Medea, Algeria, Nov. 24–25, pp. 1–6.

[63] Su, Y., Li, J., Yu, B., Zhao, Y., and Yao, J., 2021, "Fast and Accurate Prediction of Failure Pressure of Oil and Gas Defective Pipelines Using the Deep Learning Model," Reliab. Eng. Syst. Safety, **216**, p. 108016.

[64] Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., Lashari, N., Lu, H., and Sambo, C., 2022, "Integrity Assessment of Corroded Oil and Gas Pipelines Using Machine Learning: A Systematic Review," Eng. Failure Anal., **131**, p. 105810.

[65] De-León-Escobedo, D., 2023, "Risk-Based Maintenance Time for Oil and Gas Steel Pipelines Under Corrosion Including Uncertainty on the Corrosion Rate and Consequence-Based Target Reliability," Int. J. Pressure Vessels Piping, **203**, p. 104927.

[66] Liu, W., Liu, Z., Liu, Z., Xiong, S., and Zhang, S., 2023, "Random Forest and Whale Optimization Algorithm to Predict the Invalidation Risk of Backfilling Pipeline," Mathematics, **11**(7), p. 1636.

[67] Klever, F. J., and Stewart, G., 1995, "New Developments in Burst Strength Predictions for Locally Corroded Pipelines," ASME Paper No. IPC2002-27191.

[68] Leis, B. N., and Stephens, D. R., 1997, "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, Honolulu, HI, May 25–30.

[69] Canadian Standards Association, 1999, "*CSA Z662-99, Oil and Gas Pipeline Systems*," Toronto, ON, Canada.

[70] Breiman, L., 2001, "Random Forests," Mach. Learning, **45**(1), pp. 5–32.

[71] Robles-Velasco, A., Cortés, P., Muñuzuri, J., and Onieva, L., 2020, "Prediction of Pipe Failures in Water Supply Networks Using Logistic Regression and Support Vector Classification," Reliab. Eng. Syst. Saf., **196**, p. 106754.

[72] Luque, A., Carrasco, A., Martín, A., and de las Heras, A., 2019, "The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix," Pattern Recognit., **91**, pp. 216–231.