Structures 65 (2024) 106653

Contents lists available at ScienceDirect

Structures

journal homepage: www.elsevier.com/locate/structures

Machine learning-based Shapley additive explanations approach for corroded pipeline failure mode identification

Mohamed El Amine Ben Seghier [a,b,*], Osama Ahmed Mohamed [c], Hocine Ouaer [d]

[a] Laboratory for Computational Mechanics, Institute for Computational Science and Artificial Intelligence, Van Lang University, Ho Chi Minh City, Vietnam
[b] Faculty of Mechanical - Electrical and Computer Engineering, School of Technology, Van Lang University, Ho Chi Minh City, Vietnam
[c] College of Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates
[d] Faculty of Hydrocarbons and Chemistry, University M'hamed Bougara de Boumerdes, Algeria

ARTICLE INFO

Keywords:
Oil and gas pipelines
Corrosion/crack defect
Failure modes
Machine learning
Classification
Shapley additive explanations

ABSTRACT

Rapid failure mode identification of oil and gas pipelines can prevent catastrophic consequences, improve fast intervention and enhance the design safety of these critical systems. This paper proposes explainable-based machine learning models using to determine the failure mode of corroded pipelines as a function of geometric configurations, material properties, and corrosion defect details. To determine the best identification model, this study examined eight machine learning models, including Nave Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, Adaptive Boosting, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Category Boosting, based on a comprehensive experimental database for steel pipelines with various corrosion/crack defect configurations. Furthermore, the Shapley additive explanations approach is utilized to rank the input variables for failure mode identification and explains the machine learning model predicting a specific failure mode for a given sample. In identifying the failure mode of corroded pipelines, the proposed Extreme Gradient Boosting model indicated the highest accuracy in term of performance evaluation compared to all other proposed models. In addition, the model-explanation findings show that the important parameter influencing the failure mechanism of corroded pipelines is the depth of corrosion defects followed by the pipeline wall thickness. The proposed framework is adaptable enough to allow further use of experimental results for having new insights.

1. Introduction

Transporting hydrocarbons is a fundamental phase in delivering natural gas, crude oil, and other products to end-users customers [1]. Nowadays, buried steel pipelines are widely used as infrastructure systems for transporting hydrocarbons since they are the most cost-effective and safest alternative to other modes of transportation [2]. Steel pipes, however, are prone to a number of failure modes and deterioration due to the hostile surrounding environment and the elapsed operating time, with corrosion, cracking, and earth movement being the most typical causes [3–5]. Corrosion impact on steel pipeline is by far the most studied phenomenon, since statistics reveal that this phenomenon accounts for the majority of pipe failure (i.e. more than 38 % of pipe failure due to internal and exterior corrosion) [6–9]. Thus, steel pipes are highly prone to corrosion and with the restricted available control approaches again this phenomena as pipe-networks are immensely large, and the surrounding environment is harsh, corrosion process becomes inevitable[10–12]. The most frequent failure modes of steel pipelines, particularly those caused by corrosion defects or cracks, are rupture and leak. Both failure modes have detrimental effects, but the consequences of rupture are more hazardous for the environment and public safety than those of leaks [13,14]. Various studies have used the concept of structural reliability over the past decades to deal with the safety of steel pipelines that are subjected to specific failure causes in accordance with standards and codes (i.e. B31 G [15]), while others are limited to deterministic evaluation of the failure mode based on a specific regulation that account the crack/defect orientation (i.e. PRCI MAT-8 [16,17], CorLAS [18,19], Ln-Sec [20,21], the modified Battelle model [22,23], and PAFFC [24]). However, there is an absence of precise equations or models explicitly supplied by the traditional standards, norms, and regulations that may characterize the relationship between the failure modes and the corroded pipeline information. Therefore,

* Corresponding author at: Laboratory for Computational Mechanics, Institute for Computational Science and Artificial Intelligence, Van Lang University, Ho Chi Minh City, Vietnam.
E-mail address: benseghier@vlu.edu.vn (M.E.A. Ben Seghier).

https://doi.org/10.1016/j.istruc.2024.106653
Received 3 May 2023; Received in revised form 6 May 2024; Accepted 25 May 2024
Available online 14 June 2024

effective forecasting models of pipe failure modes subjected to corrosion and crack defects are critical for continued safety monitoring of these systems.

Lately, machine learning algorithms have been proposed as powerful tools for addressing pipeline-related issues such as corrosion modeling, reliability and safety analysis, risk assessment, burst pressure evaluation, and others [25–28]. Based on the available data for a given problem, such advanced techniques can develop a complicated relationship between the input and output characteristics [29,30]. All of the new paradigm-based machine-learning algorithms are proven to be more accurate than the existing standards and correlations for oil and gas challenges, where a paradigm can be either a classification or a regression, depending on the problem nature [31,32]. Nevertheless, the majority of the paradigms suggested to handle pipeline concerns were regression problems. For example, various machine-learning techniques, such as support vector regression, artificial neural network, response surface method, adaptive neurofuzzy inference system, and others, are used to predict corrosion, crack, and burst pressure of oil and gas pipelines, while current research combines machine learning with structural reliability to estimate the safety levels of these structures [33,34]. Several researchers used similar models to examine failure causes and risk assessment of steel pipes. Sun and Zhou [35] recently investigated failure modes (i.e. leak and rupture) classification using five machine learning models, with promising results of the applicability of such approaches to deal with this important problem. To the best of our knowledge, this was the only study to distinguish oil and gas pipeline failure modes. Nevertheless, the study was limited to the use of machine learning models, with no model interpretation. It is commonly known that the factors affecting the paradigm output have a significant impact on how well a paradigm performs in terms of providing the best predictions. Furthermore, interpretable paradigms depending on the model's input variables can provide new insights and explain the contribution of each variable.

In general, machine learning transfers the input variables to a gauge space where regression coefficients are recovered and utilized to produce operators that characterize the link between the input and output variables. Various works [36–39] investigated the machine learning paradigms explanation using local estimations and game theory concepts in the case of a linear relationship. SHapley Additive ex-Planations (SHAP) is an effective method for the interpretation of the machine learning paradigms based on the used input variables. This method's advantages include the ability to explain numerous supervised learning paradigms and to provide a relevance value for each input variable for a specific prediction/classification. As a result, the SHAP approach can be an effective tool for providing a global performance interpretation of the produced machine learning models for the identification of the failure modes of oil and gas pipelines.

Motivated by the preceding arguments, the purpose of this work is to apply various machine learning approaches, including Nave Bayes (NB), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Category Boosting (CatBoost) to identify (i.e. a classification problem) the failure modes of steel pipelines used for oil and gas transportation. Following the evaluation of the optimal model performance, the SHAP approach is utilized to interpret the best machine learning paradigm in order to describe the complex nonlinear behavior of steel pipeline failure modes, where the morphology of the corrosion/crack defect is also explored. To do this, a big database comprised of 250 datasets is used after statistical analysis and pre-processing. The remainder of this work is organized as follows: Section 2 briefly describes the used machine learning techniques, including the SHapley Additive ex-Planations approach, while Section 3 introduces and analyzed the employed database to perform this study comprehensively. Section 4 describes the implementation procedure as well as reporting and discussing the obtained results. Section 5 concludes by summarizing the key findings of this study.

2. Methods

2.1. Description of machine learning techniques

In the first part of this study, the optimal failure mode classification algorithm will be selected through the performance evaluation of eight machine learning paradigms: Naive Bayes (NB), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Category Boosting (CatBoost). The following subsections provide brief explanations of these techniques.

2.1.1. Naïve Bayes

NB is a straightforward machine-learning algorithm that utilizes the Bayes theorem to sort data. If the probability associated with the predicted variable exceeds 0.5, NB classification is used to determine if the output belongs to a particular category [40]. The NB classifier operates under the assumption that the impact of a feature's value on a chosen category is independent to the values of other features [41].

$$P_r(x) = P_r(Y = f|X = x \quad) = \frac{\pi_f f_f(x)}{\sum_{i=1}^{K} \pi_i f_i(x)} \tag{1}$$

with $f_f(x)$ designates the density function of an observation originating from the f^{th} class, and π_f represents the prior probability of a randomly chosen observation originating from the f^{th} class.

2.1.2. K-nearest neighbors

The non-parametric method KNN operates without making any assumptions about the decision boundaries between the failure modes. According to this machine-learning approach, a failed pipe is determined to belong to a failure mode if K out of the attribute space's most comparable specimens do [42].

In this algorithm, if K most similar pipe-specimens in the attribute space were associated with a specific failure mode, then the pipe-specimen in question is classified under that failure mode. The following equation represents the calculation of a conditional probability for x in failure mode [43].

$$P_r(x) = P_r(Y = f|X = x \quad) = \frac{1}{K}\sum_{i \epsilon N_K} I(y_i = f) \tag{2}$$

N_K stands for the K training set points that are closest to observation f.

2.1.3. Decision tree

DT is a non-parametric classification method that employs training data to construct a tree-like classification scheme [44]. DT construction involves iteratively dividing the tree structure nodes into two secondary nodes repetitively. Effectually, decision nodes are those that can be divide in a DT while leaves are those that cannot be further divide. The root node, representing the uppermost point of the inverted tree, initiates the DT growth process.The growth of DT proceeds from the root node through several stages. In first step, the algorithm selects the best attribute allocation that optimizes a division norm based on the training dataset's various input attributes. Subsequently, the optimal decision node allocation that optimizes the splitting criterion is selected. The decision node is then divided utilizing this optimal allocation, and the process continues until a predefined stopping criterion is met.

A commonly used splitting criterion for classification trees is the Gini Impurity Index (GI), a measure of the total variance across all categories [45]. The goal is to minimize the GI to identify the best decision node allocation. The stopping criterion for classification trees typically involves ensuring that all leaves are clear. Following the decisions in the tree from the starting to the ending nodes allows for the completion of

response forecasts. Each node is associated with a quiz case, and each branch represents an outcome of the quiz [46].

2.1.4. Random forest

RF classifier comprises a collection of tree predictors, where each classifier is built using a random vector selected independently from the input vector [47]. RF employs two crucial techniques: out-of-bag appreciations and random attribute subspace [48]. The first enables far quicker tree construction, and the last offers the chance to evaluate the relative importance of each input characteristic.

For RF training, multiple subsets of datasets are randomly sampled with substitution from the training dataset. Every sampled subset is then utilized to construct a decision tree. Applying RF to a novel sample involves obtaining predictions from each DT in the forest. The final forecast is determined by the category predicted by the majority of the trees in the RF. The RF advantages compared to DT are evident in being able to treat several input attributes, more strong to transact with outliers, and less liable to overfitting. In the literature, this method is described in more detail [49].

2.1.5. Adaptive boosting

In 1995, Freund and Schapire [50] introduced AdaBoost algorithm, a groundbreaking solution to various practical classification and regression problems. AdaBoost stands out as the primary adaptive boosting algorithm, adjusting its parameters based on the actual performance in each iteration, making it highly adaptive to the data. The algorithm has gained significant attention in the machine learning domain, particularly for its application in modest trees classification, yielding superior results compared to single-tree classification [51,52]. The initial step of AdaBoost involves assigning equal weight to all observations. Subsequently, the model is reconstructed, giving less weight to observations that were correctly classified and more weight to those that were misclassified.

2.1.6. Extreme gradient boosting

Chen and Guestrin. [53] introduced a scale gradient boosting technique, a method focused on enhancing regression by working on the straightened gradient vector function of the loss function from the previous iteration. This technique laid the foundation for the gradient boosting class, from which three subsequent approaches emerged: XGBoost, LightGBM, and CatBoost. These methods repeatedly fit a sequence of trees using gradient boosting, with decision trees serving as the foundational weak learner. XGBoost, in particular, distinguishes itself by addressing misclassification errors from the previous paradigm to fit the novel paradigm [54]. However, XGBoost executes slowly since it uses sequential paradigm training. Notably, XGBoost is designed to support linear classifiers and employs Taylor expansion for the cost function, enhancing result precision [55]. XGBoost quickly became one of the most used algorithms, prominently after featured in Kaggle competitions 2014, where it gained a remarkable track record of contest victories.

The tree integration paradigm of XGBoost involves the accumulative sum of the forecasted values of a pattern in every tree, serving as the forecast of the pattern framework. By incorporating uniform terms and immediately utilizes the first and second derivative value of the loss function, XGBoost effectively minimizes the risk of over-fitting. For a detailed explanation the algorithm's steps, refer to [56].

2.1.7. Light gradient boosting machine

Lunched by Microsoft in 2017, LightGBM is a lightweight gradient booster [56]. Distinguishing itself from traditional methods, LightGBM adopts leaf-wise model development approach instead of depth-wise, drawing its foundation from DT algorithms. This unique construction method results in more precise and complex trees. The algorithm is specifically designed to address the challenges of time-consuming operations in high-dimensional datasets [57]. In essence, LightGBM can be seen as an enhanced version of XGBoost. More detail of this algorithm are stated in [56].

2.1.8. Category boosting

CatBoost is defined as an implementation of gradient boosting, which utilizes binary decision trees as fundamental foretellers [58]. What sets CatBoost apart is its effective handling of categorical input parameters, allowing for the utilization of their benefits throughout training without the need for separate preprocessing. This strategy aims to address issues related to forecast shift and gradient bias [59]. The most advantages of this technique include[60] i) Automatic handling of categorical attributes as numerical properties, ii) Incorporation of class attributes that leverages connections between them, significantly enhancing attributes dimensions, and iii) The problem of overfitting is minimized and the algorithm precision is ameliorated by adopting a fully uniform tree paradigm.

2.2. Shapley additive explanations

SHAP is a well-known useful techniques that can describe successfully the performance of a machine-learning paradigm. SHAP is based on the game theory concept, employs an additive feature attribution technique to build an interpretable paradigm, i.e. the output paradigm is defined as a linear sum of input variables [61]. To understand more this technique, assume a paradigm has $\boldsymbol{y} = (y_1, y_2 \ldots, y_k)$ vector of input variables, where k represents the whole number of input variables. The expression of the explanation paradigm $h(y')$ with less complex input y' for a veritable paradigm $g(y)$ can be expressed as follows:

$$g(\boldsymbol{y}) = h(\boldsymbol{y}') = \sigma_0 + \sum_{i=1}^{N} \sigma_i y_i' \tag{3}$$

with N being the total number of input characteristics and σ_0 designating the determined value when inputs are completely missed. A mapping function uses the relation $y = f_y(y')$ to connect the inputs y' and y. Eq. (3) only takes one solution, according to Lundberg and Lee [62], and this solution has three advantages over others. Consistency, local accuracy, and missingness are among these characteristics [62]. Consistency makes it clear that changing a significant effect merit will not cause the ascription to that merit to be reduced. Local precision ensures that the output of the function represents the sum of the benefit attributions and requires the paradigm to correspond with the output of the g function for the simplified input y′. Local precision takes place when the relation $y = f_y(y')$ is satisfied. Missingness ensures that no significance is given to missing characteristics. As $y_i' = 0$ reveals that $\sigma_i = 0$, missingness is fulfilled. For a context $w' \backslash i$ when $w_i' = 0$, $g_y'(w') - g_y'(w' \backslash i\ ) \geq g_y(w') - g_y(w' \backslash i\ )$ means $\sigma_i(g', y) \geq \sigma_i(g, y)$. Only one paradigm can satisfy these characteristics, which is:

$$\sigma_i(g, y) = \sum_{w' \subseteq y'} \frac{|w'|!(N - |w'| - 1)!}{M!} \left[g_y(w') - g_y\left(w' \backslash i\ \right) \right] \tag{2}$$

with $w' \subseteq y'$ and σ_i denote the Shapley values, and $|w'|$ is the number of non-zero entries in w'. The expression $g_y(w') = g\left(f_y(y')\right) = H[g(w)/w_T]$ is proposed as a solution to Eq. (2) by Lundberg and Lee [63], where T represents the group of non-zero indices in w', recognized as SHAP values. In order to estimate the SHAP values, a variety of strategies, including as Tree SHAP, Kernel SHAP, and Deep SHAP, can be utilized. These techniques can provide perfect explanations of local and global paradigms. Tree SHAP is used in this study to acquire SHAP values and explanations.

3. Experimental database

In order to accurately build a machine-learning paradigm for the

failure mode identification of oil and gas pipelines, a thorough experimental database report, including statistical analyses and preprocessing is presented in the current section.

3.1. Database description

A comprehensive representative database that describes and includes thorough information about the failure mode mechanism in oil and gas pipelines is reported in this study, where it consists of 250 full-scale tests on pipe specimens. The database was originally retrieved from various sources and compiled by Sun and Zhou [35], whereas the gathered data comprised of burst tested pipes with a single surface crack that runs longitudinally, in which the manufactured defect may either be a semi-elliptic crack defect (SE) or a rectangular crack defect (REC). Two primary failure mechanisms were observed overall with these pipes' testing results: leaks in 135 of the examined pipes and ruptures in 115 of them. Furthermore, the defect profiles in the dataset are divided into 200 semi-elliptic and 50 rectangular shapes (See schematic views in Table 1). According to the data report, there were five pipe specimens with internal crack defects, nineteen pipe specimens without information about the defect location, and all the other pipe specimens with external crack defects.

Table 1 lists the used database statistical characteristics, including the mean, minimum, maximum, standard deviation (SD), ranges, and quartiles (i.e. Q1, median and Q3). The database includes eight input parameters: the diameter (D, *mm*), wall thickness (Wt, *mm*), corrosion/crack depth (d, *mm*), corrosion/crack length (L, *mm*), yield strength (σ_y, *MPa*), ultimate tensile strength (σ_u, *MPa*), Sharpy V-notch (Cv, *J*), and the crack/corrosion defect shape (REC or SE). The failure mode; i.e. leak or rupture; is the intended target. Furthermore, Fig. 1 displays the input and output variable histograms. This demonstrates how the data distribution varies according to each variable, indicating non-normal distributions and distinct scales that complicate the database and necessitate the use of sophisticated pre-processing and modeling techniques. Data balancing is necessary to prevent biased models since the two classes that indicate the target output's failure modes—leak and rupture—as well as the defect shape—SE, REC—have distinct values based on the histograms.

3.2. Statistical analyses database

More statistical analysis is carried out in order to understand the relationship between the input variables and the target output and to set up the database for integration with machine learning techniques. To begin with, the database underwent data analysis to exclude out imperfect samples, missing data, and inaccurate data. Following data analysis, it was found that all of the information in the database is correct without missing information. After that, a correlation matrix and scatterplot-based distribution are displayed as a visual depiction of the database complexity.

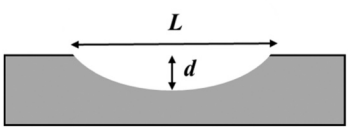
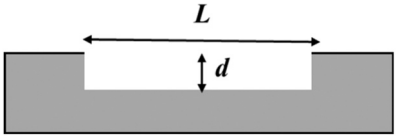
In order to identify the highly correlated variables, correlation matrix of the numerical input variables based on the Pearson correlation coefficient (r) is depicted in Fig. 2 for the scenario where $r \in [-1, 1]$ (i.e. full range) and $r \in [-1, -0.5] \cup [0.5, 1]$(i.e. The case of highly correlated variables). The pipeline design parameters (diameter (D) and wall thickness (Wt)) and defect geometries (depth (d) and length (L) of the defect) appear to have average correlation, which make sense as corrosion defects are located on the pipe-surface. The defect depth (d) and wall thickness (Wt) showed the highest correlation (i.e. $r = 0.89$), which is expected as the manufactured corrosion/crack defect values are inversely proportional to the pipe wall thickness ($0 < d < WT$). The yield strength (σ_y) and ultimate tensile strength (σ_u) also displayed a strong correlation ($r = 0.98$). From a mechanical perspective, these two parameters are significant because they reveal details about the steel pipeline failure process. The diameter (D), defect length (L), and Sharpy V-notch (Cv) on one side and the yield strength (σ_y) and ultimate tensile strength (σ_u) on the other side have negative correlation values, which shows these parameters impact in reducing the pipes' strength resistance.

Fig. 3 visualizes the scatterplot and distribution between the failure modes (i.e. L or R) and the input variables, which are subdivided into three subplots depending on crack/corrosion characteristics, material properties, and pipeline design geometries. Overall, it is clear that the distribution of the failure modes along the input variables is similar, and the proportion of samples with leak and rupture is quite close one to another. This figure illustrates that while the distribution of failure modes in each remains balanced, the number of defects with SEC (referred to as 0 in the Fig. 3(a)) is significantly lower than SE (referred to as 1 in the Fig. 3(a)) –unbalanced data-.

4. Framework implementation and performance analysis

From the assembled and analyzed database, the machine learning techniques described in the previous section were employed to identify the failure mode (i.e. Rupture or Leak) of oil and gas pipelines with different corrosion/crack shapes. To address the presented problem (i.e. classification), the corrosion/crack defect shapes and failure mode were converted to the eight input variables and output variables, respectively, in order to explore machine learning models and investigate the effect of corrosion/crack morphology (i.e. SE or REC) on failure mode. As there is

Table 1
Statistical attribution of the oil and gas pipeline failure mode identification database.

	min	Q1	median	Q3	max	mean	SD	range
D (*mm*)	76.4	178	230	565	1422.4	368.0808	302.5513	1346
Wt (*mm*)	3.2	5.3	6.8	9.7	21.7	8.2424	4.4985	18.5
d (*mm*)	0.8	3.7	5.05	6.2	17.8	5.7756	3.3326	17
L (*mm*)	17	65	74.75	143	609.6	122.2688	111.4618	592.6
σ_y (*MPa*)	246	440.075	661	852	1096.3	654.4564	245.1653	850.3
σ_u (*MPa*)	454	584	823.5	964	1179	792.6852	211.9484	725
C_v (*J*)	19.2	43.1	63.5	100	261	77.3616	53.7030	241.8
Defect Shape	Semi-elliptic defect (SE)				Rectangular defect (REC)			
								
Failure mode	Leak (L)				Rupture (R)			

SD: Standard deviation.

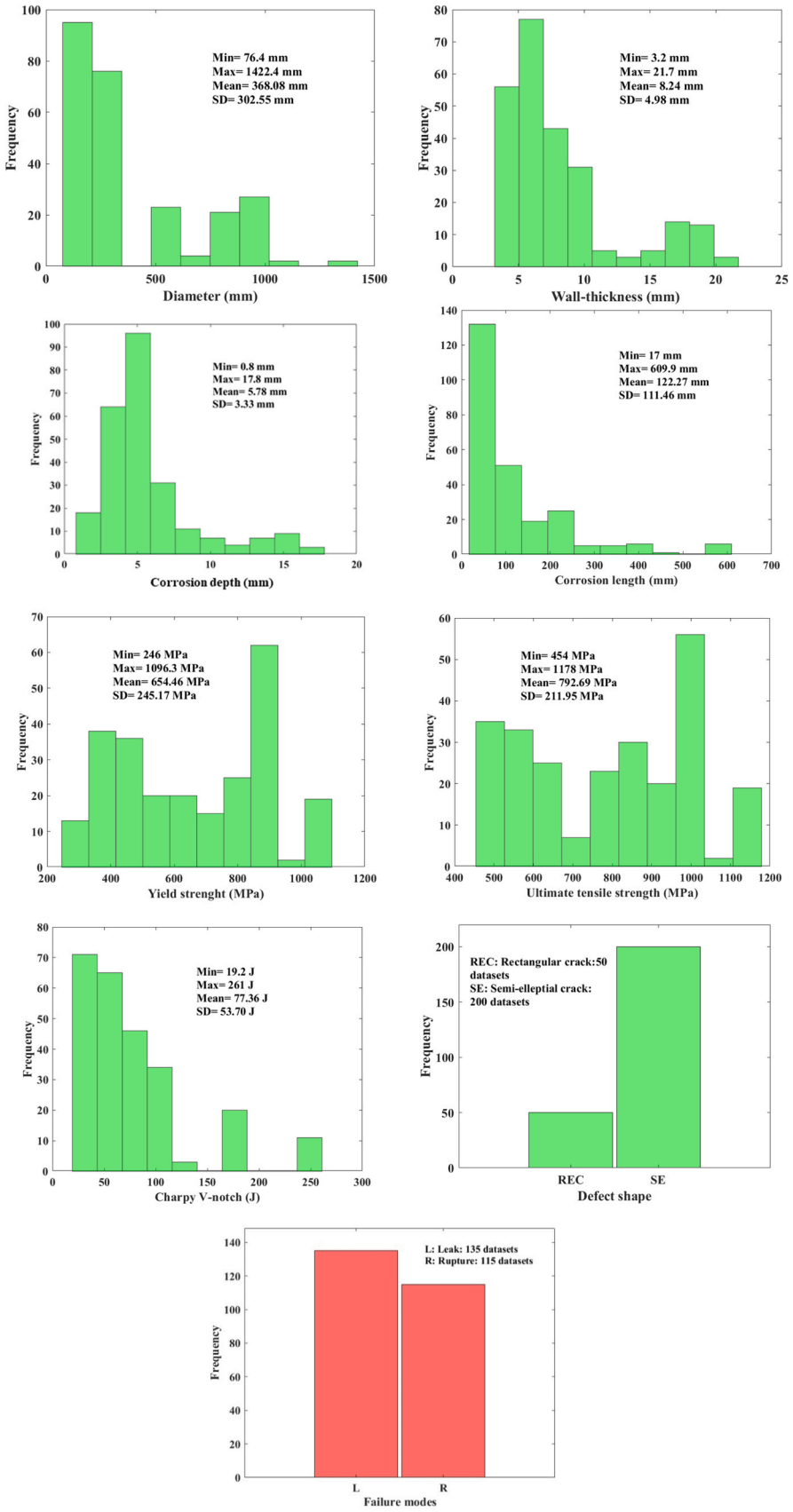


Fig. 1. Histograms of the input and output variables.

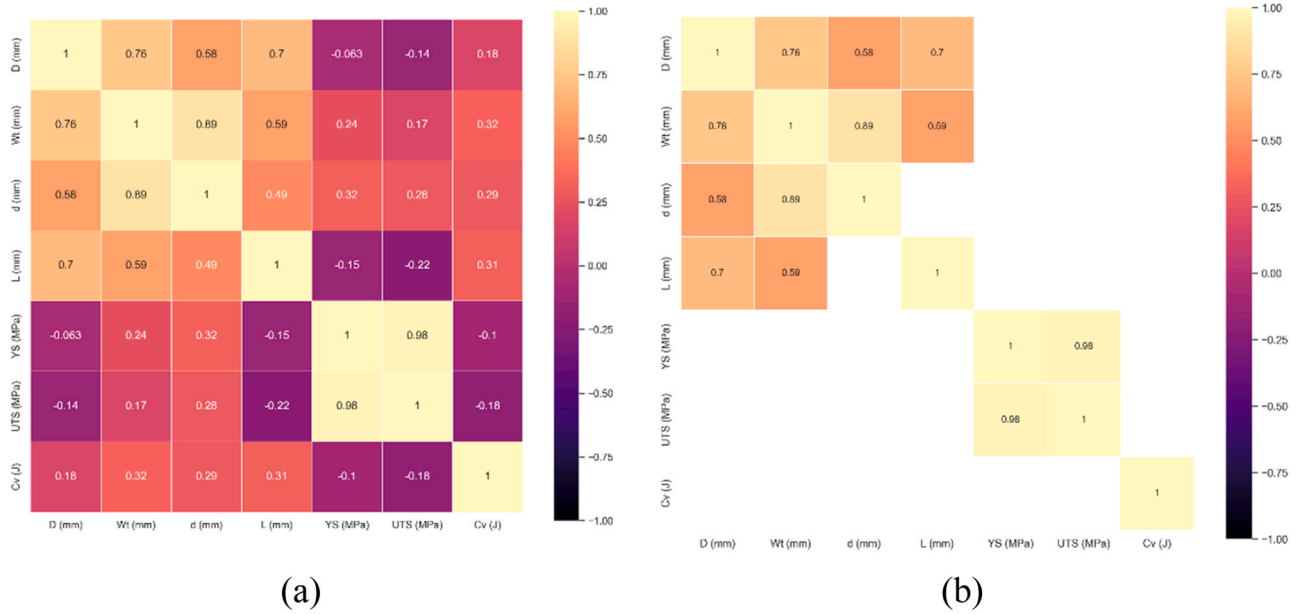


Fig. 2. Correlation matrix between the input variables (a) $r \in [-1, 1]$; (b) $r \in [-1, -0.5] \cup [0.5, 1]$.

imbalanced classes with respect to the corrosion/crack shape (i.e. SE: 200 samples and REC: 50 samples), a balancing process of the data was conducted using Synthetic Minority Oversampling Technique (SMOTE). This latter was first proposed by Chawla et al. [64] to handle unbalanced dataset classification concerns. This was done to prevent low classification accuracy when there was an unequal distribution of categories in a dataset- imbalance-. SMOTE creates synthetic cases as opposed to over-sampling in the traditional sense. To find the k neighbors, the Euclidean distance between each sample (x_l) in the minority class and the other samples in the same class must first be calculated. Next, from each of the k neighbors, a random sample (x_m) is selected. In the end, $x_n = x_l + (x_l - x_m)$ can be used to calculate the new sample, x_n, where λ is a random number in the range of 0–1. It should be noted that all of the developed codes for the presented framework were scripted in open-source Python platform. Thereafter, the complete datasets were randomly divided into 70 % to create the classification models (training phase), and the remaining 30 % to evaluate the prediction model's performance (testing phase). Consequently, it is thought that using the test datasets can reflect the model's performance as these datasets are unknown to the models.

In order to summarize the machine-learning model's performance, confusion matrix is used. This latter is a detailed table that compares the observed and predicted classes (i.e. failure modes), where the number of observations known to be in failure mode i but projected to be in failure mode j is equal to the number of elements in the confusion matrix (i.e. C_{ij} $i = 1{:}2$, $j = 1{:}2$). To explain simply, the diagonal elements in the confusion matrix indicate the failure modes successfully anticipated by the machine-learning model, whereas the off-diagonal elements represent the failure modes incorrectly predicted. The confusion matrix comprises three performance measures to evaluate the model's performance: accuracy, precision, and recall, which can be explained as follows:

1. The precision denotes the percentage of failure modes accurately assigned by the machine-learning model.
2. The recall denotes the percentage of real failure modes accurately assigned by the machine-learning model.
3. The accuracy is the ratio of the number of predicted failure modes correctly to the total failure mode predictions.

Among the three measures, the accuracy represent a global measure of the machine learning method's performance, whereas precision and recall are specific to each failure mode. A model with high accuracy, precision, and recall indicates that it can correctly identify the failure modes of oil and gas pipelines. Furthermore, the framework of the applied models is depicted in Fig. 4.

4.1. Machine learning performance for the identification of failure modes

Figs. 5 and 6 show the performance results based on the confusion matrix utilizing the eight machine learning models, from which the following points can discussed:

- The performance of the model varies depending on the failure mode under consideration, as shown in Figs. 5 and 6. Findings emphasize the significance of separating the data into training and testing sets, whereas training the machine learning model based only on the complete data sets may not produce sufficient results for unknown data (for example, the XGBoost model had 100 % accuracy using the training datasets, where 92 % accuracy during the testing phase).
- Based on the testing phase results, XGBoost had the highest accuracy (92 %), followed by KNN (84 %), and CatBoost (84 %). As a result, XGBoost model is recommended for identifying the failure mode of oil and gas pipelines subjected to corrosion/crack defects.
- The diagnosis of the failure mode of oil gas pipelines is a complicated non-linear problem, with the XGBoost model having 0.92 % recall and precision throughout the testing phase. Only DT has similar recall to XGBoost model but with very low precision.
- In general, gradient boosting approaches (such as XGBoost, LightGBM, and CatBoost) outperformed other machine learning models, with the exception of the KNN models. AdaBoost, DT, and Nave Bayes, on the other hand, produce the lowest performance (accuracy of 80 %).

Apart from the confusion matrix, the F1-score—which can be expressed by Eq.(3)—is also employed as an indication to evaluate the overall efficiency of the ML-models mentioned above.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

The F1-score results are illustrated in Fig. 7 and indicating both phases (i.e. training, testing and overall). In accordance with the

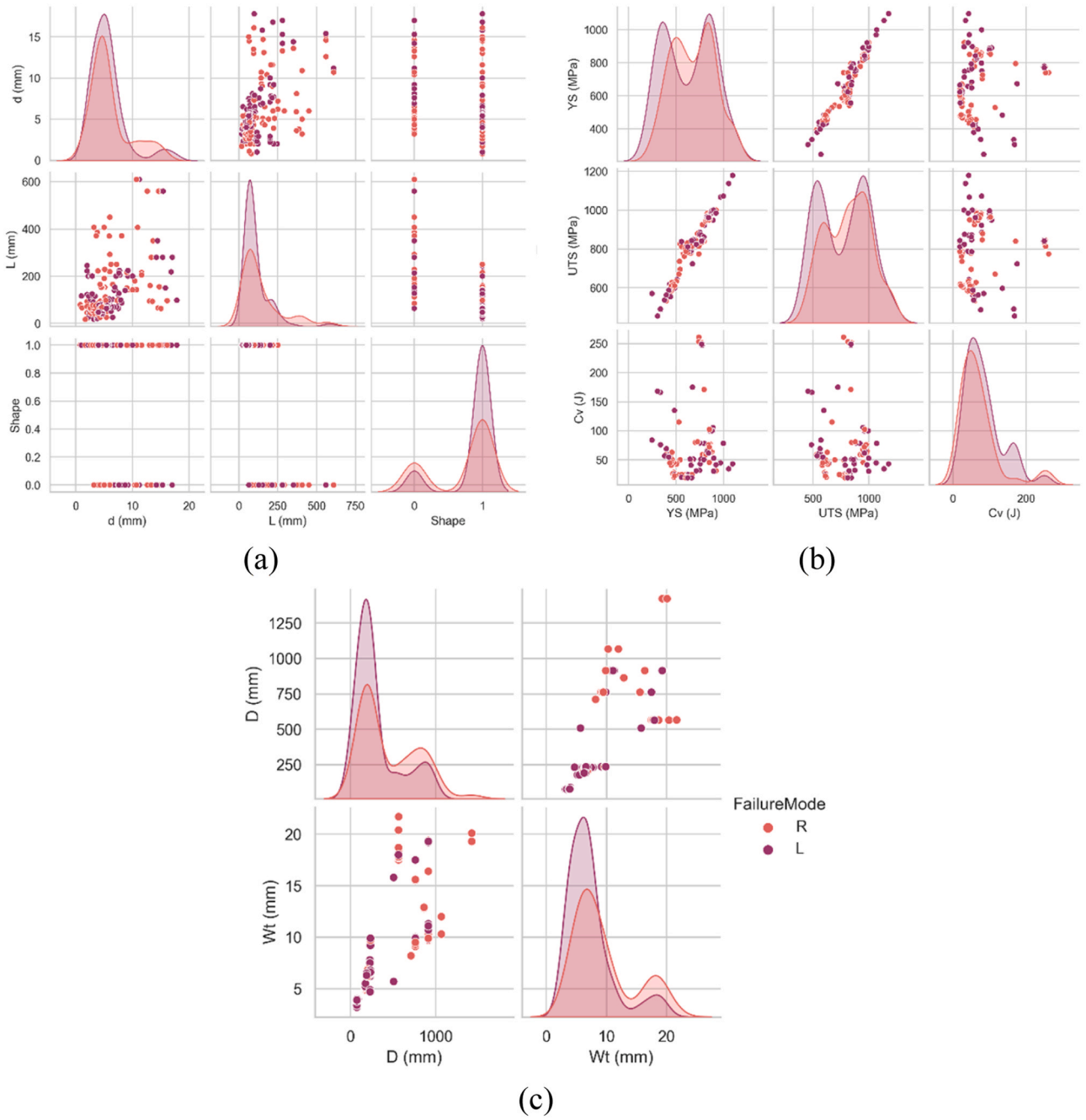


Fig. 3. Distribution of failure modes in relation to the input variables. (a) Failure mode compared to crack/corrosion characteristics, (b) failure mode compared to material properties, and (c) failure mode compared to pipeline design geometries.

findings, the boosting strategies outperform the other models, with XGBoost demonstrating the best performance among all the ML-models. In comparison to NB, KNN, DT, RF, AdaBoost, LightGMB, and CatBoost, respectively, the latter enhanced the classification results with 37.93 %, 14.2 %, 4.4 %, 3.9 %, 4.61 %, 5.45 %, and 5.12 %.

4.2. Identifying critical factors for corroded pipeline failure modes

In the final section of this study, a sensitivity analysis based on relevancy factor and SHAP values is performed to assess the importance of the input variables on failure mode identification using the best machine learning model from the previous section (i.e. XGBoost), which will aid in evaluating overall model performance. Fig. 8 depicts the relevancy factor results, which are used to observe the prioritized factors depending on their impact on the failure mode (i.e. leak and rupture). Because the defect shape is a categorical input variable, it has the least impact on the failure modes, and it was discovered that removing this factor had no effect on the results (Fig. 8(b)). Furthermore, the defect has a more complicated profile than the one stated in the database. Noting that all of the machine-models were re-run, and the same findings were obtained, hence this variable was removed during the SHAP analysis. Given the importance of corrosion depth and length in decreasing the safety levels of oil and gas pipelines [65–67], these two variables naturally demonstrated to be ranked first and second on Fig. 8 (a) and (b). Pipeline material properties exhibited a similar moderate response when compared to the wall thickness, which may be due to the latter's direct relationship to the failure mode mechanism (i.e., the ratio d/Wt is directly proportional to the strength of the pipe to withstand

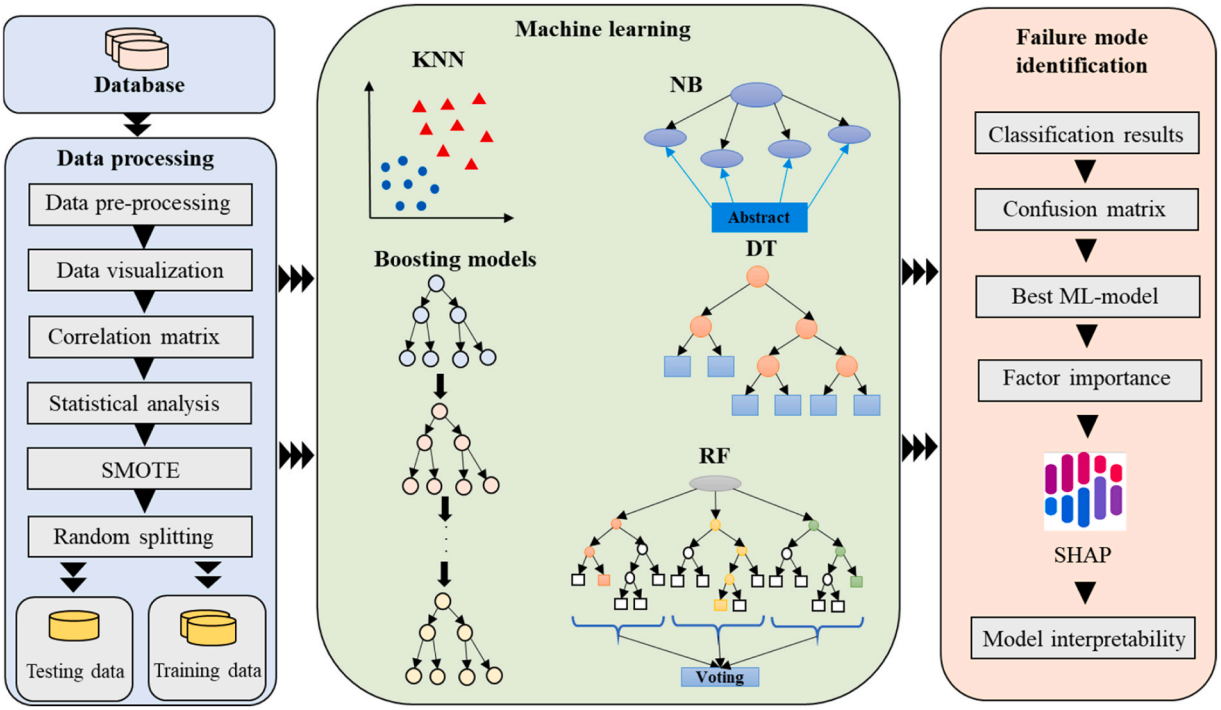


Fig. 4. Flowchart of the proposed machine learning based framework for failure mode identification of oil and gas pipelines.

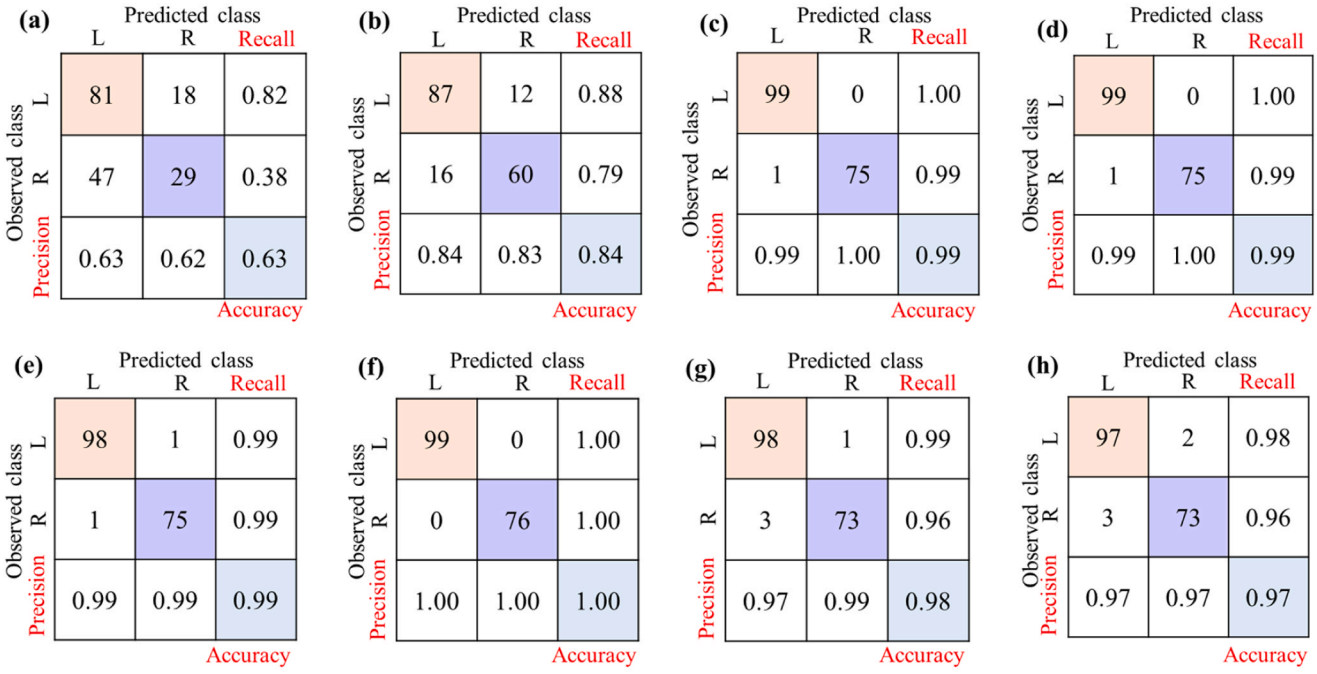


Fig. 5. ML-evaluation report based on confusion matrix during training phase: (a) NB, (b) KNN, (c) DT, (d) RF, (e) AdaBoost, (f) XGBoost, (g) LightGBM, and (h) CatBoost.

failure modes such as leak or rupture [68]).

The SHAP approach is used to explain the impact of the input variables on the developed XGBoost model for the identification of the failure modes for oil and gas pipelines. To begin, the global importance factor of the input variables is investigated and represented in Fig. 9. Given that SHAP values are path-independent, meaning that the significance values remain constant regardless of which features are deleted first, the values of the global important factor are therefore calculated by averaging the absolute Shapley values for each factor across the data. The input variable impact on the model XGBoost output is then sorted depending on its importance, from the highest mean absolute SHAP value to the lowest. Fig. 9, which provides insights into the influence of the input variables on each failure class (i.e. Leak and Rupture), shows that the impact is similar for all inputs, which can be attributed to the balanced data values of both failure mode classes. The global importance factor (Fig. 9(a)) indicates that the corrosion/crack depth (d, *mm*) has the most influence (i.e. mean|SHAP|= 3.7) during the development of the XGBoost model, followed by the pipe diameter (i.e.

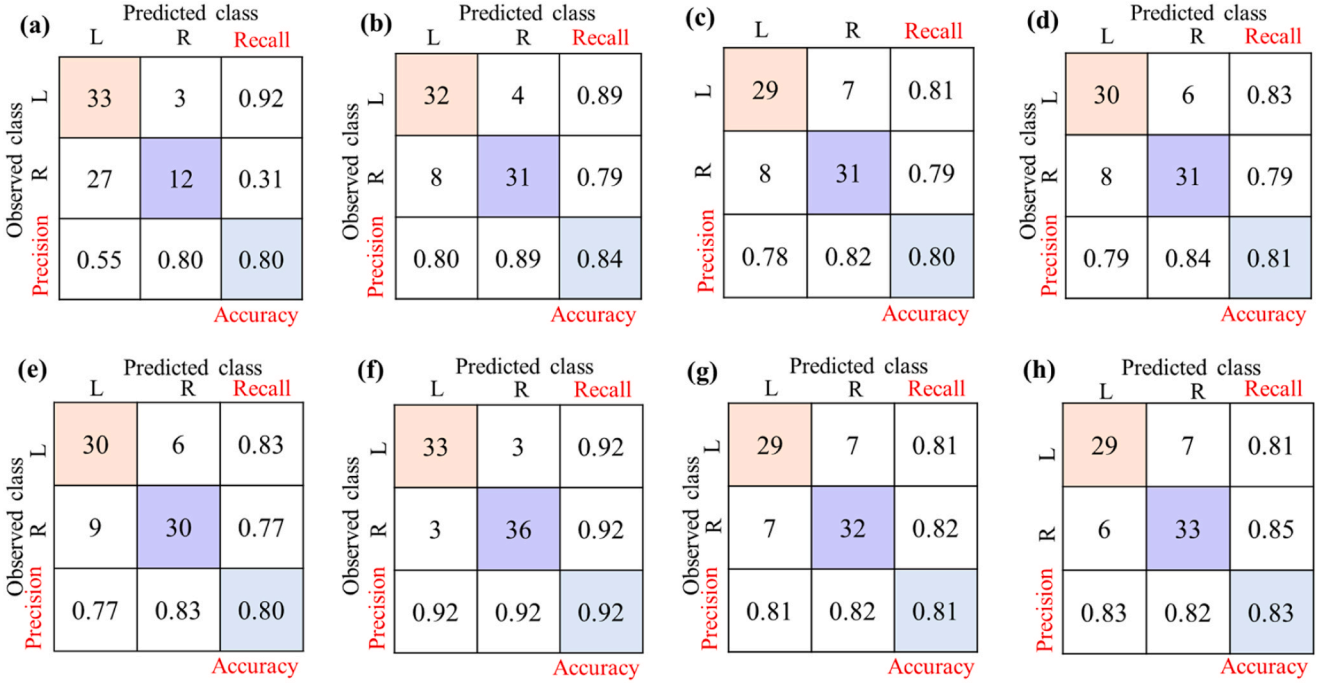


Fig. 6. ML-evaluation report based on confusion matrix during testing phase: (a) NB, (b) KNN, (c) DT, (d) RF, (e) AdaBoost, (f) XGBoost, (g) LightGBM, and (h) CatBoost.

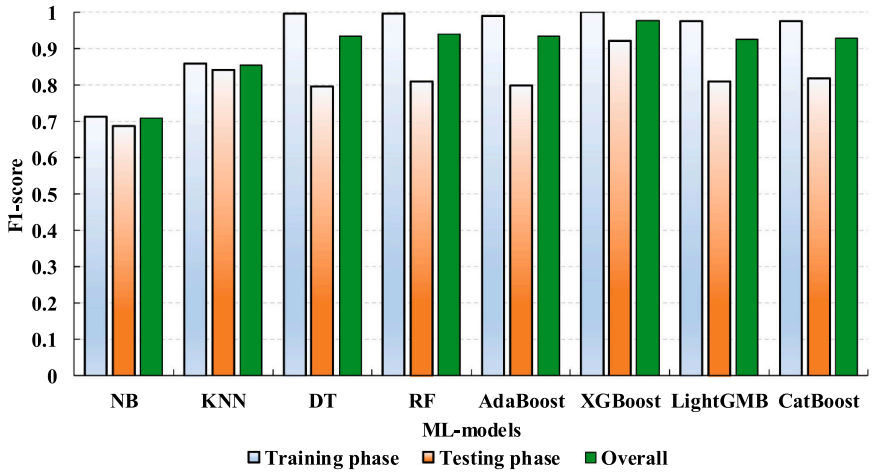


Fig. 7. Performance evaluation of the ML-models based on F1-score.

mean|SHAP|= 2.22), while Sharpy V-notch (Cv, *J*) has the least impact (i.e. mean|SHAP|= 0.2). However, the corrosion depth (d, *mm*) has mean|SHAP|= 1.99 for the leak failure mode and mean|SHAP|= 1.702 for the rupture failure mode, which are slightly closed, according to the individual (associated to the failure mode) important factor. The remaining variables exhibit this as well.

Fig. 10 (a) depicts a waterfall graphic that decomposes the influence of each input variable on the projected failure mode using SHAP values. The figure's base value represents the average value of the data, but because the presented problem is a classification problem, and the base value is negative (−0.359), it indicates that the plot refers to the predictive influence of the input variables on class 0 (i.e. leak). Consequently, SHAP waterfall plot shown in Fig. 10 (a) illustrates how each variable contributes to the final output model; the red and blue arrows in Fig. 10 denote positive (or towards leak) and negative (or towards rupture) contributions, respectively. The final prediction *f(x)* is simply the sum of the SHAP values plus the base value, which indicates the final prediction probability, in which the prediction probability of the failure mode is greater than 95 % after using the logistic function $\left(\frac{1}{1+e^{x}}\right)$, indicating a high potential for leak failure mode identification. Except for the corrosion/crack defect depth and length, most of the variables contribute to the prediction of the leak failure mode, which is to be expected given that corrosion depth and length can have a significant impact on the safety of oil and gas pipelines and produce rupture failure mode. Fig. 10 (b) depicts a beeswarm plot to analyze how each SHAP value for each variable contributes to the failure mode. This summary figure shows the range and distribution of input variables effect on failure mode prediction. Each point on the plots represents a Shapley value for the input variables and an instance, with the y-axis representing the order of importance of the input variables. The dots are expected to be colored based on the input variable's value, ranging from low (blue) to high (red), while the density shows the distribution of points in the dataset. It is obvious that the depth of the corrosion/crack

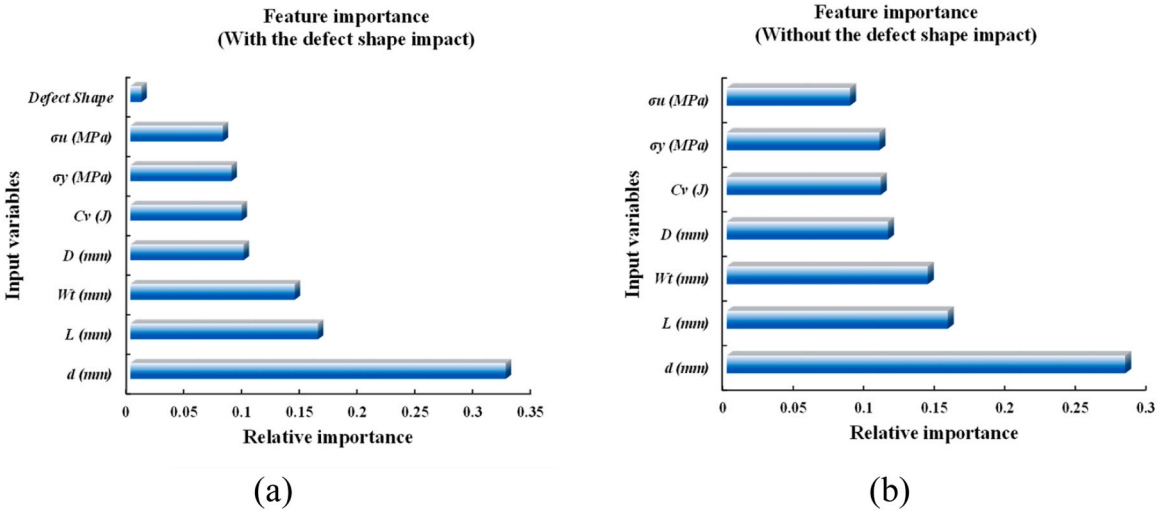


Fig. 8. Feature importance of the input variables. (a) including the defect shape effect; (b) excluding the defect shape effect.

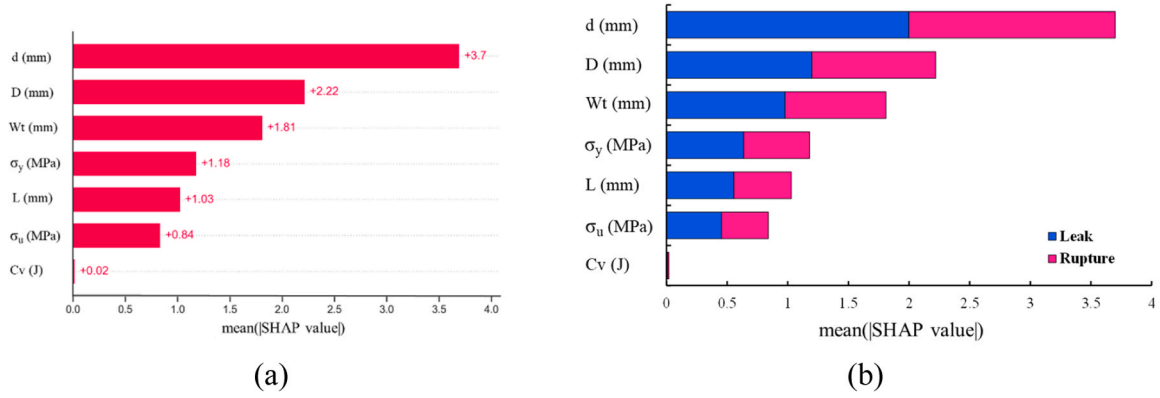


Fig. 9. Global importance factor using SHAP based on XGBoost model. (a) For each input variable, (b) for each variable and each failure mode.

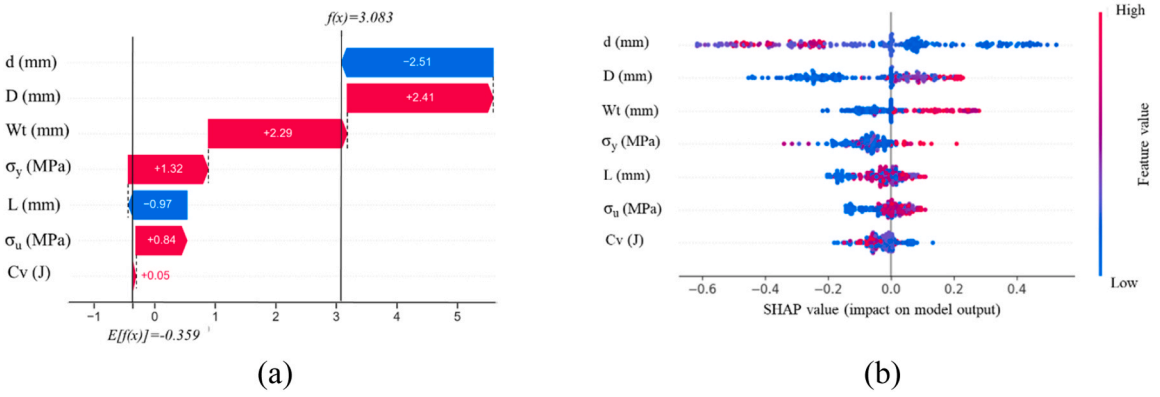


Fig. 10. SHAP based explanation of the XGBoost model (a) waterfall plot, (b) beeswarm plot.

defect has the most impact on failure mode identification, where the SHAP values indicate that any value of this variables can significantly contribute to the failure appearance. It is also worth noting that the Sharpy V-notch (Cv, *J*), ultimate tensile strength (σ_u, *MPa*) and defect length (L, *mm*) and has a relatively low-centered influence in comparison to the other factors, in the contribution for the model development.

5. Conclusions

Oil and gas pipelines that are subjected to a range of loads and sever conditions over time will acquire corrosion defects on their surface, which can lead to either a rupture or a leak. This study investigated explainable machine learning models' ability to identify the failure modes of corrosion- and crack-related pipeline failures. This is accomplished by using a large database comprising 250 examples of pipe-burst test, where these steel pipes contain manufactured corrosion defects.

The latter have two distinct shapes: rectangular (50 samples) and semi-elliptic (200 samples). In total, eight variables are considered as input parameters (i.e. The pipe design geometry, material properties, defect geometries, and shape), whereas the target variable is the failure modes. In order to address this issue, the performance of eight machine learning models (i.e. NB, KNN, DT, RF, AdaBoost, XGBoost, LightGBM, and CatBoost) are examined. The performance evaluation of these machine-learning models is assessed using the F1-score metric and confusion matrix that includes three metrics: global accuracy, precision, and recall. The study's findings demonstrated that XGBoost, followed by KNN and CatBoost models indicated the highest accuracy during both stages (training and testing phases). For the testing phase, the suggested XGBoost model exhibited a 92 % overall accuracy in determining the pipeline failure modes. According to the sensitivity analysis and the interpretability of the XGBoost model based on feature importance analysis and SHAP values revealed that the corrosion depth is the most crucial parameter that contributes to the failure mode in oil and gas pipelines, followed by the corrosion defect length, and pipe design geometries (D and Wt). The corrosion defect geometries were discovered to favor rupture over the leak, despite the fact that an increase value can produce both failure modes. The rest of the variables demonstrate that an increasing can have a favorable impact on pipeline safety.

The overall results indicate that the supplied data are insufficient to investigate the impact of defect shape on failure mode identification. Identifying the failure mode of oil and gas pipelines, on the other hand, is a difficult task that includes numerous variables and mechanisms, and establishing decision boundaries between these failure modes failure requires a comprehensive database with much larger datasets and variation to include all scenarios and ranges. Field engineering can benefit from the developed models in a framework of an integrated in an easy-to-use interface/software that can give an initial assessment of the severity of the corrosion defects based on the provided inspection reports, which will help interventions such as the appropriate selection of maintenance and repair actions in order to increase structural efficiency.

CRediT authorship contribution statement

Mohamed El Amine Ben Seghier: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft. **Osama Ahmed Mohamed:** Investigation, Methodology, Project administration, Validation, Writing – original draft. **Hocine Ouaer:** Formal analysis, Validation, Visualization, Writing – original draft.

Declaration of Competing Interest

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

Acknowledgements

Ossama Mohamed acknowledges financial support from Abu Dhabi University's Office of Research and Sponsored Programs. Grant number: 19300803.

References

[1] Sim S, Cole IS, Choi YS, Birbilis N. A review of the protection strategies against internal corrosion for the safe transport of supercritical CO_2 via steel pipelines for CCS purposes. Int J Greenh Gas Control 2014;29:185–99. https://doi.org/10.1016/j.ijggc.2014.08.010.

[2] Vanaei HR, Eslami A, Egbewande A. A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. Int J Press Vessel Pip 2017;149:43–54. https://doi.org/10.1016/j.ijpvp.2016.11.007.

[3] Li X, Wang J, Chen G. A machine learning methodology for probabilistic risk assessment of process operations: a case of subsea gas pipeline leak accidents. Process Saf Environ Prot 2022;165:959–68.

[4] Ben Seghier MEA, Mustaffa Z, Zayed T. Reliability assessment of subsea pipelines under the effect of spanning load and corrosion degradation. J Nat Gas Sci Eng 2022:104569.

[5] Li X, Zhang Y, Abbassi R, Khan F, Chen G. Probabilistic fatigue failure assessment of free spanning subsea pipeline using dynamic Bayesian network. Ocean Eng 2021;234:109323.

[6] Qin G, Huang Y, Wang Y, Cheng YF. Pipeline condition assessment and finite element modeling of mechano-electrochemical interaction between corrosion defects with varied orientations on pipelines. Tunn Undergr Sp Technol 2023;136:105101.

[7] Farh HMH, Ben Seghier MEA, Zayed T. A comprehensive review of corrosion protection and control techniques for metallic pipelines. Eng Fail Anal 2022:106885.

[8] Wang Y, Li R, Xia A, Ni P, Qin G. An integrated modeling method of uncertainties: application-orientated fuzzy random spatiotemporal analysis of pipeline structures. Tunn Undergr Sp Technol 2023;131:104825.

[9] Hussein Farh HM, Ben Seghier MEA, Taiwo R, Zayed T. Analysis and ranking of corrosion causes for water pipelines: a critical review. Npj Clean Water 2023;6:65.

[10] Papavinasam S, Revie RW, Friesen WI, Doiron A, Panneerselvam T. Review of models to predict internal pitting corrosion of oil and gas pipelines. Corros Rev 2006;24:173–230.

[11] Lam C, Zhou W. Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. Int J Press Vessel Pip 2016;145:29–40. https://doi.org/10.1016/j.ijpvp.2016.06.003.

[12] Han Z, Li X, Zhang R, Yang M, Ben Seghier MEA. A dynamic condition assessment model of aging subsea pipelines subject to corrosion-fatigue degradation. Appl Ocean Res 2023;139:103717.

[13] Zhou W, Xiang W, Cronin D. Probability of rupture model for corroded pipelines. Int J Press Vessel Pip 2016;147:1–11. https://doi.org/10.1016/j.ijpvp.2016.10.001.

[14] Xie M, Tian Z. Risk-based pipeline re-assessment optimization considering corrosion defects. Sustain Cities Soc 2018;38:746–57.

[15] Asme-B31G. Manual for determining the remaining strength of corroded pipelines. Am Soc Mech Eng 2004:1–64. https://doi.org/10.1520/G0154-12A.

[16] Anderson TL. Development of a modern assessment method for longitudinal seam weld cracks. Cat. No. PR-460-134401 Pipeline Res Counc Inc 2016;95.

[17] T.L. Anderson, Assessing crack-like flaws in longitudinal seam welds: a state-of-the-art review, pipeline Res. Counc. Int. Cat. No. PR-460–134506-R02. Available from: ⟨Https//www.Prci.org/Research/DesignMaterialsConstruction/2774/MAT-8/4537/119445.Aspx⟩; 2017.

[18] C.E. Jaske, J.A. Beavers, Integrity and remaining life of pipe with stress corrosion cracking. Pr. Report, Cat; 2001.

[19] S.J. Polasik, C.E. Jaske, T.A. Bubenik, Review of engineering fracture mechanics model for pipeline applications. In: Proceedings of the Int. Pipeline Conf., American Society of Mechanical Engineers; 2016: p. V001T03A038.

[20] Shannon RWE. The failure behaviour of line pipe defects. Int J Press Vessel Pip 1974;2:243–55.

[21] W.A. Maxey, J.F. Kiefner, R.J. Eiber, A.R. Duffy, Ductile fracture initiation, propagation, and arrest in cylindrical vessels. Conshohocken, PA: ASTM International West; 1972.

[22] Kiefner JF. Modified equation aids integrity management. Oil Gas J 2008;106:78–82.

[23] Kiefner JF. Modified Ln-Secant equation improves prediction. Oil Gas J 2008;106:64–6.

[24] B.N. Leis, N.D. Ghadiali, Pipe Axial Flaw Failure Criteria (PAFFC): Version 1.0 users manual and software. Columbus, OH (United States): Battelle Memorial Inst.; 1994.

[25] Ben Seghier MEA, Knudsen OØ, Skilbred AWB, Höche D. An intelligent framework for forecasting and investigating corrosion in marine conditions using time sensor data. Npj Mater Degrad 2023;7:91.

[26] Keshtegara B, el M, Ben Seghier A. Modified response surface method basis harmony search to predict the burst pressure of corroded pipelines. Eng Fail Anal 2018;89:177–99. https://doi.org/10.1016/j.engfailanal.2018.02.016.

[27] Taiwo R, Zayed T, Ben Seghier MEA. Integrated intelligent models for predicting water pipe failure probability. Alex Eng J 2024;86:243–57.

[28] Taiwo R, Yussif A-M, Ben Seghier MEA, Zayed T. Explainable ensemble models for predicting wall thickness loss of water pipes. Ain Shams Eng J 2024:102630.

[29] M.E.A. Ben Seghier, B. Keshtegar, M. Taleb-Berrouane, R. Abbassi, N.-T. Trung, Advanced intelligence frameworks for predicting maximum pitting corrosion depth in oil and gas pipelines. Process Saf. Environ. Prot; n.d.

[30] Rachman A, Zhang T, Ratnayake RMC. Applications of machine learning in pipeline integrity management: a state-of-the-art review. Int J Press Vessel Pip 2021;193:104471.

[31] Soomro AA, Mokhtar AA, Kurnia JC, Lashari N, Lu H, Sambo C. Integrity assessment of corroded oil and gas pipelines using machine learning: a systematic review. Eng Fail Anal 2022;131:105810.

[32] Liu Y, Bao Y. Review on automated condition assessment of pipelines with machine learning. Adv Eng Inform 2022;53:101687.

[33] Seghier M, Keshtegar B, Correia J, Jesus ADe, Lesiuk G. Structural reliability analysis of corroded pipeline made in X60 steel based on M5 model tree algorithm and Monte Carlo simulation. Procedia Struct Integr 2018;13:1670–5.

[34] Mazumder RK, Salman AM, Li Y. Failure risk analysis of pipelines using data-driven machine learning algorithms. Struct Saf 2021;89:102047.

[35] Sun H, Zhou W. Classification of failure modes of pipelines containing longitudinal surface cracks using mechanics-based and machine learning models. J Infrastruct Preserv Resil 2023;4:1–25.

[36] Mangalathu S, Heo G, Jeon J-S. Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes. Eng Struct 2018;162:166–76.
[37] Mangalathu S, Jeon J. Stripe-based fragility analysis of multispan concrete bridge classes using machine learning techniques. Earthq Eng Struct Dyn 2019;48:1238–55.
[38] Wu Y, Zhou Y. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. Constr Build Mater 2022;330:127298.
[39] Esteghamati MZ, Gernay T, Banerji S. Evaluating fire resistance of timber columns using explainable machine learning models. Eng Struct 2023;296:116910.
[40] Berrar D. Bayes' theorem and naive Bayes classifier. Encycl Bioinforma Comput Biol ABC Bioinforma 2018;403:412.
[41] Leung KM. Naive bayesian classifier. Polytech Univ Dep Comput Sci Risk Eng 2007;2007:123–56.
[42] Peterson LE. K-nearest neighbor. Scholarpedia 2009;4:1883.
[43] Laaksonen J, Oja E. Classification with learning k-nearest neighbors. Proc Int Conf Neural Netw, IEEE 1996:1480–3.
[44] Swain PH, Hauska H. The decision tree classifier: design and potential. IEEE Trans Geosci Electron 1977;15:142–7.
[45] D. Witten, G. James, An introduction to statistical learning with applications in R, springer publication; 2013.
[46] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 1991;21:660–74.
[47] Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens 2005;26:217–22.
[48] Ben Seghier MEA, Höche D, Zheludkevich M. Prediction of the internal corrosion rate for oil and gas pipeline: implementation of ensemble learning techniques. J Nat Gas Sci Eng 2022:104425.
[49] Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 2012;67:93–104.
[50] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55:119–39.
[51] Meir R, Rätsch G. An introduction to boosting and leveraging. Adv Lect Mach Learn Mach Learn Summer Sch 2002 Canberra, Aust Febr 11–22, 2002 Revis Lect, Springe 2003:118–83.
[52] Ridgeway G. The state of boosting. Comput Sci Stat 1999:172–81.
[53] Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the twenty second Acm Sigkdd Int Conf Knowl Discov Data Min 2016:785–94.
[54] Zhang D, Qian L, Mao B, Huang C, Huang B, Si Y. A data-driven design for fault detection of wind turbines using random forests and XGboost. Ieee Access 2018;6:21020–31.
[55] Jiang H, He Z, Ye G, Zhang H. Network intrusion detection based on PSO-XGBoost model. IEEE Access 2020;8:58392–401.
[56] Zhang D, Gong Y. The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. IEEE Access 2020;8:220990–1003.
[57] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30.
[58] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018;31.
[59] Luo M, Wang Y, Xie Y, Zhou L, Qiao J, Qiu S, et al. Combination of feature selection and catboost for prediction: the first application to the estimation of aboveground biomass. Forests 2021;12:216.
[60] Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data 2020;7:1–45.
[61] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
[62] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles. arXiv Prepr. arXiv1802.03888; 2018.
[63] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–67.
[64] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–57.
[65] Taiwo R, Ben Seghier MEl Amine, Zayed T. Towards sustainable water infrastructure: the state-of-the-art for modeling the failure probability of water pipes. Water Resour Res 2023:e2022WR033256.
[66] Guillal A, Ben Seghier MEA, Nourddine A, Corria JAFO, Mustaffa ZB, Trung N-T. Probabilistic investigation on the reliability assessment of mid-and high-strength pipelines under corrosion and fracture conditions. Eng Fail Anal 2020:104891.
[67] Keshtegar B, El M, Ben A, Zhu S, Abbassi R, Trung N. Reliability analysis of corroded pipelines: novel adaptive conjugate first order reliability method. J Loss Prev Process Ind 2019;62:103986. https://doi.org/10.1016/j.jlp.2019.103986.
[68] Idris NN, Mustaffa Z, Ben Seghier MEA, Trung N-T. Burst capacity and development of interaction rules for pipelines considering radial interacting corrosion defects. Eng Fail Anal 2021;121:105124.