IJMT. Vol. 22, Iss. 1, 2026. (14-31)

International Journal of Maritime Technology

Journal homepage: ijmt.ir

Machine Learning Models Development to Predict Corroded Pipeline Behavior Considering Defects Interaction

Soheyl Hosseinzadeh[1]*, Mohammad Reza Bahaari[1], Mohsen Abyani[1]

[1] *School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran.*

ARTICLE INFO

Article History:
Received: 19 Aug 2025
Last modification: 7 Oct 2025
Accepted: 8 Oct 2025
Available online: 9 Oct 2025

Article type:
Research paper

Keywords:
Offshore Pipeline Engineering
Pipeline Integrity Management
Structural Reliability
Random Sampling
Latin Hypercube Sampling

ABSTRACT

Internal corrosion poses a significant risk to offshore pipeline operations. This study aims to utilize a combination of the Finite Element Method (FEM) and Latin Hypercube Sampling (LHS) to create a database of structural response data for corroded pipelines experiencing longitudinally interacting internal corrosion defects under internal and external pressure loading. The database includes input data such as pipeline geometry parameters, pipeline material data, corrosion defect data and loading data. This generated database will be utilized to train various advanced machine learning (ML) models to develop a predictive model capable of estimating the Maximum von Mises Stress occurring in the outermost mesh layer of a mesh ligament within the thickness of the corroded pipeline at the defected area. Such predictive capabilities of the ML model will enhance the ability to forecast leakage based on pipeline and defect specifications, thereby saving costs and time. To achieve the optimal model, various ML algorithms have been compared. Finally, to assess the prediction accuracy of the models, results of models were compared and evaluated.

ISSN: 2645-8136

DOI:

1. Introduction

Despite the rapid expansion of renewable energy, oil and gas remain the primary sources of energy, and offshore pipelines play a crucial role in ensuring a reliable energy supply [1]. Corrosion failures in offshore pipelines transporting marine oil are becoming more frequent. Internal inspections frequently reveal substantial corrosion defects. These defects can appear in different forms, including deposit corrosion, cavitation corrosion, uniform corrosion, and pitting corrosion [2]. Among these, internal pitting corrosion, mainly caused by carbon dioxide (CO2) and hydrogen sulfide (H2S), is identified as a major failure mechanism in these pipelines [3]. The depth of corrosion defects, especially pitting, has a significant impact on the failure pressure of the pipelines [4]. As artificial intelligence (AI) progresses rapidly, data-driven models utilizing machine learning (ML) algorithms have shown significant flexibility in managing high-dimensional data and intricate operating conditions. This capability is especially valuable for corrosion prevention and fault diagnosis in pipelines and equipment [5].

In this regard, numerous efforts have been undertaken to evaluate how the corrosion specifications impact the reliability and integrity of pipelines. For instance, Zhang [6] revealed that corrosion depth significantly affects the condition of pipelines during landslide events, with a notable impact on stress levels. The study identified a complex relationship between maximum stress, the position facing the landslide, and the location of corrosion around the pipe. Moreover, it was demonstrated that the axial placement of corrosion in relation to the landslide center affects the distribution of stress within the pipeline. Wang et al. [7] provided a detailed analysis of theoretical methods used to assess the remaining strength of seawater pipelines with corrosion defects. The study compared different evaluation results across various pipeline parameters and corrosion conditions, underscoring the importance of these methods for understanding how corrosion affects pipeline integrity and lifespan. Yang et al. [8] examined the reliability of gas pipelines with corrosion defects by employing ASME-B31G revised criteria alongside finite element numerical analysis methods. The study forecasts the residual strength and lifespan of pipelines, offering important insights for maintaining the safe and stable operation of natural gas pipelines. Cheng [9] highlighted the essential role of corrosion modeling in evaluating and forecasting pipeline failure and managing related risks. The complexity of pipeline corrosion stems from the interplay of multiple reactions and processes unique to the material and environmental conditions, affecting pipeline integrity. In a recent study by the same researchers, a new method called "IPA" (Internal Pressure Assessment) was introduced to evaluate the reliability of offshore pipelines with internal corrosion defects [10]. This approach is based on the probabilistic Incremental Dynamic Analysis (IDA) method developed by Vamvatsikos and Cornell [11]. However, the study focused solely on single defects and did not address the effects of defect interactions.

Although the higher risk associated with interacting corrosion defects is well recognized, it has not been extensively studied, with most research concentrating on single defects. Nevertheless, some studies have shown that interactions between two defects can affect pipelines significantly. Mustaffa et al. [12] introduced a method for assessing the reliability of corroded pipelines arranged in series, focusing on how length-scale effects impact pipeline integrity. The study emphasized the statistical correlation between corrosion features across different segments of the pipeline, noting that failure in one section can influence adjacent sections. By including a correlation distance parameter to address these effects, the research demonstrated that accounting for length-scale effects leads to a higher probability of failure for corroded pipelines compared to analyses that do not consider these factors. Kuppusamy et al. [13] examined the effect of interacting corrosion defects on the buckling strength of pipelines using numerical analysis. The study highlights how corrosion characteristics can impact the structural integrity of pipelines. Xie et al. [14] explored the interaction between corrosion and crack defects in pipelines and its effect on fatigue crack growth. The study introduced a crack propagation method to predict how corrosion influences fatigue cracks. Arumugam et al. [15] investigated the burst capacity of pipelines with multiple longitudinally aligned interacting corrosion defects using FEM. The study examined the effects of internal pressure and axial compressive stress on pipelines with two or three longitudinal internal defects. Moreover, Kuppusamy et al. [13] explored how corrosion defects in close proximity can interact, leading to a decrease in the overall strength of a pipeline, a phenomenon referred to as interacting corrosion defects. The study focused on the buckling strength of corroded pipelines with such interacting defects, offering a numerical analysis to better understand the effects on structural stability and dynamics in materials science. Zhang et al. [16] emphasized the need to consider the interaction between adjacent defects when evaluating the failure pressure of pipelines with corrosion clusters. The study introduced a "center failure location" method that focuses on the failure position of corrosion clusters, accounting for how adjacent defects affect the failure pressure of the central defect. Mondal and Dhar [17] highlighted that the proximity of interacting corrosion defects significantly affects the burst pressure of pipelines. Factors such as pipe wall thickness, corrosion depth, and defect locations are crucial in

determining the critical spacing for these interactions. Abyani and Bahari [18] examined the impact of correlations between random variables in adjacent components on the time-variant system reliability of corroded pipelines. They addressed the challenge of non-positive definiteness in the multivariate target correlation matrix by applying optimization techniques to transform it into a positive definite matrix. In a recent study by the authors, internal corrosion was identified as a significant threat to offshore pipeline services. The study evaluated the reliability of 32" oil and gas offshore pipelines with internal corrosion considering the interaction of longitudinally aligned defects interaction. The findings underscored that multiple corrosion defects pose a greater risk to offshore pipelines, potentially leading to more severe consequences than a single defect alone, highlighting the importance of considering defect interactions in reliability assessments.

Identifying and predicting pipeline failures is crucial for ensuring pipeline safety and integrity. Numerous studies have been carried out to develop techniques for estimating failure pressure, residual strength, and the probability of failure in corroded pipelines. Ossai et al. [19] developed a predictive model for internal pitting corrosion in aged non-piggable pipelines, incorporating various operational parameters. Using ten years of Ultrasonic Thickness Measurement (UTM) data, the study analyzed maximum pit depths in relation to factors such as temperature, CO2 partial pressure, flow rate, pH, sulfate ions, chloride ions, and water cut. The model's accuracy was validated with field data from twelve pipelines, demonstrating its effectiveness in predicting maximum pitting rates. This is essential for evaluating the strength and integrity of corroded pipelines and ensuring operational safety. Nizamani et al. [20] evaluated how corrosion affects the structural strength of offshore pipelines, focusing on assessing remaining strength to extend pipeline lifespan. The study employed Bayesian updating to assess the probability of failure and compared results from burst tests, including a sensitivity analysis of variables such as defect depth and thickness. additionally, Hou et al. (2019) performed a non-probabilistic time-varying reliability analysis of corroded pipelines, taking into account the interaction of various uncertainty variables [21]. The research conducted by Cui [22] introduced a management system for predicting corrosion failures in carbon steel oil and gas pipelines, improving the precision and effectiveness of corrosion failure predictions and control. This work advances the field by providing a model that simplifies the prediction process, supporting proactive maintenance for underwater oil and gas transport pipelines. Moreover, Colindres et al. [23] investigated the mechanical responses of various pipelines subjected to different corrosion types, such as external, internal, and combined defects. The study demonstrated that finite element models provided more accurate predictions of failure pressures in corroded pipelines compared to traditional methods like B31G and DNV-99. Abyani et al. [24] conducted a study on the failure pressure and remaining lifespan of corroded offshore pipelines by applying various code-based methods, including ASME B31G, modified ASME B31G, DNV RP-F101, and FFS Level-1.

In the contemporary era, ML tools are considered as decisive elements in costly industries and can play a supportive role to significantly improve the prediction of pipeline behavior, fostering heightened reliability and enabling proactive maintenance strategies. To predict the potential failure of corroded pipelines using machine learning (ML) tools, it is essential to review the application of ML in pipeline integrity management systems (PIMS). A comprehensive review of ML applications in pipeline integrity management is provided by Rachman et al. [5], which is directly related to the task. In a similar effort, a review study by Soomro et al. highlighted the essential role of ML in assessing hydrocarbon fluid integrity in oil and gas pipelines, particularly given the severe impacts of corrosion. Their study emphasizes that unlike conventional deterministic and probabilistic models, ML techniques such as ANN, Support Vector Machine (SVM), and hybrid models have been better suited to handle the complex nature of pipeline degradation. The study offers an extensive evaluation of current ML methods, datasets, and variables, and proposes future directions for researchers and practitioners to improve pipeline integrity assessment. [25]. Various facets of integrity management, encompassing the utilization of ML for predictive analysis and risk assessment in pipeline systems, are covered in the review. Additionally, a data-driven ML approach for corrosion risk assessment is presented by Ossai [26], emphasizing the efficacy of ML techniques, such as ANN, in forecasting future states of corrosion defect depth growth. Valuable insights into the application of machine learning specifically for corrosion risk assessment, pertinent to the task of predicting failure in corroded pipelines, are offered by this study. Data-driven methods, including ANN and SVM, for predicting the burst strength of corroded line pipelines subjected to internal pressure are discussed by Cai et al. [27]. This research, addresses the prediction of pipeline strength, crucial for assessing the failure likelihood of corroded pipelines. The study provides a practical application of ML tools for predicting the strength of corroded pipelines, contributing to failure prediction in such systems. ML data analytics based on distributed fiber sensors for pipeline feature detection are the focus of Zhang et al. [28], aiming to enhance ML algorithms for the detection and size prediction of major pipeline structural changes and corrosion types. This research is relevant as it underscores the use of ML for detecting and predicting corrosion types, fundamental for failure

prediction in corroded pipelines. In a recent research, Abyani et al. employed ML techniques, specifically Gaussian Process Regression (GPR) and Multi-Layer Perceptron (MLP), to predict the failure pressure of corroded offshore pipelines, demonstrating their remarkable performance compared to other models, while also observing that Maximum Von Mises Stress (MVMS) of the pipeline increases with water depth at low levels of IP but decreases at high IP levels [29]. Although there are other recent researches related to ML and pipeline management, such as [30], [31], and [32] which have furnished valuable insights and methodologies for utilizing ML to address the challenges of predicting failure in corroded pipelines. This paper has aimed to use generated FEM analysis big data to train a ML model in order to predict the behavior of corroded pipelines having pipeline and defects data including the defect interaction effect.

Subsequently, pipelines as key conduits transporting hydrocarbons transportation play a key role in the oil and gas industry. Offshore pipelines particularly, are vital due to their vulnerability to corrosive and severe conditions. Internal corrosions are the primary cause of failure for the offshore pipelines, posing a significant threat to their reliability and performance. It's probable that Corrosion defects interact when they are in proximity which intensify the damage, making pipelines much weaker than those with isolated defects. To ensure safe and efficient operations and to predict pipeline failures, it is essential to estimate the behavior of offshore pipelines affected by interacting defects. This research establishes a novel, data-driven framework by integrating Latin Hypercube Sampling, advanced finite element analysis, and machine learning to develop a high-fidelity predictive model for the structural response of pipelines with interacting corrosion defects. The final outcome is a robust and computationally efficient machine learning tool capable of accurately forecasting pipeline reliability, providing a significant contribution to the field of pipeline integrity management that supports proactive and informed decision-making.

2. Methodology

The primary objective of this research is to construct a model for forecasting the likelihood of failure and the Maximum Von Mises Stress (MVMS) in pipelines featuring two interacting, longitudinally aligned corrosion flaws. Figure 1 illustrates the proposed methodological workflow, with each stage detailed below.

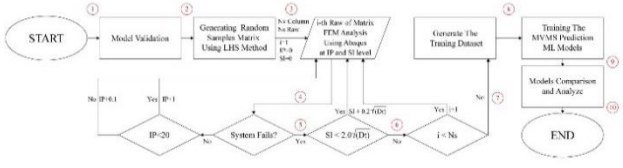


Figure 1. Diagram of the suggested approach

Step 1: FEA results were validated by comparing them against data from an experimental study.

Step 2: Latin Hypercube Sampling (LHS) technique was employed to generate 200 realizations of the 8 random variables [33], creating an input matrix for the subsequent analysis.

Step 3: For a given realization (*i*), the MVMS was computed using ABAQUS FEA software. The analysis began at an internal pressure (IP) of 0 MPa and a longitudinal defect spacing (S_l) of 0 mm.

Step 4: For a fixed defect spacing, the internal pressure was incrementally increased. If the pipeline did not fail, IP was increased by 1 MPa increments up to 20 MPa. Beyond 20 MPa, the increment was refined to 0.1 MPa to accurately pinpoint the failure pressure. This approach efficiently tracks failure while conserving computational resources.

Step 5: After completing the pressure cycle for a given spacing, the effect of defect interaction was assessed by increasing the longitudinal spacing by $0.2\sqrt{Dt}$ This process was repeated for each realization until the spacing reached the maximum limit of $2\sqrt{Dt}$.

Step 6: Steps 3 through 5 were repeated for all 200 realizations.

Step 7: The FEA results were compiled into a comprehensive dataset. The input features include pipeline geometry (*OR, t*), defect characteristics (*d, l, S_l*), material properties (*E, EYS, EUS*), and loading conditions (*IP, EP*). The target output variable is the MVMS.

Step 8: Several well-known regression machine learning models were selected and trained on the generated dataset to predict MVMS based on the input parameters.

Step 9: The performance of the trained models was compared and evaluated on a held-out test set. Techniques like k-fold cross-validation were used to assess accuracy and prevent overfitting.

Step 10: The best-performing machine learning model was identified and presented as the final predictive tool for estimating MVMS in pipelines with longitudinally aligned corrosion defects.

3. Random Sampling

Abyani and Bahaari [34] demonstrated that the LHS technique effectively evaluates the reliability of corroded pipelines while requiring fewer random samples than the MCS method. Additionally, TDA analysis was applied to investigate how different parameters influence the sensitivity of corroded pipelines' MVMS, aiding in parameter selection [10]. Table 1 outlines the probabilistic properties of the random variables obtained through the LHS method and specify 95% Confidence Intervals (CI) of each. A total of 200 random samples were generated for each parameter, with studies by Hosseinzadeh et al. [35] and

Abyani and Bahaari [10] confirming the adequacy of this sample size.

Table 1. Statistical attributes of the examined random variables

Parameter	Sign	Dist.	Unite	Mean	COV	95% CI	Reference
Defect Depth	d	N	mm	10	0.1	[8.04, 11.96]	[36]
Defect Length	l	N	mm	200	0.05	[180.4, 219.6]	[37]
Pipeline Outer Radius	OR	N*	mm	406.4	0.03	[382.50, 430.30]	[37]
Pipeline Thickness	t	N	mm	20	0.05	[18.04, 21.96]	[36]
External Pressure	EP	G**	MPa	0.6	0.03	[0.5647, 0.6353]	[38]
Young's modulus	E	N	MPa	210000	0.05	[189420, 230580]	[39]
Engineering Yield Stress	EYS	L***	MPa	464.5	0.056	[414.03, 520.77]	[36]
Engineering Ultimate Tensile Strength	EUS	N	MPa	563.8	0.03	[530.67, 596.93]	[40]

** Normal*
*** Log-Normal*
****Gamble (max)*

4. Numerical modeling

For stress analysis on offshore pipelines corroded by interacting defects, a 3D finite element model has been generated utilizing ABAQUS 6.14. Within this model, a dataset for burst pressure is generated, considering variations in pipeline geometry, materials, and defect characteristics. Corroded pipe models are created as a quarter pipe model according to symmetry rules utilizing solid elements and featuring two identical defects on the inner surface. The details of the FEM model's properties and its validation is extensively covered in earlier research by the authors.[35]. A summary of the FEM model development procedure for an offshore pipeline suffering from interacting corrosion defects is outlined below:

4.1. Defect shapes and configurations

The simulated corrosion defects represent localized metal loss, which encompasses both pitting corrosion and general metal loss. To define rational models, interaction rules available in the literature were studied to estimate suitable longitudinal spacing levels between defects. Various interaction rules have been established to demonstrate interacting defects, with specific limits defined for this purpose. These rules are categorized into longitudinal and circumferential types, based on the defects' size and position. Both industry standards and academic studies contribute interaction rules that determine the distance at which defects influence one another. Idris et al. [41] have gathered significant interaction rules from existing literature which provide the engineering basis for determining the distance at which defects influence one another. This review categorizes these rules into longitudinal and circumferential types, based on the defects' size and position. Consequently, this study evaluates the interaction of defects arranged longitudinally, as depicted in Figure 2.

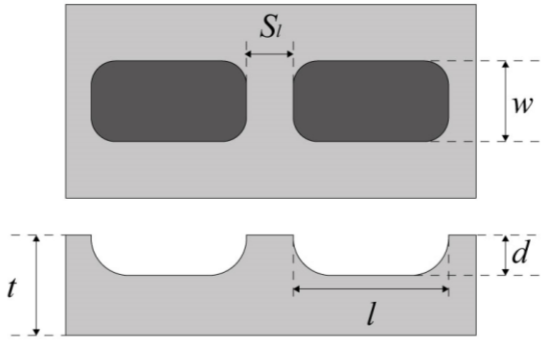


Figure 2. Defects' geometry and arrangement scheme

As a result, the current research aims to evaluate the influence of interaction of defect on pipeline performance. Rectangular-shaped defects rounded at edges will be aligned along the pipeline length. As a quarter model has been modeled due to symmetry rules, a single defect was created (see Figure 3). The gap between defects will gradually increase from 0 to $2\sqrt{Dt}$, with intervals of $0.2\sqrt{Dt}$, resulting in 11 distinct distances for every samples.

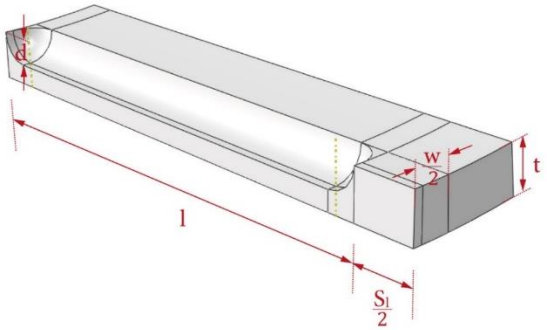


Figure 3. defect geometry design in Abaqus

4.2. Material properties

The present paper analyzes a pipeline constructed from API-5L-X65 steel, a common material for hydrocarbon transport. The material's behavior is modeled using the stress-strain relationship defined by the Ramberg-Osgood equation (See Eq. (1)) [42].

$$\varepsilon_E = \frac{\sigma_E}{E} + S_0 \left(\frac{\sigma_E}{\sigma_{EY}}\right)^N \quad (1)$$

The true stress is then determined by [43]:

$$\varepsilon_T = Ln(1 + \varepsilon_E) \quad (2)$$
$$\sigma_T = \sigma_E(1 + \varepsilon_E) \quad (3)$$

Here, σ_{EY}, σ_T, and ε_T represents the engineering yield stress, true stress, and true strain, while S_0 and N are material-specific constants. For API-5L-X65 steel, S_0 is set to 0.003 and N is set to 22.12 [44]. A full true stress–strain curve derived from the mentioned equation is shown in Figure 4:

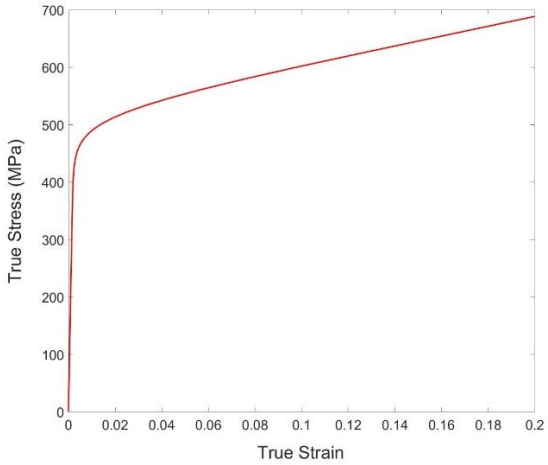


Figure 4. API 5L X65 true stress–strain curve

4.3. Boundary conditions and loads

The pipeline is subjected to internal and external pressures on its respective surfaces. IP begins at 0 MPa and steadily rises until it meets the failure criteria. Additionally, water depth-induced hydrostatic pressure is applied to the pipeline. In the study area, a water depth of 60 m is assumed (maximum depth in the midline area of the Persian Gulf), with seawater density considered to be 1020 $\frac{kg}{m^3}$ [10].

The pipeline model, representing a quarter of the complete corroded pipeline, is developed based on symmetry rules. The longitudinal faces are restricted to movement exclusively in the Z and Y directions and allowed to rotate solely around the X axis. Additionally, the freedom of the pipe section is restricted to translate in the X and Y direction and rotate around the Z axis. A point positioned in Z=0 and X=Y=OR is treated as rigid. Further details of these BC are depicted in Figure 5.

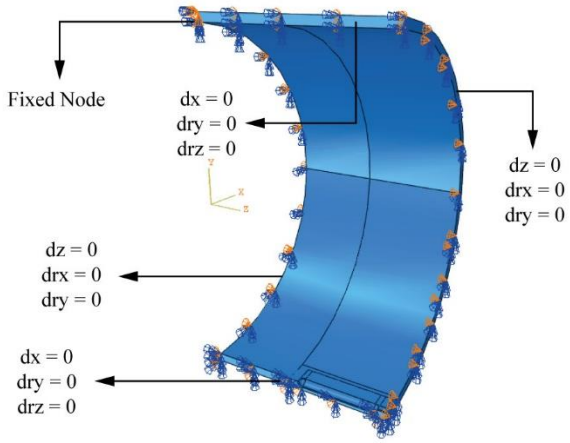


Figure 5. Pipeline model load condition and boundary condition

4.4. Meshing Structure

Hexahedral elements, such as Abaqus C3D8, were used in this study and also, a mesh sensitivity analysis, combined with model validation have been carried out. The model was partitioned into three areas: the defect area, defect-free area, and transition area (see Figure 6). In the defect-free area, a coarser mesh was utilized compared to the defect zone to minimize computational demands. A transition area was defined too, to ensure a smooth integration of the mesh structure utilizing a single biased meshing rule (illustrated in Figure 6).

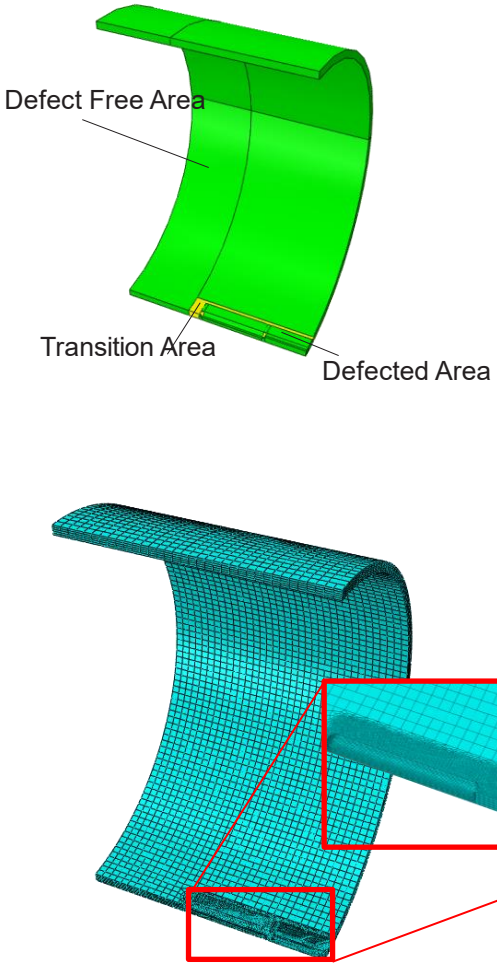


Figure 6. Meshing model system

4.5. Failure criteria

According to earlier studies [35][10][45][46], this research determines that pipeline failure happens when the Von Mises stress across the thickness of the ligament exceeds the true ultimate strength of the element. This method employs a recognized standard to evaluate the critical condition and identify failure in the model.

4.6. Model Verification

The model results were validated using an experimental data from Benjamin et al. [47], which involved an IDTS3 specimen – a pipeline with two longitudinally aligned corrosion defects matching the current model. The verification process is detailed in the authors' recent paper [35]. Model performance was evaluated by analyzing the failure pressure as the mesh size decreased, with results compared to the experimental data. As shown in Table 2, a minimum error of 1.39% was observed for mesh sizes smaller than one-fifth of the pipeline thickness.

Table 2. The verification model results for different mesh layer counts in pipeline thickness

Number of mesh layers in thickness	1	2	3	4	5	6	7	8
Failure Pressure (Mpa)	22.9	22.7	20.8	20.7	20.6	20.6	20.6	20.6
Error (%)	11.29	10.51	2.34	1.86	1.39	1.39	1.39	1.39

To optimize the model a sensitivity analysis was performed to assess the minimum acceptable length for the Pipe model. The FEM model consists of 3 parts, the length of the defect zone is equal to the defect length and the length of part 2 which is a transition part to help the meshing algorithm has been selected 10 percent of the defect length. Therefore, to obtain the optimum length of the model, the length of the 1st part has been reduced from 20D to 0.5D and the burst pressure of the model has been assessed in each case as shown in Table 3. Also, to assess the effect of defects longitudinal spacing the effect of model length was observed in cases where the longitudinal space between defects is 0 and 250 mm (approximately equal to $2\sqrt{Dt}$ for a 32" pipeline).

Table 3. The failure pressure in various model length

Sl (mm)	*L/D* 20	10	5	2.5	1	0.5
0	20	20	20	20	18.5	14.5
20.5	20.6	20.6	20.6	20.6	20	17.6
250	21.25	21.25	21.25	21.25	21.25	21.18

According to the sensitivity analysis results it can be concluded that by choosing a length equal to 2.5 times of pipe diameter for part 3 which is about 100 mm the results are converged.

5. ML Algorithms

In this section, various Artificial Intelligence (AI) models have been examined to find the best ML method for predicting the MVMS, based on the main parameters of pipeline geometry (outer radius and thickness), defect geometry (depth, and length), loading condition (internal and external pressure), and material specification (Ultimate Tensile Strength, Yield Stress, and young modulus).

AI is a branch of computer science aimed at designing systems that handle tasks requiring human-like cognitive abilities, such as solving problems and making decisions. [48]. ML falls under the broader category of AI. that deals with algorithms enabling computers to learn from data and make predictions [49]. Regression is a specific type of ML technique used to model relationships between variables when the goal is to predict continuous numerical outcomes [50]. In continuation, various regression ML models have been assessed to find the best AI model to predict the MVMS in corroded pipelines with longitudinal interacting defects. Scikit-Learn (sklearn) is utilized to create regression ML models [51]. This Python library is renowned for its versatility in ML, offering a uniform interface for a broad spectrum of tasks. The integrity of other Python libraries such as NumPy and pandas, makes Scikit-Learn an invaluable resource for streamlining the development and deployment of ML models across diverse applications.

In the present study, 5 Extensively employed ML models such as Linear Regression (LR), Stochastic Gradient Descent (SGD), K-nearest Neighbors (KNN), Decision Tree Regression (DTR), and Neural Network (NN) have been utilized. 80% of input data has been used for training the models. Also, 20% of the data will be reserved for the test dataset, while the remaining 80% will be allocated for training the ML model. The test dataset plays a vital role in assessing the model's performance. The input features have been transformed to have a mean value of 0 and a standard deviation of 1. This method is a typical preprocessing procedure to enhance the data compatibility with ML algorithms, particularly those that are influenced by the feature scales, like SVM and k-nearest neighbors. Standardization has been done to avoid biases related to parameter values that can occur when features are on different scales.

5.1. Linear Regression (LR)

Linear regression is a fundamental ML algorithm, that establishes a linear model that establishes a connection between a dependent variable (the target) and several independent variables (predictors) by fitting a linear equation [52]. The model seeks to minimize the discrepancy between its predictions and the actual

values in a training dataset, utilizing the least squares method. Once trained, the model can provide forecasts for new data. It is a widely applied and easily interpretable approach for comprehending and forecasting relationships between variables in situations where those relationships are linear, although it may not excel when dealing with non-linear associations. Evaluation metrics like Mean Squared Error and R-squared are used to gauge its effectiveness. This relationship is represented by a linear equation, typically in the form Eq.(4):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + + \beta_n x_n + \varepsilon \quad (4)$$

Where: y is the dependent variable. $x_1, x_2, \ldots, x_n$ are the independent variables. β_0 is the intercept. β_1, $\beta_2, \ldots, \beta_n$ are the coefficients for each independent variable, representing their respective slopes and ε is the error term.

5.2. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) regression is a widely employed optimization algorithm within the realm of machine learning and statistical modelling. .This model is well-suited for addressing large-scale optimization challenges associated with substantial training sets, SGD is well known for its simplicity and prevalence in stochastic optimization methods [53]. The algorithm iteratively refines an objective function with precision, rendering it a potent tool for minimizing objective functions across various applications [54]. Moreover, SGD exhibits an implicit regularization mechanism, ensuring that solutions derived through its application generalize effectively, regardless of the number of parameters involved [55].

5.3. K-Nearest Neighbors (KNN)

K-nearest neighbor (KNN) regression represents a supervised learning algorithm designed for regression tasks. This straightforward method determines the output of a new instance by considering the average or weighted average of its k-nearest neighbors within the feature space. The KNN algorithm is recognized for its simplicity and performance that rivals more intricate regression techniques[56]. Notably, KNN regression can be extended to support interval regression. In this approach, a novel method leveraging tolerance intervals determines the hyper-parameter K for each instance, balancing precision and uncertainty arising from limited sample size [57]. Additionally, KNN regression finds application in multidimensional query processing, exemplified by the k-nearest neighbor (kNN) query, which identifies the k-nearest points using distance metrics from a given location [58]. In a simple case with one feature, the predicted value for a test data point would be the average of the target values of its 'k' nearest neighbors [59] (Eq. (5)).

$$y = \frac{1}{k}\sum_{i=1}^{k} y_i \quad (5)$$

Where y is the predicted target value for the test data point. k is the number of nearest neighbors. And y_i is the target value of the i-th nearest neighbor.

5.4. Decision Tree Regression (DTR)

Decision tree regression stands out as a versatile tool employed for regression tasks, where the primary objective is to predict a continuous value. This algorithm, though simple, proves effective and applicable to both classification and regression problems [60]. Its operational principle involves recursively partitioning data into subsets based on input feature values and subsequently fitting a simple model within each subset. The outcome is a tree-like model where the leaves signify predictions. Decision tree regression finds extensive application across various domains, encompassing sales prediction, stock price forecasting, and retail promotion planning [61]. decision tree regression models exhibit versatility in handling both numerical and categorical data, showcasing robustness to outliers owing to their non-parametric nature [62]. Notably, the resultant model is interpretable, offering insights into the significance of different features in the prediction process.

5.5. Neural Network (NN)

Neural network regression emerges as a potent tool for modeling and forecasting continuous values. Successfully implemented in diverse applications [63]. What sets neural networks apart is their capacity to capture intricate relationships within data, rendering them well-suited for nonlinear regression tasks [64]. Additionally, they demonstrate superior prediction accuracy compared to traditional regression models [65].

The architecture of a neural network is composed of elementary processing units called neurons, which perform mathematical functions [66]. Neurons are structured into different layers, which generally include an input layer, multiple hidden layers, and an output layer. [67]. The input layer accepts the initial data, which is then handled by the hidden layers before reaching the output layer that delivers the final output. The weights of the connections between neurons are adjusted during training, allowing the network to refine its learning based on the data. As the number of hidden layers and nodes in a neural network grows, both its complexity and processing time increase markedly. [68]. Also, Tirumala & Narayanan [69] found that when hidden layers have the same number of nodes, performance is better than when hidden layers have different numbers [69]. In the present study the input layer consists of pipeline and defects characteristics

and outer layer is the MVMS value analyzed by FEM. For instance, a schematic of an NN model including 2 hidden layers has been depicted in Figure 7.

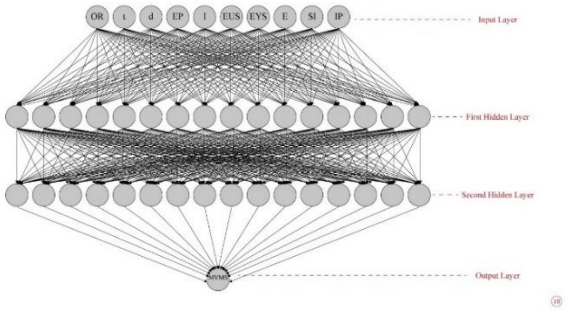


Figure 7. NN model schematic

6. Model Performance Assessment Methods

Various metrics are typically employed to assess the effectiveness of regression models, offering insights into their accuracy, precision, and ability to generalize. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination R-squared (R2) are extensively employed metrics for gauging model performance [70]. MAE represents the average magnitude of errors between predicted and observed values, calculated as the absolute mean of the error (Eq. (6)) [71].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{P}_{fi} - P_{fi}| \qquad (6)$$

Conversely, RMSE quantifies the square root of the average squared differences between predicted and observed values, assigning more significance to larger errors (Eq. (7)) [72].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_{fi} - P_{fi})^2} \qquad (7)$$

Moreover, R2 or the coefficient of determination indicates the proportion of predictable variance in the dependent variable from the independent variable(s) (Eq. (8)) [73].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{P}_{fi} - P_{fi})^2}{\sum_{i=1}^{N}(\hat{P}_{fi} - \bar{P}_{fi})^2} \qquad (8)$$

where $\hat{P}_{fi}$ and P_{fi} are the predicted and actual failure pressures, respectively, and $\bar{P}_{fi}$ is the average value of the actual failure pressures and N is the number of variables.

7. Results and discussion

7.1 LHS-FEM Dataset

As mentioned before the taking advantages of ABAQUS software a FEM model were developed to analyze the MVMS value in 121 IP levels (from 1 MPa to 20 MPa in increments of 1 MPa and from 20.1 MPa to 30 MPa in increments of 0.1 MPa) and 11 spacing levels (ranging from 0 to $2.0\sqrt{Dt}$ in increments of $0.2\sqrt{Dt}$) for each of 200 random variables. A sample of ABAQUS results has been illustrated in Figure 8 illustrating MVMS contours. The results show the highest MVMS values has been occurred in the defect and defects spacing area.

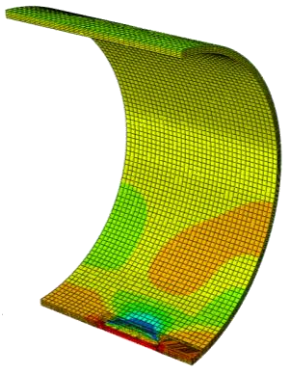


Figure 8. Example of ABAQUS FEM analysis results

Having these Valuable data, a data set matrix has been generated containing 10 input variables including Outer Radius (OR), Thickness (t), defect depth (d), defect length (l), Engineering Ultimate Strength (EUS), Engineering Yield Strength (EYS), Young Module (E), longitudinal Spacing (S_l), and Internal Pressure (IP) and Maximum Von Mises Stress (MVMS) as output data which leads to 11 columns and 266,200 rows.

Figure 9 illustrates the count histograms corresponding to each input data. The initial eight data, comprising $OR, t, d, EP, l, EUS, EYS,$ and E, are generated through LHS, adhering to a defined distribution. The histogram for IP similar to S_l demonstrates a uniform distribution when IP<20. However, beyond IP>20, the counts increase as the IP interval decreases from 1 to 0.1, but a declining trend emerges subsequently due to observed failures.

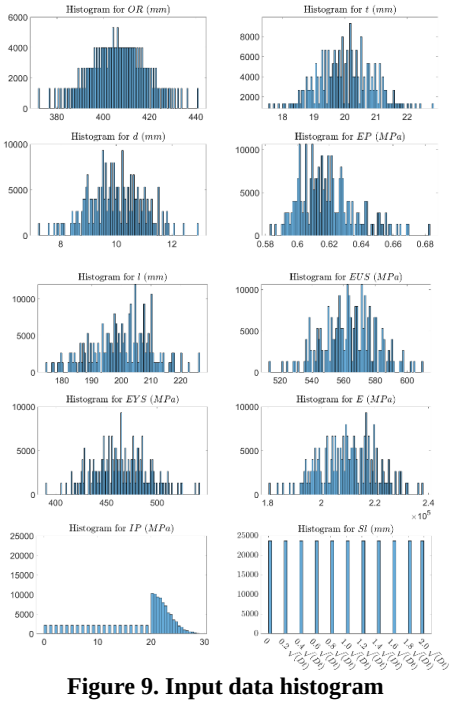


Figure 9. Input data histogram

7.2 AI Model Development

The machine learning dataset contains 266,200 samples, with each sample containing 10 input and one output variable. Out of these, 212,960 samples are used for training the prediction models, and 53,240 samples are used for evaluating the models' performance, maintaining an 8:2 ratio. This division supports detailed analysis of how the input variables relate to the output, leading to the creation of reliable ML models for accurate predictions.

Considering the significance of hyperparameters in attaining high-performing models, in this paper, k-fold cross-validation techniques were applied to evaluate the performance of the proposed model, ensuring its resilience and dependability across numerous train-test divisions. K-fold cross-validation stands as a widely embraced approach for assessing the predictive efficacy of regression models [74]. This method entails segmenting the dataset into k subsets, utilizing k-1 subsets for training, and reserving one subset for validation, iterating this process k times to ensure robustness and guard against overfitting [75]. The resulting k-fold cross-validation error estimation serves as a crucial metric for evaluating the model's generalization capabilities [76], proving especially beneficial in data-scarce scenarios by leveraging all available data for training and validation purposes [77]. Moreover, it serves as a tool for selecting optimal hyperparameters, such as λ in regularization techniques [78].

Also, a grid search was utilized to methodically investigate hyperparameter combinations aimed at optimizing model performance. Grid search is a well-known hyperparameter optimization technique within regression models. Grid search methodology follows an exhaustive exploration of a predetermined subset of hyperparameter values to pinpoint the combination that elicits the finest model performance [79]. Grid search is esteemed for its simplicity and broad applicability in parameter optimization [80], proving particularly advantageous in fine-tuning hyperparameters across a spectrum of ML models.

In the LR model, grid search explores different settings of the 'fit_intercept' hyperparameter. The "fit_intercept" parameter in linear regression denotes the constant factor within the regression formula. It permits the regression line to intersect the y-axis at a designated position rather than originating from the origin. This parameter is crucial for capturing the baseline value of the dependent variable when all independent variables are set to zero. It constitutes a fundamental aspect of the linear regression model, enhancing its capability to capture the relationship between independent and dependent variables with greater flexibility [81] When "fit_intercept" is set to True, the linear regression model will estimate an intercept, or bias term, along with the coefficients for the input features. This intercept represents the value of the dependent variable (target) when all input features are zero.

The SGD model has been optimized assessing the loss function, maximum iteration (max_iter), alpha and epsilon parameters. The loss function represents the function to be minimized during the training of the SGD model [82]. On the other hand, alpha parameter controls the regularization strength in the SGD model. It is the coefficient that multiplies the regularization term [55]. Furthermore, the epsilon parameter is related to the margin of tolerance for the model and it controls the width of the epsilon-insensitive tube [83]. Finally, the last analyzed hyperparameter (max_iter) sets the maximum number of iterations for the solver to converge. It specifies the maximum number of iterations taken for the solver to converge or reach a stopping criterion [84].

To enhance KNN model performance, the grid search explores optimal values for Number of neighbors to consider (n_neighbors), The weight function used in prediction (weight), and the power parameter (p) for the Minkowski distance metric. The weight function used in prediction can be 'uniform' (all neighbors have equal weight) or 'distance' (weight points by the inverse of their distance) (Altay, Ulaş, and Alyamaç 2020). For ($p = 1$), the distance measured is the Manhattan distance; for ($p = 2$), it is the Euclidean distance [85].

Exploring optimal values for DTR model hyperparameters, different configurations for the maximum depth of the tree (max_depth), the least number of samples needed to divide an internal node (min_samples_split), and the smallest number of samples required at a leaf node (min_samples_leaf) were analyzed. Increasing max_depth allows the tree to delve deeper, capturing more intricate relationships within the training data. However, this also heightens the risk of overfitting, particularly if the tree becomes excessively deep and captures noise in the data [86]. On the other hand, higher values for min_samples_split yield a tree with fewer splits, preventing the model from becoming overly tailored to the training data, thereby aiding in overfitting control. However, setting this parameter too high may lead to underfitting, as the tree may fail to capture essential patterns in the data [87]. min_samples_leaf, similar to min_samples_split, regulates the minimum number of samples at the terminal (leaf) nodes. Larger values for min_samples_leaf result in more conservative models, impeding the creation of small leaf nodes. While this helps mitigate overfitting, excessively high values may contribute to underfitting [88].

The grid search for the NN model probes the best values for 3 hyperparameters. First, the optimum number of neurons in each hidden layer of the neural network is assessed, second the activation function used in each neuron is specified and third the alpha parameter is evaluated which is the regularization

strength, controlling the penalty on the complexity of the neural network [89].
The Grid search configurations and results have been summarized in Table 3.

Table 4. Grid search configurations and results

Model	Hyperparameters	Range	Final Value
LR	Fit Intercept	[True, False]	True
SGD	loss	[squared_loss, huber, epsilon_insensitive]	epsilon_insensitive
	alpha	[0.0001, 0.001, 0.01]	0.0001
	epsilon	[0.1, 0.01, 0.001]	0.1
	max_iter	[1000, 2000, 3000]	1000
KNN	n_neighbors	[3, 5, 7]	3
	weights	['uniform', 'distance']	distance
	p	[1, 2]	2
DTR	max_depth	[None, 5, 10, 15]	None
	min_samples_split	[2, 5, 10]	5
	min_samples_leaf	[1, 2, 4]	1
NN	hidden_layer_sizes	[(100,), (50, 50), (25, 25, 25)]	(50, 50)
	activation	['relu', 'tanh']	tanh
	alpha	[0.0001, 0.001, 0.01]	0.01

To ensure the robustness and reproducibility of the machine learning models developed in this study, key specifications were rigorously implemented throughout the modeling process. The models were trained using an 80/20 train/test split, with a fixed random seed (random_state=42) applied to all stochastic processes to guarantee consistent results. Hyperparameter tuning was performed using 5-fold cross-validation on the training set only. The neural network model utilized the Adam solver for optimization. Crucially, feature scaling was fitted exclusively on the training data to prevent any data leakage, and the same scaling parameters were then applied to the test set. These measures ensured that the reported performance metrics provide an unbiased estimate of the models' generalization capability.

7.3 Models Evaluation

The validity of the ML models has been evaluated by comparing the model results vs the MVMS values calculated by FEM, the results are derived from an analysis of the test dataset, constituting 20% of the total data. The comparison among model actual and predicted MVMS value (MPa) are illustrated in the Figure 10 to Figure 14. As can be seen in Figure 10 and Figure 11 ,The LR and SGD models demonstrates loos fit in the lower MVMS values. Notably, the KNN algorithm outperforms these methods and has a better performance in MVMS values greater than 50 MPa. DTR and NN (Figure 13 and Figure 14) models exhibit good performance and almost mirror FEM values while the NN model produces even less noise data in comparison with the DTR model.

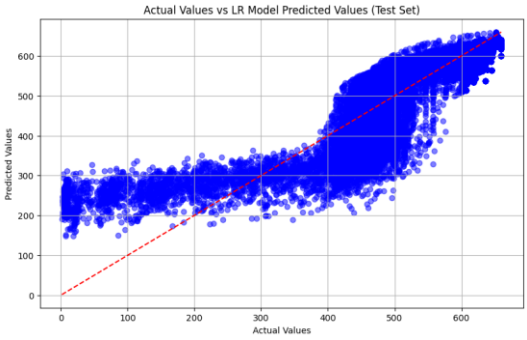


Figure 10. LR Model Prediction Accuracy

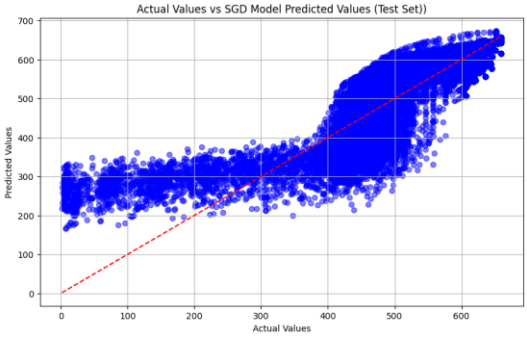


Figure 11. SGD Model Prediction Accuracy

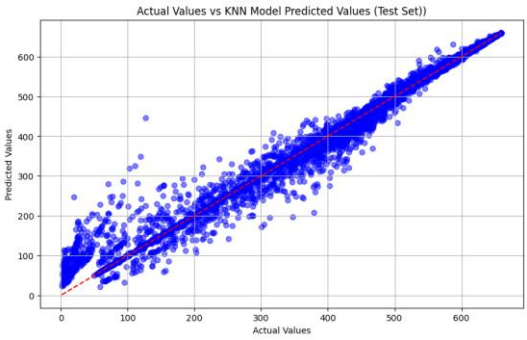


Figure 12. KNN Model Prediction Accuracy

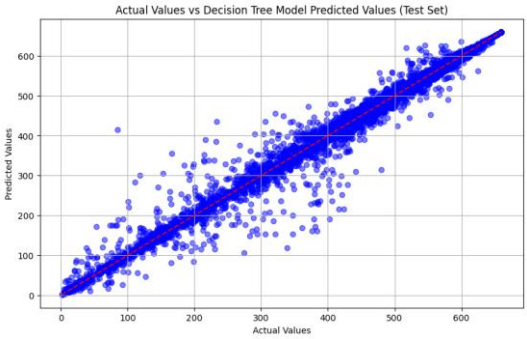


Figure 13. DTR Model Prediction Accuracy

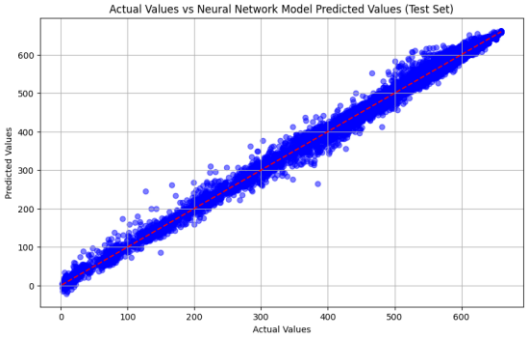


Figure 14. NN Model Prediction Accuracy

In order to have a better insight into model performances, Evaluation metrics such as $RMSE$, MAE, and R^2 are explored and compared in Figure 15, offering a comprehensive insight into the model's accuracy and predictive capabilities. The results show that KNN, DTR, and NN models present metrics lower than 10 for a data range of 0 to 600 which demonstrates a satisfying fit. The DTR and NN have presented the best performances according to metrics and NN has an RMSE of 15% lower than DTR. Also, according to Figure 14, the dispersion of data from the $X = Y$ line is a lower than DTR predictions shown in Figure 13. Therefore, the NN model has been selected to predict the MVMS data.

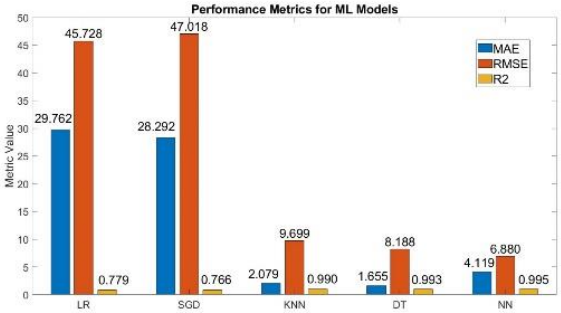


Figure 15. Model's performance metrics

7.4. Parameters Importance Order Assessment

Assessing random parameters' importance order is a way to assess the significance of uncertainties of each variable in evaluating the reliability. Various methods have been presented for evaluating predictor variable significance, such as recursive feature elimination [90], permutation importance (PI) ranking across diverse ML models [91], and utilizing ML approaches for feature selection and evaluation [92]. PI is a widely adopted technique for evaluating the relative significance of parameter features within ML models. PI provides a pragmatic and effective means to delineate the hierarchy of feature importance [93]. This method gauges the influence of each parameter feature on model performance by quantifying the rise in prediction error when the variable's values undergo random permutation while maintaining other variables constant [94].

Consequently, the Random Variables parameter importance has been calculated using permutation for the NN and DTR models to specify the role of each feature in finding the $MVMS$ values. The PI value of each feature in the NN and DTR model has been illustrated in Figure 16 and Figure 17 respectively. As the results demonstrate the t value can influence the $MVMS$ value more than other random features. In The second stage both models take a material parameter in their importance order nevertheless, NN model consider and DTR consider EYS and EUS respectively as their more effective parameter. Defect depth in both models stand in third place of importance order. A significant different in the models is about defects length which is the 4th effective parameter in NN model while it has the least effect on the DTR results.

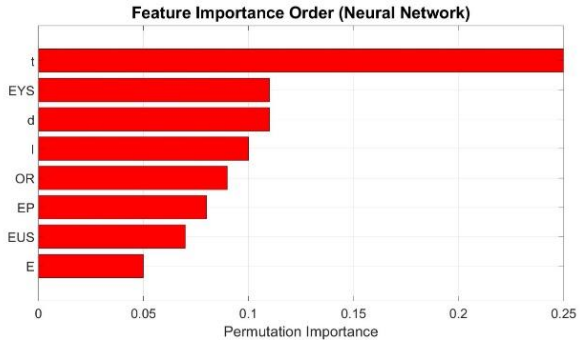


Figure 16. Features' importance order for NN model

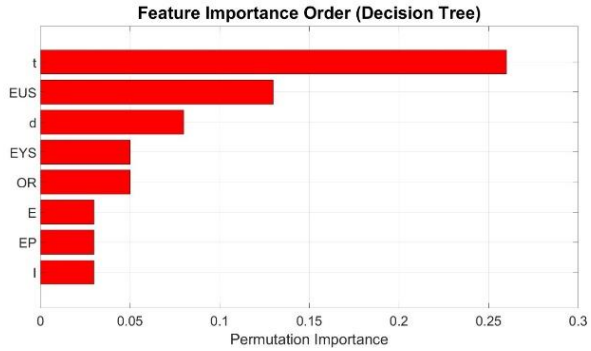
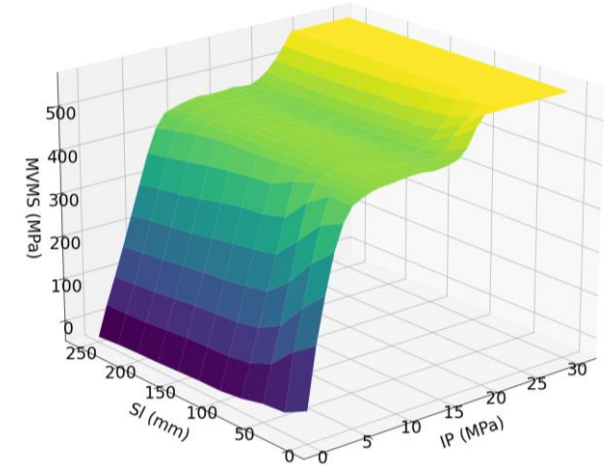


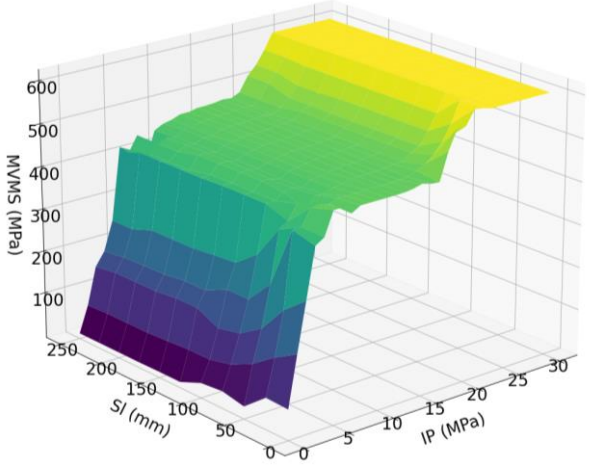
Figure 17. Features' importance order for DTR model

7.5. Model Results

To assess the performance of ML models in predicting the MVMS based on pipeline and defect characteristics, the predicted MVMS values by NN and DTR models were compared. In Figure 18, the MVMS was predicted at different longitudinal spacing and internal pressure levels using both models. The results are presented in a 3D scatter plot, where the X and Y axes represent longitudinal spacing and internal pressure, respectively, and the Z axis shows the predicted MVMS for specified internal pressure and spacing levels when other variables are at their mean values (as listed in Table 1).

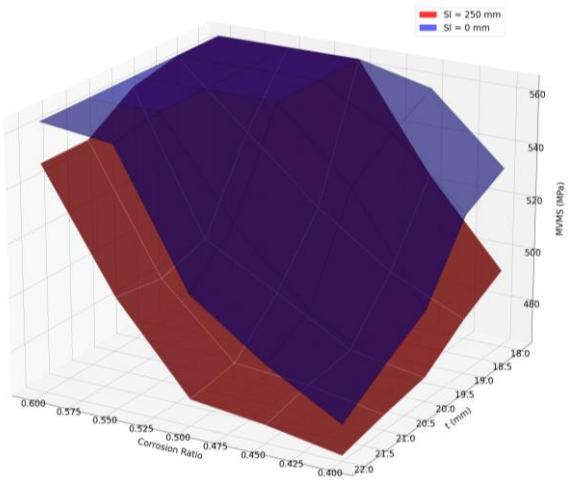


(A)

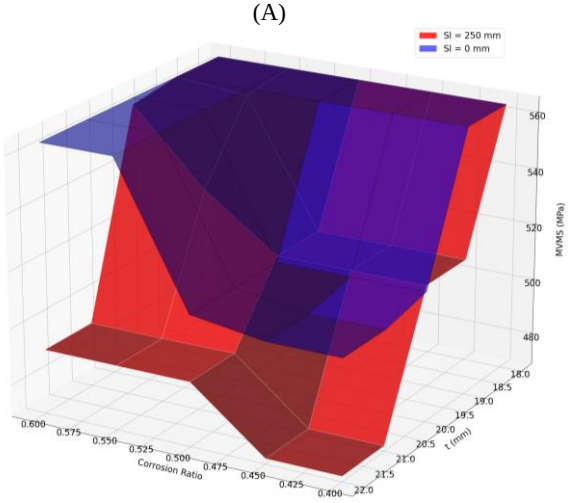


(B)

Figure 18. Predicted MVMS by NN (A) and DTR (B) models in different longitudinal spacings and internal pressures



(A)



(B)

Figure 19. Pipeline thickness and corrosion rate effect on Predicted MVMS using NN (A) and DT (B) models

As observed, the DTR model, due to its classification-based approach, produces jagged results and shows a decrease in precision at certain points despite its general accuracy. In contrast, the NN model provides a smoother surface with better accuracy. As expected, in both models, MVMS decreases with higher longitudinal spacing, reflecting reduced defect interaction, and increases with greater internal pressure.

Furthermore, the effect of pipeline thickness and corrosion rate on the MVMS at two longitudinal spacing levels ($S_l = 0$ *and* $S_l = 250\ mm$) with other input parameters at their mean values is illustrated in Figure 19. As seen, the DTR results are rough due to the inherent behavior of DTR models. Both models show a reduction in MVMS values as the corrosion rate decreases and indicate that MVMS is inversely proportional to pipeline thickness. The results also reveal that the DTR model predictions are more conservative; for instance, when $t = 18\ mm$, the DTR model predicts pipeline failure at a 40% corrosion rate, whereas the NN model predicts failure at corrosion rates above 50%. Therefore, it can be concluded that the NN model performs better in predicting MVMS in corroded pipelines.

The ML model can also be beneficial in assessing interaction rules. As previously mentioned, various interaction rules provided in the literature can determine the maximum space between defects where the interaction affects pipeline performance. The NN model was employed to evaluate the maximum defect spacing level where interaction impacts the MVMS in different defect geometries.

For this analysis, all input data were set to their mean values. Spacing and MVMS were assessed across various defect depths and lengths under different internal pressures (0:1:20 for $IP < 20$ and 20.1:0.1:30 for $IP > 20$). The pressure at which the MVMS exceeds the true ultimate strength value was defined as the burst pressure (see Figure 20).

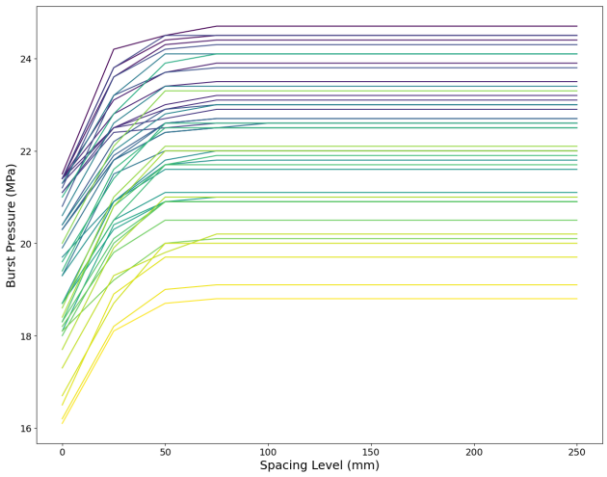


Figure 20. Burst Pressure in different Spacing levels

The spacing level at which interaction does not influence the results was defined as the maximum interacting spacing level, as illustrated in Figure 21. The results indicate an approximate increase in effective spacing with an increase in defect depth, consistent with interaction rules in the literature. However, changes in defect length do not exhibit a significant relationship with variations in the maximum interacting spacing level.

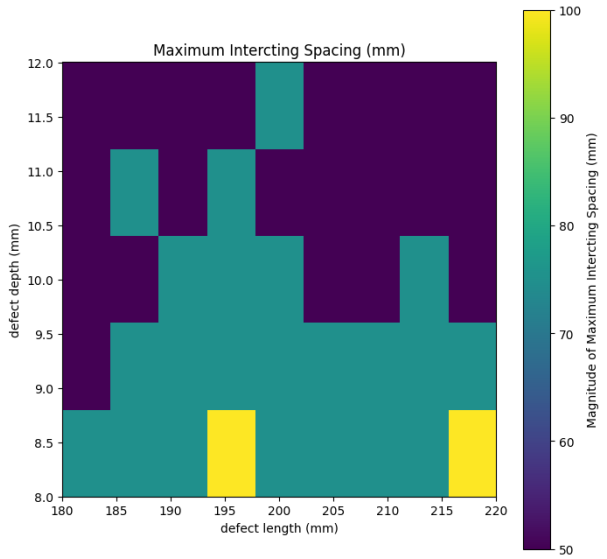


Figure 21. Maximum Spacing Between defects which the interaction is effective on MVMS assessed by NN model

8. Conclusions

In this paper, a study was conducted to generate an LHS-FEM database aimed at training an ML model capable of predicting MVMS in the outermost mesh layer of a ligament within the defective area of the pipeline thickness, considering the interaction of two longitudinally aligned defects.

Using ABAQUS software and leveraging Abaqus Python scripting an FEA dataset was generated. The LHS method was used to generate 200 random variables to study the MVMS for each one at 121 internal pressure (IP) levels (ranging from 1 to 20 MPa and from 20.1 to 30 MPa) and 11 spacing levels (ranging from 0 to $2\sqrt{Dt}$ in steps of $0.2\sqrt{Dt}$).

Using the LHS-FEM generated big data, 5 well-known and rigorous ML models including LR, SGD, KNN, DTR, and NN were trained to predict MVMS value. The database includes input data such as pipeline geometry parameters like outer radius (OR) and pipeline thickness (t), pipeline material data including engineering yield stress (EYS), engineering ultimate strength (EUS), and Young's modulus (E), as well as corrosion defect data encompassing defect depth (d), defect length (l), and levels of longitudinal defect spacing (S_l). To achieve the best hyperparameters in each model a gridsearch were executed. Also, the performance of the models was evaluated using k-fold cross-validation techniques, ensuring its resilience and dependability across various train-test divisions. The predictive models presented in this paper are validated for use within the 95% confidence intervals of the input parameters.

To evaluate the performance and accuracy of the models, the first 20% of the dataset, reserved as test data, was used for comparison with the actual FEM results to define the best model with the most accurate outcomes. In the subsequent step, the model's ability to generalize its results was assessed using the learning curve method.

The results indicated that the DTR and NN models exhibited the best prediction accuracy among the assessed models. The prediction performances of the NN and DTR models were compared, revealing that the NN model produced smoother and more accurate results. The findings indicated that the MVMS increases with a rise in spacing level when other parameters remain constant. Additionally, the MVMS decreases with a reduction in the corrosion rate and an increase in pipeline thickness. The NN results were also used to assess the maximum defect spacing level where the interaction significantly affects MVMS. The results demonstrated that the depth of the defect impacts the defect interacting space.

9. References

1- Hosseinzadeh S, Gatmiri B. Bearing behavior assessment of wind turbines' s shallow foundations, comparison of gravity-based foundations and suction buckets. Ocean Syst Eng [Internet]. 2025 [cited 2025 Oct 5];15(3):241. Available from: http://techno-press.org/content/?page=article&journal=ose&volume=15&num=3&ordernum=1

2- Sampath S, Bhattacharya B, Aryan P, Sohn H. A Real-Time, Non-Contact Method for in-Line Inspection of Oil and Gas Pipelines Using Optical Sensor Array. Sensors. 2019;19(16):3615.

3- Demoz A, Papavinasam S, Omotoso O, Michaelian K, Revie RW. Effect of Field Operational Variables on Internal Pitting Corrosion of Oil and Gas Pipelines.

Corrosion. 2009;65(11):741–7.
4- Colindres SC, Méndez GT, Velázquez JC, Cabrera-Sierra R, Angeles-Herrera D. Effects of Depth in External and Internal Corrosion Defects on Failure Pressure Predictions of Oil and Gas Pipelines Using Finite Element Models. Adv Struct Eng. 2020;
5- Rachman A, Zhang T, Chandima Ratnayake RM, Ratnayake RMC. Applications of Machine Learning in Pipeline Integrity Management: A State-of-the-Art Review. Int J Press Vessel Pip. 2021 Oct 1;193:104471.
6- Zheng Y, Zhang Y, Lin J. System reliability analysis for independent and nonidentical components based on survival signature. Probabilistic Eng Mech. 2023 Jul 1;73:103466.
7- Wang Y, Wharton JA, Shenoi RA. Mechano-electrochemical modelling of corroded steel structures. Eng Struct. 2016 Dec 1;128:1–14.
8- Yang Y, Wang GH, Qu Z, Zhang H, He J, Chen H. Reliability Analysis of Gas Pipeline With Corrosion Defect Based on Finite Element Method. Int J Struct Integr. 2021;
9- Cheng YF. Pipeline Corrosion. Corros Eng Sci Technol Int J Corros Process Corros Control. 2015;50(3):161–2.
10- Abyani M, Bahaari MR. A new approach for finite element based reliability evaluation of offshore corroded pipelines. Int J Press Vessel Pip. 2021 Oct 1;193:104449.
11- Vamvatsikos D, Allin Cornell C. Incremental dynamic analysis. Earthq Eng Struct Dyn [Internet]. 2002 Mar 1 [cited 2023 Feb 10];31(3):491–514. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/eqe.141
12- Mustaffa Z, Gelder P v., Dawotola AW, Yu S, Kim DK. Reliability Assessment for Corroded Pipelines in Series Considering Length-Scale Effects. Int J Automot Mech Eng. 2018;
13- Kuppusamy CS, Karuppanan S, Patil S. Buckling Strength of Corroded Pipelines With Interacting Corrosion Defects: Numerical Analysis. Int J Struct Stab Dyn. 2016;16(09):1550063.
14- Xie M, Wang Y, Xiong W, Zhao J, Pei X. A Crack Propagation Method for Pipelines With Interacting Corrosion and Crack Defects. Sensors. 2022;
15- Arumugam T, Rosli MKAM, Karuppanan S, Ovinis M, Lo M. Burst Capacity Analysis of Pipeline With Multiple Longitudinally Aligned Interacting Corrosion Defects Subjected to Internal Pressure and Axial Compressive Stress. Sn Appl Sci. 2020;2(7).
16- Zhang H, Sun M, Zhang J, Zhang Y, Li B, Zhai K. Study on Assessment Method of Failure Pressure for Pipelines with Colony Corrosion Defects Based on Failure Location. Process 2023, Vol 11, Page 3134 [Internet]. 2023 Nov 2 [cited 2024 May 25];11(11):3134. Available from: https://www.mdpi.com/2227-9717/11/11/3134/htm
17- Fekete G, Varga L. Extension of Pit Corrosion Effect on Pipelines. Period Polytech Mech Eng. 2011;55(1):15.
18- Abyani M, Bahaari MR. Effects of correlation between the adjacent components on time dependent failure probability of corroded pipelines. Struct Infrastruct Eng [Internet]. 2020;0(0):1–14. Available from: https://doi.org/10.1080/15732479.2020.1811990
19- Ossai CI, Boswell B, Davies IJ. Predictive Modelling of Internal Pitting Corrosion of Aged Non-Piggable Pipelines. J Electrochem Soc. 2015;162(6):C251–9.
20- Nizamani Z, Mustaffa Z, Wen LL. Determination of Extension of Life of Corroded Offshore Pipelines Using Form and Monte Carlo Structural Reliability. 2015;
21- Hou X, Wang Y, Zhang P, Qin G. Non-Probabilistic time-varying reliability-based analysis of corroded pipelines considering the interaction of multiple uncertainty variables. Energies. 2019;12(10).
22- Cui J. Studying Corrosion Failure Prediction Models and Methods for Submarine Oil and Gas Transport Pipelines. Appl Sci. 2023;13(23):12713.
23- Capula Colindres S, Méndez GT, Velázquez JC, Cabrera-Sierra R, Angeles-Herrera D. Effects of depth in external and internal corrosion defects on failure pressure predictions of oil and gas pipelines using finite element models. Adv Struct Eng. 2020 Oct 1;23(14):3128–39.
24- Abyani M, Karimi M, Shahgholian-Ghahfarokhi D. Failure assessment of corroded offshore pipelines using code-based approaches and a combination of numerical analysis and artificial neural network. Int J Press Vessel Pip [Internet]. 2024;209(April):105194. Available from: https://doi.org/10.1016/j.ijpvp.2024.105194
25- 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 Jan 1;131:105810.
26- Ossai CI. A Data-Driven Machine Learning Approach for Corrosion Risk Assessment—A Comparative Study. Big Data Cogn Comput [Internet]. 2019 Jun 1 [cited 2025 Jul 20];3(2):1–22. Available from: https://doi.org/10.3390/bdcc3020028
27- Cai J, Jiang X, Yang Y, Lodewijks G, Wang M. Data-Driven Methods to Predict the Burst Strength of Corroded Line Pipelines Subjected to Internal Pressure. J Mar Sci Appl. 2022;
28- Zhang P, Venketeswaran A, Bukka SR, Sarcinelli E, Lalam N, Wright R, et al. Machine Learning Data Analytics Based on Distributed Fiber Sensors for Pipeline Feature Detection. 2023;
29- Abyani M, Bahaari MR, Zarrin M, Nasseri M. Predicting failure pressure of the corroded offshore pipelines using an efficient finite element based algorithm and machine learning techniques. Ocean Eng. 2022 Jun 15;254:111382.
30- Biswas S, Rajan H. Fair Preprocessing: Towards

Understanding Compositional Fairness of Data Transformers in Machine Learning Pipeline. 2021;
31- Aditiyawarman T, Setiawan Kaban AP, Soedarsono JW. A Recent Review of Risk-Based Inspection Development to Support Service Excellence in the Oil and Gas Industry: An Artificial Intelligence Perspective. Asce-Asme J Risk Uncert Engrg Sys Part B Mech Engrg. 2022;
32- Zhang C, Ye Z. Water Pipe Failure Prediction Using AutoML. Facilities. 2020;
33- McKay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. 1979 May;21(2):239.
34- Abyani M, Bahaari MR. A comparative reliability study of corroded pipelines based on Monte Carlo Simulation and Latin Hypercube Sampling methods. Int J Press Vessel Pip [Internet]. 2020;181(August 2019):104079. Available from: https://doi.org/10.1016/j.ijpvp.2020.104079
35- Hosseinzadeh S, Bahaari MR, Abyani M. Reliability Assessment for pipelines corroded by longitudinally aligned defects. Ocean Eng. 2024;
36- Zhang P, Su L, Qin G, Kong X, Peng Y. Failure probability of corroded pipeline considering the correlation of random variables. Eng Fail Anal [Internet]. 2019 May 1 [cited 2021 Jul 17];99(February):34–45. Available from: https://doi.org/10.1016/j.engfailanal.2019.02.002
37- Leira BJ, Næss A, Brandrud Næss OE. Reliability analysis of corroding pipelines by enhanced Monte Carlo simulation. Int J Press Vessel Pip. 2016;144:11–7.
38- Teixeira AP, Palencia OG, Soares CG. Reliability Analysis of Pipelines With Local Corrosion Defects Under External Pressure. J Offshore Mech Arct Eng. 2019;
39- Zarrin M, Asgarian B, Abyani M. Probabilistic seismic collapse analysis of jacket offshore platforms. J Offshore Mech Arct Eng [Internet]. 2018 Jun 1 [cited 2025 Aug 30];140(3):1–11. Available from: https://dx.doi.org/10.1115/1.4038581
40- Hosseinzadeh S, Bahaari MR, Abyani M. Corrosion Defects Interaction Impact on Failure Pressure of Offshore Pipelines. In: ICOPMAS 2022. 2022.
41- Idris NN, Mustaffa Z, Ben Seghier MEA, Trung NT. Burst capacity and development of interaction rules for pipelines considering radial interacting corrosion defects. Eng Fail Anal. 2021 Mar 1;121.
42- Bai Y, Yu Z. Pipeline On-Bottom Stability Analysis Based on FEM Model. Proc Int Conf Offshore Mech Arct Eng - OMAE [Internet]. 2011 Oct 31 [cited 2023 Feb 28];4:329–33. Available from: /OMAE/proceedings-abstract/OMAE2011/44366/329/357531
43- Song B, Sanborn B. Relationship of compressive stress-strain response of engineering materials obtained at constant engineering and true strain rates. Int J Impact Eng. 2018 Sep 1;119:40–4.
44- Hosseinzadeh S, Bahaari M, Abyani M, Taheri M. Data-Driven Remaining Useful Life Estimation of Subsea Pipelines Under Effect of Interacting Corrosion Defects. Appl Ocean Res. 2025;
45- Zhang YM, Tan TK, Xiao ZM, Zhang WG, Ariffin MZ. Failure assessment on offshore girth welded pipelines due to corrosion defects. Fatigue Fract Eng Mater Struct [Internet]. 2016 Apr 1 [cited 2023 Feb 27];39(4):453–66. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/ffe.12370
46- Mondal BC, Dhar AS. Burst pressure of corroded pipelines considering combined axial forces and bending moments. Eng Struct [Internet]. 2019 May 1 [cited 2023 Feb 24];186:43–51. Available from: http://dx.doi.org/10.1016/J.ENGSTRUCT.2019.02.010
47- Benjamin AC, Vieira RD, Diniz JLC, Freire JLF, De Andrade EQ. Burst Tests on Pipeline Containing Interacting Corrosion Defects. Proc Int Conf Offshore Mech Arct Eng - OMAE. 2008 Nov 11;3:403–17.
48- Kallem SR. Artificial Intelligence Algorithms. IOSR J Comput Eng. 2012;
49- Yves, Kodratoff., Ryszard, S. M. Machine learning: an artificial intelligence approach volume III. 1990.
50- Nelder JA. Regression Analysis by Example. Chatterjee S, Price B, editors. Biometrics [Internet]. 2023 Nov 17;35(1):355–6. Available from: http://www.jstor.org/stable/2529957
51- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res. 2011;12:2825–30.
52- Stephenson WR. Simple Linear Regression. 2003.
53- Shamir O. Stochastic Gradient Descent for Non-Smooth Optimization: Convergence Results and Optimal Averaging Schemes. 2012;
54- Ighalo JO, Adeniyi AG, Marques G. Application of Linear Regression Algorithm and Stochastic Gradient Descent in a Machine-learning Environment for Predicting Biomass Higher Heating Value. Biofuels Bioprod Biorefining. 2020;
55- Goldt S, Advani M, Saxe AM, Krzakala F, Zdeborová L. Dynamics of Stochastic Gradient Descent for Two-Layer Neural Networks in the Teacher–student Setup*. J Stat Mech Theory Exp. 2020;
56- Zaidi NA, Squire D, Suter D. BoostML: An Adaptive Metric Learning for Nearest Neighbor Classification. 2010;
57- Hamed MM, Serrurier M, Durand N. Simultaneous Interval Regression for K-Nearest Neighbor. 2012;
58- Nodarakis N, Pitoura E, Sioutas S, Tsakalidis AK, Tsoumakos D, Tzimas G. Efficient Multidimensional AkNN Query Processing in the Cloud. 2014;

59- Enas GG. Choice of the smoothing parameter and efficiency of k-nearest neighbor classification. Vol. 12, Computers & Mathematics With Applications. 1986. p. 235–44.
60- Lamrini B. Contribution to Decision Tree Induction With Python: A Review. 2021;
61- Reddy SRT, Malathi P. Design and Implementation of Sales Prediction Model Using Decision Tree Regressor Over Linear Regression Towards Increase in Accuracy of Prediction. 2022;
62- Guo T, Kutzkov K, Ahmed ME, Calbimonte JP, Aberer K. Efficient Distributed Decision Trees for Robust Regression. 2016;
63- Al-Mahasneh AJ, Anavatti SG, Garratt M, Pratama M. Applications of General Regression Neural Networks in Dynamic Systems. 2018;
64- Wong SF, Wong KYK. Wavelet Network for Nonlinear Regression Using Probabilistic Framework. 2004;
65- Gante D V, Silva DL, Leopoldo MP. Forecasting Construction Cost Using Artificial Neural Network for Road Projects in the Department of Public Works and Highways Region XI. 2022;
66- Ferreira R, Martiniano A, Ferreira A, Romero M, Sassi RJ. Container Crane Controller With the Use of a NeuroFuzzy Network. 2016;
67- Araghinejad S. Artificial Neural Networks. 2013;
68- Sarkar A, Mandal JK. Comparative Analysis of Tree Parity Machine and Double Hidden Layer Perceptron Based Session Key Exchange in Wireless Communication. 2015;
69- Tirumala SS, Narayanan A. Hierarchical Data Classification Using Deep Neural Networks. 2015;
70- Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast [Internet]. 2006;22(4):679–88. Available from: https://www.sciencedirect.com/science/article/pii/S0169207006000239
71- Willmott CJ, Matsuura K. Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Clim Res. 2005;
72- Willmott CJ, Robeson SM, Matsuura K. A Refined Index of Model Performance. Int J Climatol. 2011;
73- Zhang R, Guo Z, Meng Y, Wang S, Li S, Niu R, et al. Comparison of ARIMA and LSTM in Forecasting the Incidence of HFMD Combined and Uncombined With Exogenous Meteorological Variables in Ningbo, China. Int J Environ Res Public Health. 2021;
74- Hadjisolomou E, Stefanidis K, Herodotou H, Michaelides MP, Papatheodorou G, Papastergiadou E. Modelling Freshwater Eutrophication With Limited Limnological Data Using Artificial Neural Networks. Water. 2021;
75- Zhang Y, Wang T, Liu K, Xia Y, Lu Y, Jing Q, et al. Developing a Time Series Predictive Model for Dengue in Zhongshan, China Based on Weather and Guangzhou Dengue Surveillance Data. PLoS Negl Trop Dis. 2016;
76- Fushiki T. Estimation of Prediction Error by Using K-Fold Cross-Validation. Stat Comput. 2009;
77- Behroozi-Khazaei N, Nasirahmadi A. A Neural Network Based Model to Analyze Rice Parboiling Process With Small Dataset. J Food Sci Technol. 2017;
78- Ye H, Bellotti T. Modelling Recovery Rates for Non-Performing Loans. Risks. 2019;
79- Irmawati, Chai R, Basari, Gunawan D. Optimizing CNN Hyperparameters for Blastocyst Quality Assessment in Small Datasets. Ieee Access. 2022;
80- Qi Y, Liu H, Zhao J, Xia XH. Prediction Model and Demonstration of Regional Agricultural Carbon Emissions Based on PCA-GS-KNN: A Case Study of Zhejiang Province, China. Environ Res Commun. 2023;
81- Verbeek M. Using Linear Regression to Establish Empirical Relationships. Iza World Labor. 2017;
82- Quan J, Yan B, Sang X, Zhong C, Li H, Qin X, et al. Multi-Depth Computer-Generated Hologram Based on Stochastic Gradient Descent Algorithm With Weighted Complex Loss Function and Masked Diffraction. Micromachines. 2023;
83- Acharjee A, Finkers R, Visser RGF, Maliepaard C. Comparison of Regularized Regression Methods for ~Omics Data. J Postgenomics Drug Biomark Dev. 2012;
84- Chen X, Liu Q, Tong XT. Dimension Independent Excess Risk by Stochastic Gradient Descent. Electron J Stat. 2022;
85- Rani P. A Review of Various KNN Techniques. Int J Res Appl Sci Eng Technol. 2017;
86- Birajdar MR, Sewatkar CM. Machine Learning Approach to Predict the Thermal Performance of Closed-loop Thermosyphon. Heat Transf. 2023;
87- Cai C, Dong H, Wang X. Expectile c Forest: A New Nonparametric Expectile Regression Model. Expert Syst. 2022;
88- SUNESH, Balhara AK, Dahiya N, Himanshu, Singh RP, Ruhil AP. Machine Learning Algorithms for Predicting Peak Yield in Buffaloes Using Linear Traits. Indian J Anim Sci. 2022;
89- Meybodi MR, Beigy H. New Learning Automata Based Algorithms for Adaptation of Backpropagation Algorithm Parameters. Int J Neural Syst. 2002;
90- Araya SN, Fryjoff-Hung A, Anderson A, Viers JH, Ghezzehei TA. Advances in Soil Moisture Retrieval From Multispectral Remote Sensing Using Unoccupied Aircraft Systems and Machine Learning Techniques. Hydrol Earth Syst Sci. 2021;
91- Jiang D, Xu Y, Yang L, Gao J, Wang K. Forecasting Water Temperature in Cascade Reservoir Operation-Influenced River With Machine Learning Models. Water. 2022;
92- Pant J, Pant P, Bhatt AK, Pant H V, Pandey N. Feature Selection Towards Soil Classification in the Context of Fertility Classes Using Machine Learning. Int J Innov Technol Explor Eng. 2019;

93 -Huang F, Zhang Y, Zhang Y, Shangguan W, Nourani V, Li Q, et al. Towards Interpreting Machine Learning Models for Predicting Soil Moisture Droughts. Environ Res Lett. 2023;
94- Zhang Y, Dian Y, Zhou J, Peng S, Hu Y, Hu L, et al. Characterizing Spatial Patterns of Pine Wood Nematode Outbreaks in Subtropical Zone in China. Remote Sens. 2021;