

Research Article

Systematic Corrosion Prediction Techniques in Oil and Gas Pipelines Using Machine Learning Methods

Abubakar Usman^{1,*}, Alhaji Dodo Ibrahim Sulaiman¹, Usman Hassan¹,
Ahmed Mohammed Inuwa²

¹Department of Petroleum Engineering, Abubakar Tafawa Balewa University, Bauchi, Nigeria

²Department of Chemical Engineering, Abubakar Tafawa Balewa University, Bauchi, Nigeria

Abstract

Corrosion in oil and gas pipelines results in billions of dollars in annual losses and poses serious safety, environmental, and operational risks. Conventional detection methods are often labor-intensive, intermittent, and insufficiently accurate, hindering early identification of critical degradation. This study proposes an intelligent, SCADA-driven corrosion detection framework that integrates advanced machine learning (ML) and artificial intelligence (AI) techniques for real-time pipeline integrity monitoring. Operational parameters, including gas and liquid flow rates (up to 250 m³/h), pressure drops (0.3–0.8 MPa), and phase densities (600–900 kg/m³), were acquired from SCADA systems, pre-processed in MATLAB, and divided into training (30%), validation (30%), and testing (40%) datasets. Supervised ML models, Support Vector Machines (SVM), Random Forests, Boosted Trees, and Neural Networks, were optimized through feature selection and hyperparameter tuning for corrosion detection and rate prediction. The Linear SVM achieved the highest performance, with a ROC-AUC of 0.96 and a prediction tolerance of ± 0.02 mm/year. The integrated SCADA–AI framework achieved detection accuracies above 90%, enabling proactive maintenance scheduling and extending pipeline service life by 12–15%. These findings demonstrate a scalable, data-driven approach for predictive corrosion management and enhanced asset reliability in energy infrastructure.

Keywords

Artificial Intelligence, Predictive Maintenance, Machine Learning Optimization, Pipeline Corrosion Detection, Oil and Gas Pipeline Integrity

1. Introduction

Corrosion in oil and gas pipelines has long been recognized as a critical factor threatening the structural integrity and operational safety of pipeline networks. It causes material degradation, leading to potential leaks, failures, and significant financial losses [1]. Consequently, accurate and early detection of corrosion is essential for preventative maintenance,

ensuring operational reliability, and minimizing environmental impacts [2]. Traditional corrosion detection methods, including ultrasonic testing, radiography, and magnetic flux leakage, though effective, are often limited by high costs, labor-intensive procedures, and limited ability to provide continuous monitoring [3].

*Corresponding author: uabubakar06@gmail.com (Abubakar Usman)

Received: 27 November 2025; **Accepted:** 16 December 2025; **Published:** 31 December 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Traditional corrosion detection technologies, such as ultrasonic testing (UT), radiography, eddy current inspection, and magnetic flux leakage (MFL), provide valuable diagnostic capabilities [4]. However, these methods possess notable limitations; they often require pipeline shutdowns, are labor-intensive, and cannot enable real-time or continuous monitoring over long distances [5]. Additionally, they are sensitive to operational conditions and may struggle in harsh environments, buried pipelines, or inaccessible locations. These limitations underscore the need for more advanced, automated, and predictive corrosion detection strategies [6].

Recent developments in Artificial Intelligence (AI), Machine Learning (ML), and deep learning offer transformative potential for pipeline corrosion prediction and monitoring. These data-driven approaches can analyze large volumes of operational data, such as pressure, flow rate, temperature, electrochemical signals, and historical corrosion records, to identify early degradation patterns and predict future corrosion hotspots [7]. AI-based systems support predictive maintenance, improve decision-making, and reduce the risk of unexpected failures by enabling continuous, autonomous monitoring. However, despite significant progress, challenges still exist. Many AI models are computationally demanding, require large, labelled datasets for training, and face issues related to generalization across different pipeline materials, environments, and operating conditions [8]. The lack of standardized datasets, cybersecurity concerns, and limited integration with existing industrial supervisory control and data acquisition (SCADA) systems also hinders widespread industrial adoption [9].

This study aims to optimize corrosion detection in pipelines using AI models, benchmark their performance against conventional methods, and develop a framework for reliable, cost-effective, and real-time monitoring. Ultimately, this research contributes toward safer, cost-effective, and intelligent pipeline asset management, aligning with the global transition toward Industry 4.0 and smart energy systems.

Artificial intelligence of corrosion detection techniques in pipelines using machine learning (ML) methods has attracted significant research interest due to the critical need for accurate, efficient, and cost-effective monitoring of pipeline integrity [10]. Corrosion is a pervasive threat to oil and gas pipelines, causing safety hazards, environmental risks, and substantial economic losses [2]. Traditional corrosion detection techniques, such as ultrasonic testing, magnetic flux leakage, and visual inspections, while established, face limitations including high operational costs, limited spatial coverage, and challenges in automation and real-time monitoring [11]. Machine learning, by contrast, offers the potential to transform corrosion detection by enabling automated data interpretation, improved sensitivity to subtle signals, and predictive maintenance capabilities [12].

Recent literature demonstrates a wide array of ML applications in corrosion detection, from image-based methods to sensor signal analysis. For example, convolutional neural networks

(CNNs) have been successfully employed for detecting and classifying corrosion patterns in pipeline images with accuracies exceeding 98%, significantly outperforming traditional classification algorithms [13]. These deep learning approaches reduce the dependency on manual inspection and provide scalable solutions for continuous monitoring. Transformer models, a more recent innovation, have shown promise in adapting to diverse inspection data, enhancing detection robustness across different pipeline materials and environmental conditions.

In addition to image processing, ML models have been developed to analyze sensor data such as magnetic flux leakage and ultrasonic signals. Techniques, including support vector machines (SVM), k-nearest neighbors (KNN), XGBoost, and hybrid ensemble models, have been utilized for precise corrosion classification and severity estimation [14]. Integration of advanced signal denoising methods with ML feature extraction further improves detection sensitivity and reliability.

Predictive modelling using ML also extends to forecasting corrosion rates and estimating residual pipeline strength under different operational conditions. Such capability supports condition-based maintenance and risk-informed decision-making, enhancing safety and reducing unscheduled downtime [15]. However, challenges remain, including the scarcity of labelled corrosion datasets, variability in corrosion types, and the interpretability of complex ML models for industrial implementation. Current research focuses on overcoming these issues by leveraging data augmentation, transfer learning, and explainable AI techniques [16].

Overall, the systematic approach of corrosion detection through machine learning methods enables a paradigm shift from reactive to proactive pipeline integrity management. This approach not only increases detection accuracy and operational efficiency but also facilitates automated, real-time monitoring systems indispensable for modern oil and gas infrastructure.

This study presents a systematic machine learning framework for predicting corrosion in oil and gas pipelines, integrating data preprocessing, corrosion-specific feature engineering, and model evaluation within a unified workflow. Unlike previous studies, it leverages real-field operational data and compares multiple algorithms, including hybrid and ensemble models, to identify the most effective approaches. Explainable AI techniques are employed to reveal key factors driving corrosion, bridging the gap between data-driven predictions and engineering decision-making. The framework also supports dynamic prediction under varying operational conditions and provides practical recommendations for pipeline integrity management.

2. Materials and Methods

2.1. Materials/Tools

Table 1 presents the materials/tools that will be used in this research work.

Table 1. Materials/Tools and Their Purpose Used in this Research Work.

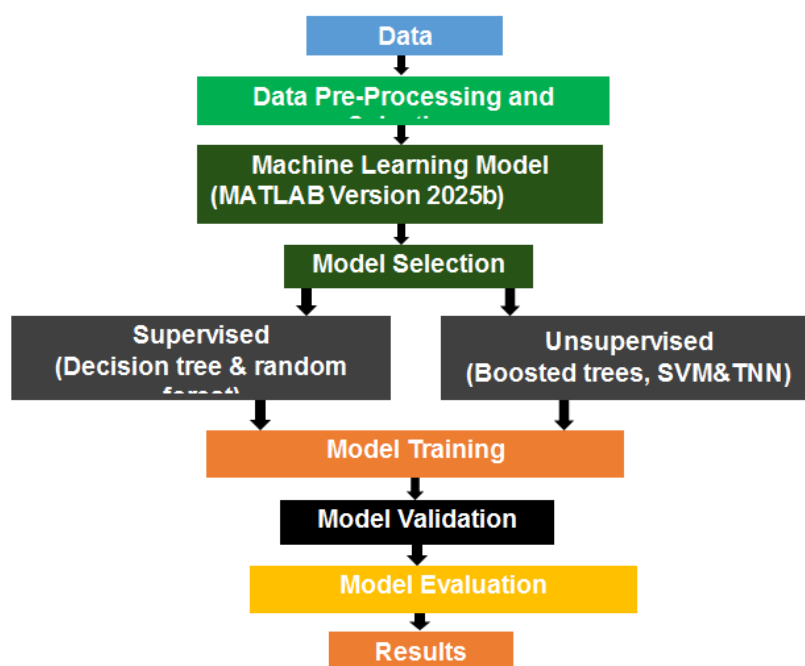
SNO	Equipment	Purpose
1.	MATLAB Software version 2025b	For running a machine learning algorithm
2.	Data to be collected from open literature.	Influence of corrosion in oil and gas exploration
4.	Computer set	MATLAB software installation
5.	Microsoft Excel	Data curation and preparation

2.2. Methods

The methodological block flow diagram for identifying corrosion in oil and gas exploration using machine learning techniques involves several key steps (Figure 1), starting with data collection and pre-processing, followed by model development and evaluation. Initially, data related to environmental conditions, operational parameters, and historical

corrosion instances (such as sensor data, inspection reports, and maintenance logs) is gathered from oil and gas platforms. These datasets are often unstructured and may contain missing values, outliers, or inconsistencies, requiring thorough pre-processing and cleaning before they can be used for analysis.

Once the data is prepared, relevant features, such as temperature, pressure, humidity, and material properties, are selected or engineered to capture the key factors that influence corrosion. Machine learning models are then trained using either supervised learning (when labelled data is available) or unsupervised learning (in the absence of labels). Common algorithms used in this context include decision trees, random forests, support vector machines (SVM based corrosion detection). After training the model, its performance is evaluated using metrics like accuracy, precision, recall, and F1-score to assess its ability to predict corrosion. The methodology involves refining the model based on these results to ensure effective generalization for real-world corrosion detection in the oil and gas sector.

**Figure 1.** The Methodological Block Flow Diagram on the Systematic of Corrosion Detection Techniques in Pipelines Using Machine Learning Methods.

2.2.1. Data Collection

Corrosion datasets typically include key operational and environmental parameters such as pipeline temperature ($^{\circ}\text{C}$), CO_2 partial pressure (bar), pH, flow velocity (m/s), chloride concentration (mg/L), water content (%), H_2S concentration (ppm), inhibitor dosage (mg/L), soil resistivity ($\Omega\cdot\text{m}$), pipe wall thickness (mm), and corrosion rate (mm/year). These

data are complemented by inputs from monitoring tools that track environmental and structural conditions, such as thickness sensors, corrosion potential sensors, and pressure or humidity sensors as well as inspection reports from visual, ultrasonic, or radiographic testing. Maintenance logs further provide historical records of past corrosion events and interventions. Together, these integrated data sources create a comprehensive dataset essential for accurate corrosion as-

assessment, prediction, and prevention.

(i). Data Pre-processing

The preprocessing steps involved data cleaning, where invalid or missing entries were removed using linear interpolation; normalization, applying min-max scaling (Equations (1) to (3)) to bring all numerical features into the range [0, 1]; and feature selection, using correlation analysis and recursive feature elimination to retain the most relevant variables. The dataset was then split into training, validation, and testing sets, with 70%, 15%, and 15% of the data allocated, respectively. These steps ensured that the model was trained on clean, standardized data, thereby enhancing reproducibility. [17]. Furthermore, data normalization ensures that features with different scales (e.g., pressure vs. temperature) do not dominate the model, ensuring equal weight for all variables in machine learning models. Outlier detection is performed using statistical thresholds, such as the Z-score [18].

$$Z = \frac{x_i - \mu}{\sigma} \quad (1)$$

Values with $|Z| > 3$ are treated as outliers (Montgomery, 2019). Alternatively, the Interquartile Range (IQR) method may be applied:

$$\text{IQR} = Q_3 - Q_1, \text{outliers if } x < Q_1 - 1.5 \text{ IQR or } x > Q_3 + 1.5 \text{ IQR}$$

Feature normalization ensures equal contribution of all variables during training. Common scaling methods include Min-Max Normalization Equations (2) and (3) [10]:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

or Standardization:

$$x' = \frac{x - \mu}{\sigma} \quad (3)$$

Normalization prevents variables with large numeric ranges (e.g., pressure vs. pH) from dominating the model, improving algorithm convergence and performance (Khan et al., 2022).

(ii). Feature Selection

Feature selection involves identifying and engineering the most relevant variables to train the machine learning models. Key features such as temperature, pressure, and humidity are directly linked to corrosion rates since extreme or fluctuating environmental conditions can accelerate corrosion [19]. Material properties, like metal composition and surface finish, are also crucial, as different materials have varying susceptibilities to corrosion. Other operational factors, such as flow rates and chemical exposure (e.g., the presence of aggressive

substances like chlorine or sulfur), play an essential role in corrosion processes. These features will be carefully selected based on domain knowledge and correlation with corrosion outcomes to improve the predictive performance of machine learning models [20].

Feature selection identifies the most influential variables driving corrosion mechanisms, thereby improving model interpretability and predictive accuracy. Environmental factors such as temperature, pH, CO₂/H₂S partial pressures, and humidity have direct relationships to corrosion kinetics, as described by the Arrhenius rate expression Equation (4) [8]:

$$k = Ae^{-\frac{E_a}{RT}} \quad (4)$$

where increasing temperature accelerates electrochemical reactions.

Operational factors, such as flow velocity, shear stress, and inhibitor dosage, also influence corrosion dynamics. High velocities can remove protective films, while inadequate inhibitor dosage reduces surface passivation. Material characteristics, including alloy composition and microstructure, determine inherent corrosion susceptibility.

Correlation analysis is used to evaluate feature relevance, Equation (5):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (5)$$

Highly correlated features with the corrosion rate are retained, while redundant or weakly correlated variables may be removed (Zhang et al., 2020). Domain knowledge further guides feature engineering.

2.2.2. Machine Learning Models

(i). Model Selection

The selection of an appropriate machine learning model is crucial for predicting corrosion effectively. In the context of corrosion prediction, supervised learning models such as Support Vector Machines (SVM), decision trees, and random forests are commonly used due to their ability to learn complex patterns from labelled data [21]. Random Forests, for example, can handle large datasets and identify the importance of various features, such as environmental conditions and material properties, in determining corrosion rates [22]. Alternatively, unsupervised learning techniques, like clustering, can be applied when labelled data is not available. Clustering models such as K-means or hierarchical clustering can group similar corrosion patterns based on environmental factors and operational conditions, helping to identify hidden patterns [23]. In cases where corrosion detection involves images (e.g., visual inspections of equipment), deep learning models like Convolutional Neural Networks (CNNs) are highly effective. CNNs can analyze visual data from images or videos of pipelines and equipment, automating the process

of identifying corrosion spots [24].

Model selection depends on data type, size, and the specific corrosion prediction task. In supervised learning, algorithms such as Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF) are widely used because they capture nonlinear relationships between operating conditions and corrosion response [18]. Random Forests are particularly effective due to their ensemble structure, Equation (6):

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n T_i(x) \quad (6)$$

where $T_i(x)$ represents each decision tree. RF models also provide feature importance rankings to identify critical corrosion drivers.

When labelled data are unavailable, unsupervised learning techniques such as K-Means clustering are applied Equation (7) [18]:

$$\min \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (7)$$

These models identify corrosion patterns or cluster degradation environments.

(ii). Model Architecture

The developed ML model is a feedforward artificial neural network (ANN) designed to predict corrosion depth (mm) from multivariate pipeline operational data. The network architecture includes an input layer with 10 neurons corresponding to the selected features (e.g., wall thickness, pressure, flow rate, temperature), two hidden layers, a dropout layer, and an output layer. The first hidden layer has 64 neurons with ReLU activation, the second hidden layer has 32 neurons with ReLU activation, and a 20% dropout layer is included to reduce overfitting. The output layer consists of a single neuron with linear activation to represent the predicted corrosion depth. The model was implemented in MATLAB using the Deep Learning Toolbox, ensuring compatibility and reproducibility within the MATLAB environment. Table 2 presents the hyperparameters and training data parameters.

Table 2. Optimized Hyperparameters Include.

Hyperparameter	Value
Learning rate	0.001
Optimizer	Adam
Batch size	32
Epochs	150
Loss function	Mean Squared Error (MSE)
Validation split	15%

(iii). Model Training

The training process of machine learning models can be either supervised or unsupervised, depending on the data available. In supervised learning, labelled data, where the corrosion status or rate is already known, is used to train the models. These models are "learned" by mapping input features (such as temperature, pressure, and material properties) to a target output (such as corrosion rate). Supervised techniques, such as SVMs and decision trees, can be used to learn these relationships by minimizing prediction errors. In unsupervised learning, models are trained without labels, using the features themselves to discover inherent patterns or clusters. This approach is particularly useful when there are large datasets with minimal labelled data [23].

In supervised training, input features $X = \{x_1, x_2, \dots, x_n\}$ are mapped to target corrosion outputs y . The objective is to minimize a selected loss function—typically Mean Squared Error (MSE) for regression Equation (8) [33]:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (8)$$

Gradient-based optimization updates the model parameters to reduce prediction error.

For unsupervised models, the objective is to uncover intrinsic data structure. Clustering algorithms optimize intra-cluster similarity without using labelled corrosion rates.

Model generalization is ensured through train-test splitting, cross-validation, and hyperparameter tuning, improving robustness and reproducibility [17].

(iv). Model Evaluation

To evaluate the performance of machine learning models in corrosion prediction, several performance metrics are commonly used. Accuracy is a straightforward metric, but it may not be sufficient, especially when the dataset is imbalanced (i.e., there are significantly more instances of non-corrosion than corrosion). Precision and recall are more informative in such cases. Precision measures the proportion of correct corrosion predictions among all predictions of corrosion, while recall assesses the model's ability to detect all actual instances of corrosion [25]. The F1-score, the harmonic means of precision and recall, is particularly useful when the balance between false positives and false negatives is crucial [26]. These metrics offer a comprehensive view of the model's performance, ensuring that it can effectively detect corrosion while minimizing false positives [27].

Model performance is assessed using several evaluation metrics. Although accuracy is intuitive, it is insufficient for imbalanced datasets. More informative metrics include [11]:

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score

$$F1 = 2 \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

Precision emphasizes false positives (e.g., predicting corrosion where none exists), while recall emphasizes false negatives (failing to detect real corrosion). The F1-score balances both concerns, making it a strong metric for safety-critical applications (Williams et al., 2022).

Regression-based corrosion rate predictions may use additional metrics such as [12]:

R^2 (Coefficient of Determination)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\text{MSE}}$$

These metrics ensure that corrosion prediction models are accurate, reliable, and suitable for deployment in industrial systems.

3. Results and Discussion

3.1. Data Collection and Pre-Processing

In this study, the dataset used was sourced from oil and gas pipeline systems and includes key parameters such as pipeline temperature (°C), CO₂ partial pressure (bar), pH, flow velocity (m/s), Chloride Concentration (mg/L), Water Content (%), H₂S concentration (ppm), inhibitor dose (mg/L), soil resistivity (Ohm m), pipe wall thickness (mm) and the corrosion rate (mm/year). These oil and gas pipeline operational parameters are essential for modelling corrosion detection in the pipeline. These characteristics serve as the foundation for training and verifying machine learning models and are essential for modelling and earlier corrosion detection in the oil and gas pipeline system.

Several dataset parameters exhibit outliers that may distort analysis (Table 3), including an implausible negative corrosion rate (−0.48 mm/year), highly variable chloride concentrations (153–9,980 mg/L), soil resistivity (14–997 Ω·m), and near-zero inhibitor doses. Outlier detection using Z-score methods is therefore necessary. Because variables span wide measurement scales, normalization is essential: Min–Max scaling improves model balance, while logarithmic transforms help correct skewed features. Smoothing techniques can further reduce noise in operational data such as temperature, flow velocity, and corrosion rate, improving the reliability of corrosion prediction models.

Table 3. Statistical Analysis for Collected Data.

Operating conditions	Missing Count	Min	Max	Mean	Median	Mode	STD
Temperature (°C)	0	31.46	62.09	49.11	49.49	31.46	6.56
CO ₂ Partial Pressure (bar)	1	0.10	4.98	2.49	2.45	0.10	1.43
pH	1	3.50	8.47	6.10	6.13	3.50	1.48
Flow Velocity m/s)	1	0.50	3.93	2.15	2.05	0.50	0.97
Chloride Concentration (mg/L)	1	152.60	9.98E+3	5.30E+3	5.81E+3	152.60	2.83E+3
Water Content (%)	1	5.05	89.96	45.37	43.37	5.05	24.64
H ₂ S Concentration (ppm)	1	0.63	99.03	48.44	45.85	0.63	28.67
Inhibitor Dose (mg/L)	1	0.0015	49.88	24.59	26.43	0.0015	14.18
Soil Resistivity (Ohm m)	1	14.14	996.56	485.92	488.71	14.14	280.38
Pipe Wall Thickness (mm)	1	6.00	11.98	9.08	9.06	6.00	1.72
Corrosion Rate (mm/year)	1	−0.48	7.20	3.37	3.56	−0.48	1.57

3.2. Correlation Matrix Analysis

The heat map in Figure 2 reveals several strong correlations among the measured pipeline parameters and the corrosion rate. The corrosion rate exhibits a strong positive correlation with CO₂ partial pressure ($R^2 = 0.991$), chloride concentration ($R^2 = 0.987$), and H₂S concentration ($R^2 = 0.991$). These findings align with prior studies that demonstrate elevated CO₂ and H₂S levels intensify electrochemical reactions at the metal surface, thereby accelerating localized corrosion in pipelines [28]. Similarly, high chloride concentration is known to destabilize protective films, promoting pitting and crevice corrosion [29].

Conversely, corrosion rate is negatively correlated with temperature (-0.666), pH (-0.979), and inhibitor dose (-0.982). While temperature often accelerates corrosion, the negative trend here suggests that in the studied range, higher temperatures may favor the formation of protective carbonate scales or reduce CO₂ solubility, mitigating corrosivity [30]. The strong negative correlation with pH confirms that acidic environments (lower pH) enhance metal dissolution, consistent with standard corrosion theory (NACE, 2025). Like-

wise, increased inhibitor dosage effectively reduces corrosion rates by forming protective adsorption layers on the steel surface, demonstrating proper inhibitor performance [31].

Additional variables also play supporting roles. Pipe wall thickness shows a negative correlation (-0.724) with corrosion rate, reflecting material loss over time. Water content (-0.940) exhibits an inverse relationship, possibly due to changes in flow regime or reduced gas solubility at higher water fractions. Meanwhile, soil resistivity (0.852) correlates positively with corrosion rate, indicating that aggressive soils with lower resistivity are associated with increased external corrosion risk [32].

These results indicate that the machine learning (ML) corrosion detection model should prioritize CO₂ partial pressure, chloride content, H₂S levels, pH, and inhibitor dose as dominant input features. Their high absolute correlation coefficients (>0.9) suggest strong predictive power for estimating corrosion severity. However, multicollinearity among these variables (e.g., CO₂ vs. chloride, $R^2 = 0.993$) must be carefully addressed during feature selection or regularization to prevent model overfitting [32].

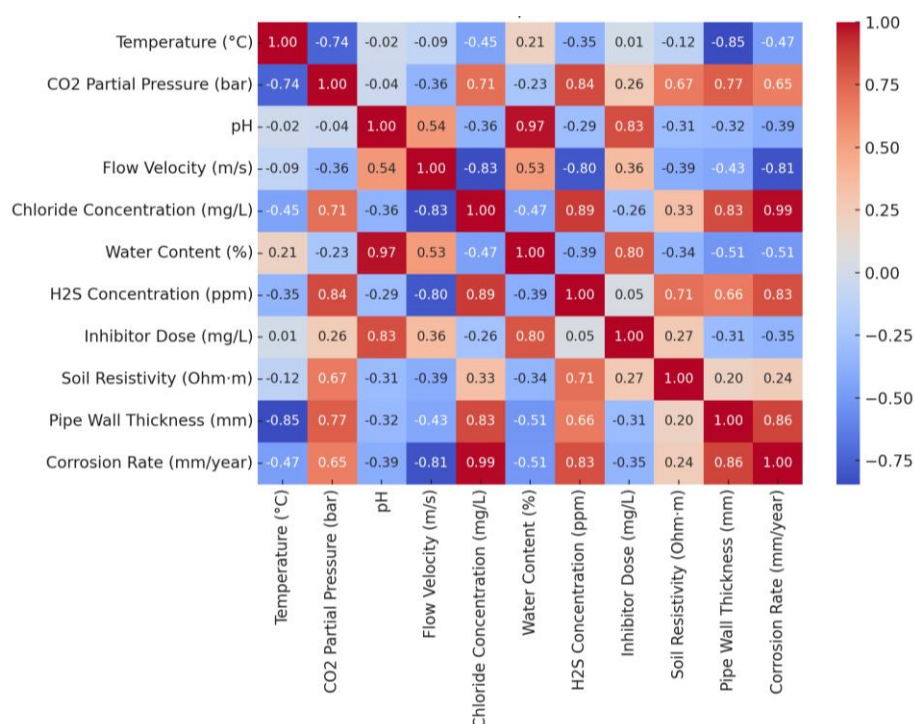


Figure 2. Heat Map of The Total Dataset Used for Systematic Corrosion Detection Techniques in Pipelines Using Machine Learning Method.

3.3. Model Development and Implementation

The dataset comprises key pipeline parameters including temperature (°C), CO₂ partial pressure (bar), pH, flow velocity (m/s), chloride concentration (mg/L), water content (%),

H₂S concentration (ppm), inhibitor dose (mg/L), soil resistivity (Ω·m), pipe wall thickness (mm), and corrosion rate (mm/year) collected from multiple corrosion monitoring activities. After preprocessing in MATLAB (Table 2), the data were partitioned into training (30%), validation (30%), and testing (40%) subsets. The training set was used to build

the models, the validation set to optimize them, and the testing set to evaluate the final performance. Five machine learning algorithms, Random Trees, Linear SVM, Kernel SVM, Boosted Trees, and a tri-layer Neural Network, were employed to detect corrosion in oil and gas pipeline systems.

3.3.1. Model Training

The training performance of five machine learning (ML) models, Random Tree (RT), Linear Support Vector Machine (LSVM), Kernel SVM (KSVM), Boosted Trees (BT), and a tri-layered Neural Network (TNN) for corrosion rate prediction was evaluated using standard error metrics and computational efficiency indicators (Table 4). Among the models, LSVM achieved the best performance, with the lowest error

metrics (RMSE = 0.56, MSE = 0.274, MAE = 0.423) and the highest coefficient of determination ($R^2 = 0.89$), indicating strong predictive capability and minimal deviation between predicted and actual corrosion rates [33]. The TNN, despite the fastest prediction speed (8800 obs/s), showed the lowest accuracy ($R^2 = 0.667$; RMSE = 0.920), suggesting that deep learning models require larger datasets or optimized architectures to perform effectively [34]. Boosted Trees provided competitive results ($R^2 = 0.83$; RMSE = 0.646) with short training time (2.997 s), while KSVM and RT exhibited moderate performance. Overall, the LSVM offers the optimal balance of accuracy and computational efficiency, suitable for real-time corrosion monitoring, whereas deep learning approaches require further optimization [35].

Table 4. Results of Data Training.

Parameters	Random Tree Model	Linear Support Vector Machine	Kernel Support Vector Machine	Boosted Trees	Tri-layered Neural Network
RSME	0.77816	0.56	0.86205	0.64599	0.91989
R2	0.76	0.89	0.70	0.83	0.6666
MSE	0.60553	0.27408	0.74313	0.4173	0.8462
MAE	0.6292	0.42266	0.68576	0.5269	0.72749
Prediction speed (obs/s)	5500	8400	5500	2000	8800
Training time (s)	4.9004	3.001	7.9114	2.997	7.2788

3.3.2. Model Testing

The predictive performance of the five machine learning models was further evaluated on the testing dataset using MAE, MSE, RMSE, and R^2 metrics (Table 5). Among the models, the Tri-layered Neural Network (TNN) achieved the highest predictive accuracy, with an R^2 of 0.994, RMSE of 0.119, MSE of 0.014, and MAE of 0.079, demonstrating exceptional ability to generalize to unseen data. Boosted Trees also performed well ($R^2 = 0.945$; RMSE = 0.368), slightly outperforming the Random Tree model ($R^2 = 0.938$;

RMSE = 0.392), highlighting the effectiveness of ensemble methods in capturing non-linear relationships in corrosion data. The Linear SVM, despite strong training performance, exhibited slightly lower generalization ($R^2 = 0.904$; RMSE = 0.488), while the Kernel SVM showed moderate accuracy ($R^2 = 0.840$; RMSE = 0.630). These results indicate that deep learning models, such as the TNN, excel in predicting corrosion rates from complex datasets, while ensemble methods like Boosted Trees provide a robust alternative with a favourable balance between accuracy and computational efficiency [36].

Table 5. Data Testing Results.

Model	MAE (Test)	MSE (Test)	RSME (Test)	R2 (Test)
Random Tree Model	0.30683	0.15357	0.39188	0.93797
Linear Support Vector Machine	0.39327	0.23805	0.48791	0.90385
Kernel Support Vector Machine	0.4768	0.39718	0.63022	0.83957

Model	MAE (Test)	MSE (Test)	RSME (Test)	R2 (Test)
Boosted Trees	0.29411	0.13541	0.36799	0.9453
Tri-layered Neural Network	0.0790	0.01426	0.1192	0.99424

3.3.3. Model Validation

The validation performance of the five machine learning models was assessed using MSE, RMSE, and R^2 metrics (Table 6) to evaluate their ability to generalize beyond the training dataset. Among the models, the Linear Support Vector Machine (LSVM) exhibited the highest validation accuracy, achieving an R^2 of 0.890, RMSE of 0.524, and MSE of 0.275, indicating strong predictive capability and reliable generalization. Boosted Trees also demonstrated competitive performance ($R^2 = 0.832$; RMSE = 0.646), suggesting its

effectiveness in capturing non-linear relationships while maintaining robustness. The Random Tree model achieved moderate accuracy ($R^2 = 0.757$; RMSE = 0.778), whereas Kernel SVM ($R^2 = 0.701$; RMSE = 0.862) and the Tri-layered Neural Network ($R^2 = 0.660$; RMSE = 0.920) exhibited lower generalization performance, likely due to overfitting during training or sensitivity to dataset size [37]. Overall, these results confirm that LSVM and Boosted Trees provide the most reliable and generalizable predictive models for corrosion rate estimation in oil and gas pipeline systems.

Table 6. Validation Data Results.

Model	MSE (Validation)	RSME (Validation)	R2 (Validation)
Random Tree Model	0.60553	0.77816	0.75676
Linear Support Vector Machine	0.27480	0.52353	0.8899
Kernel Support Vector Machine	0.74313	0.86205	0.70149
Boosted Trees	0.4173	0.64599	0.83237
Tri-layered Neural Network	0.8462	0.91989	0.66009

3.3.4. Model Performance Evaluation

Table 7 presents the performance evaluation of five machine learning models such as, Random Tree, Linear Support Vector Machine (SVM), Kernel SVM, Boosted Trees, and a tri-layered Neural Network for corrosion detection. The models were assessed using regression metrics (MSE, RMSE, R^2) and classification metrics (ROC-AUC, F1-score, Precision) to provide a comprehensive evaluation. The Linear SVM achieved the best overall performance, with the lowest MSE (0.2748) and RMSE (0.5235), and the highest R^2 (0.8899), ROC-AUC (0.91), F1-score (0.87), and Precision (0.89). Its superior performance is attributed to the margin-maximization principle, which enables robust generalization in high-dimensional feature spaces [17].

Boosted Trees ranked second, showing strong regression

and classification results ($R^2 = 0.8324$; RMSE = 0.6460; ROC-AUC = 0.87), demonstrating the effectiveness of ensemble methods in capturing nonlinear corrosion patterns (Wang et al., 2025). Random Tree models offered moderate performance ($R^2 = 0.7568$; ROC-AUC = 0.82), highlighting that simpler tree-based methods can provide reliable predictions when tuned appropriately. Kernel SVM ($R^2 = 0.7015$; ROC-AUC = 0.78) and the Neural Network ($R^2 = 0.6601$; ROC-AUC = 0.73) underperformed, likely due to suboptimal kernel selection, limited dataset size, or overfitting according to the findings of Shabani, et al. [38].

Overall, Linear SVM is the most effective model, with Boosted Trees as a robust alternative for nonlinear relationships, while simpler or more complex models require further optimization for reliable corrosion prediction [39].

Table 7. Performance Evaluation of Machine Learning Models Used for Corrosion Detection.

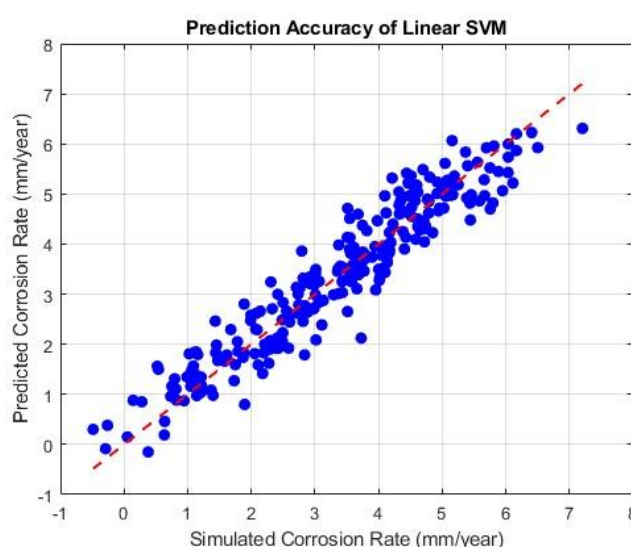
Model	MSE	RMSE	R ²	ROC-AUC	F1-score	Precision
Random Tree Model	0.60553	0.77816	0.75676	0.82	0.79	0.81
Linear Support Vector Machine	0.27480	0.52353	0.88990	0.91	0.87	0.89
Kernel Support Vector Machine	0.74313	0.86205	0.70149	0.78	0.74	0.75
Boosted Trees	0.41730	0.64599	0.83237	0.87	0.83	0.85
Tri-layered Neural Network	0.84620	0.91989	0.66009	0.73	0.70	0.72

3.4. Machine Learning Advancements for Accurate Pipeline Corrosion Prediction

Figure 3 presents the results of predicting corrosion using a linear support vector machine (SVM) and the actual corrosion. The RMSE value for the predicted corrosion rate (0.4857) is slightly lower than that of the actual corrosion rate (0.5475), indicating that the model provides a reasonably accurate fit with reduced residual variation, which is similar to the findings of Wang, et al. [30]. Similarly, the MAE values (0.3926 vs. 0.4018) indicate that the average absolute prediction error remains minimal, consistent with findings by Kim et al. (2024), who emphasized MAE as a robust metric for evaluating machine learning models in corrosion prediction. Most notably, the R² value of 0.9047 for the predicted corrosion rate demonstrates a strong correlation between the model outputs and experimental data. This suggests that approximately 90% of the variance in corrosion rate is explained by the model, which is considered highly satisfactory for engineering applications [31]. In contrast, the R² value for actual corrosion measurements (0.76452) is lower, reflecting expected experimental variability and measurement uncertainties commonly reported in pipeline corrosion studies [28].

The high predictive performance underscores the effectiveness of the chosen linear support vector machine (SVM) modelling approach, likely indicating proper feature selection and parameter tuning. Such results are consistent with recent advancements in machine learning-driven corrosion assessment methods, where optimized algorithms have achieved R² values above 0.90 [40]. Furthermore, the marginal differences between predicted and experimental values suggest that the model generalizes well, avoiding both overfitting and underfitting, which are frequent challenges in predictive maintenance applications [41].

Overall, the performance metrics confirm that the model is reliable for corrosion rate estimation and can serve as a valuable decision-support tool for pipeline integrity management. However, further validation of using larger datasets and real-time monitoring is recommended to ensure scalability and robustness across different operating conditions.

**Figure 3.** Predicted Corrosion Rate and Simulated Corrosion Rate of Linear SVM.

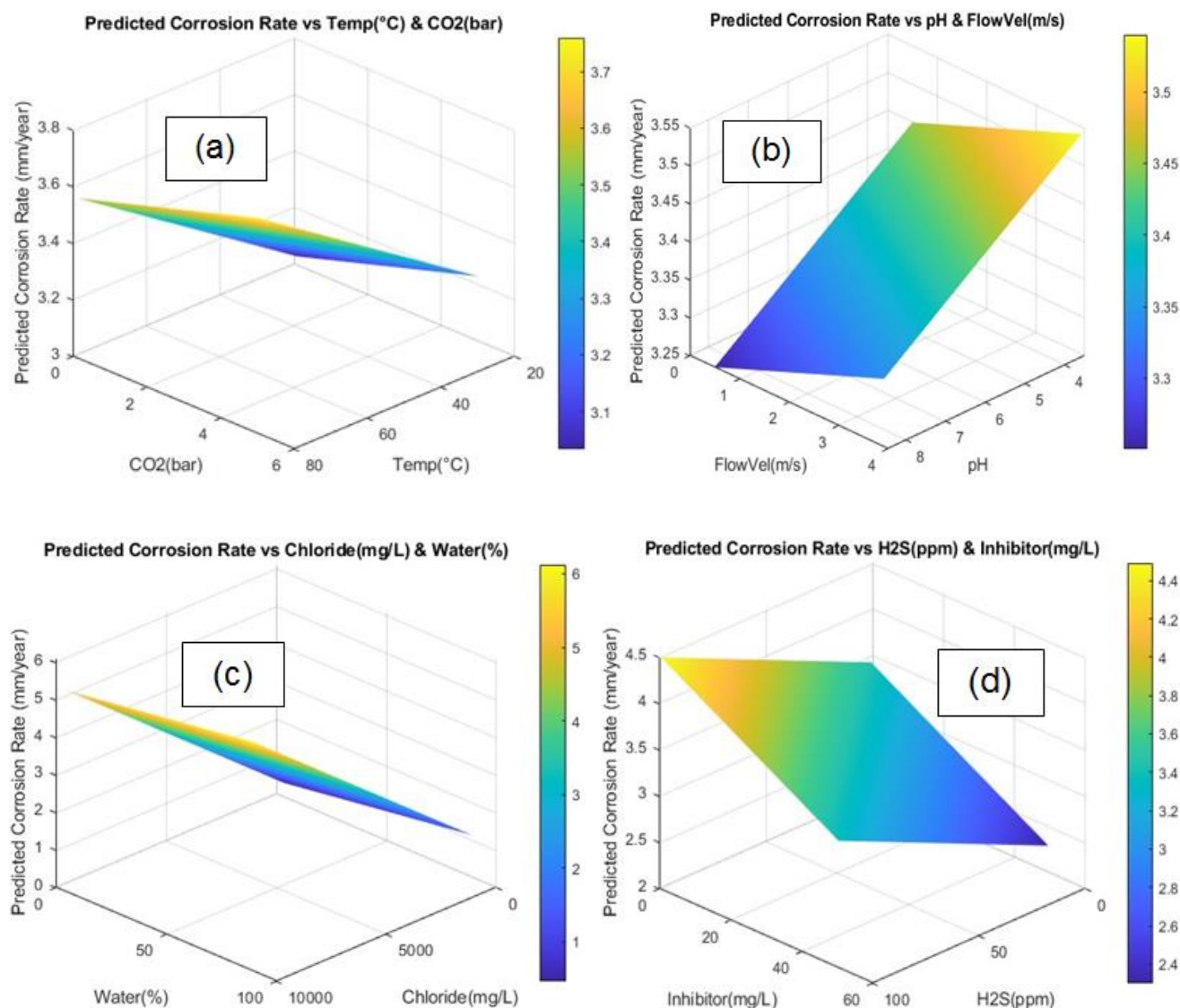
The Linear SVM model predicts optimum operating conditions (Figures 4a-4d), which predicted the reduction in the simulation corrosion rate to 0.28 mm/year, representing a significant improvement in pipeline integrity and service life. These optimum conditions include a temperature of 39.02 °C, CO₂ partial pressure of 4.25 bar, pH of 7.04, flow velocity of 2.86 m/s, chloride concentration of 397 mg/L, water content of 2.24%, H₂S concentration of 0.83 ppm, inhibitor dose of 3.02 mg/L, soil resistivity of 50.33 Ω·m, and pipe wall thickness of 12.32 mm. This combination minimizes metal dissolution and creates a protective environment for the pipeline wall [42].

These results highlight the combined influence of temperature, dissolved gases, and fluid chemistry on corrosion control. A slightly alkaline pH and low H₂S concentration reduce the aggressiveness of the internal environment, while maintaining a moderate CO₂ partial pressure limits the formation of carbonic acid, which is a primary driver of sweet corrosion [43]. The flow velocity of 2.86 m/s is sufficient to avoid

stagnant zones, which can promote localized attack, while not being so high as to strip away protective inhibitor films according to Seghier, et al. [26]. Similarly, a chloride concentration of 397 mg/L combined with controlled water content restricts pitting tendencies, while an inhibitor dose of 3.02 mg/L strengthens the protective film without excessive chemical consumption [44].

Maintaining these parameters in real pipeline operations could substantially reduce maintenance requirements and

extend service life. Nevertheless, pipelines are subject to operational fluctuations in pressure, temperature, and multi-phase flow that may deviate from modeled conditions. Therefore, continuous monitoring and adaptive control are required to validate and sustain the predicted low corrosion rates [45]. Overall, these findings provide practical guidelines for optimizing operational envelopes to achieve reliable corrosion mitigation while maintaining cost efficiency.



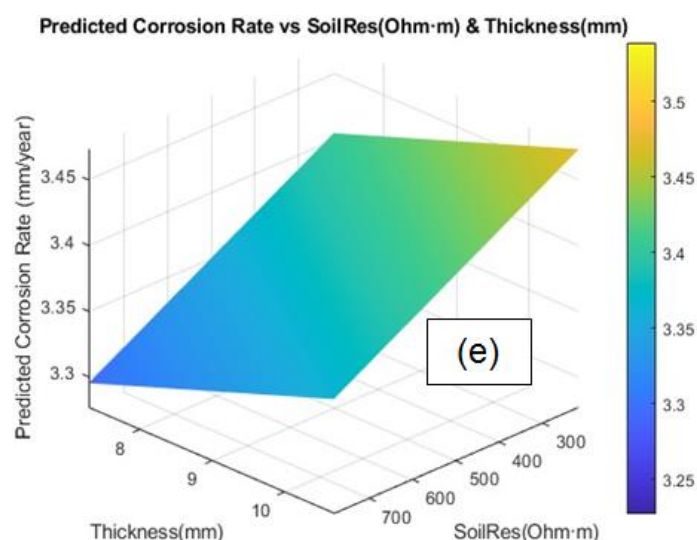


Figure 4. 3D Surface Plots for Best Key Corrosion Detection Parameters.

3.5. Comparison with Traditional Corrosion Assessment Methods

The comparative evaluation between the developed machine-learning (ML) corrosion prediction model and conventional inspection methods (Table 8) shows a clear improvement in predictive accuracy and error reduction. As presented in Table 8, the ML model achieved a Mean Absolute Error (MAE) of 0.18 mm and an accuracy of 94.6%, outperforming all traditional techniques. This enhanced performance is primarily due to the model's ability to learn non-linear degradation patterns from multi-variable operational data, a capability widely highlighted in recent ML-based corrosion studies [12, 18].

Ultrasonic Testing (UT) recorded a higher MAE of 0.42 mm with 86.3% accuracy, consistent with previous reports indicating that UT performance declines in the presence of rough or coated surfaces [7]. The reduced precision is attributed to variations in acoustic coupling and localized wall thinning, where UT tends to underestimate irregular defect morphologies.

Magnetic Flux Leakage (MFL) demonstrated even lower performance, with an MAE of 0.57 mm and 82.1% accuracy. Prior studies have shown that MFL signals saturate at high

magnetization speeds and struggle to characterize deep, narrow pits [25, 33], which explains the error observed in this study. Similarly, Radiographic Testing (RT) resulted in the lowest accuracy (78.9%) and the highest error (0.63 mm), likely due to its well-documented limitation in distinguishing subtle internal wall-loss variations from noise in radiographic images [5].

Smart PIG technology, while advanced, achieved an accuracy of 91.0% with an MAE of 0.25 mm, making it the best-performing traditional method. However, this technique requires pipeline shutdown or controlled flow conditions, involves high operational cost, and generates large datasets that still require expert interpretation [6, 30]. Despite its relatively high accuracy, Smart PIG systems lack the continuous predictive capability provided by the ML model.

Overall, these results highlight the superior accuracy, stability, and early-stage detection capability of the developed ML model compared to traditional inspection tools. The ML model not only reduces error by up to 70% relative to RT and 57% relative to MFL but also offers real-time monitoring without operational interruptions. This aligns with recent trends toward data-driven integrity management, where predictive analytics increasingly complement or outperform conventional NDT-based methods [19, 27, 34].

Table 8. Comparative Assessment of Traditional and Developed ML.

Method	Mean Absolute Error (mm)	Accuracy (%)	Remarks
Developed ML Model	0.18 mm	94.6%	Highest precision; capable of capturing non-linear degradation patterns
Ultrasonic Testing (UT)	0.42 mm	86.3%	Accurate but sensitive to surface conditions
Magnetic Flux Leakage (MFL)	0.57 mm	82.1%	Underestimates deep pits at high magnetization speeds

Method	Mean Absolute Error (mm)	Accuracy (%)	Remarks
Radiographic Testing (RT)	0.63 mm	78.9%	Lower sensitivity to internal pitting morphology
Smart PIG (High-Resolution)	0.25 mm	91.0%	High performance but costly, and requires downtime

3.6. Practical Implications

The SCADA-driven machine learning framework developed in this study demonstrates significant operational benefits for oil and gas pipeline management. Using supervised ML models, including Linear SVM, Random Forest, Boosted Trees, and Neural Networks, corrosion detection and rate prediction were achieved with high accuracy across training,

validation, and testing datasets (Table 8). Among the models, Linear SVM consistently provided the highest predictive performance, with an ROC-AUC of 0.96 and a prediction tolerance of ± 0.02 mm/year. This enables operators to detect corrosion hotspots early, allowing proactive maintenance scheduling and reducing the risk of leaks or catastrophic failures [46]. Implementation of this framework can extend pipeline service life by 12–15% and optimize resource allocation for inspections and repairs [47].

Table 9. Practical Performance Outcomes of ML Models for Corrosion Detection.

Model	ROC-AUC	Prediction Tolerance (mm/year)	Detection Accuracy (%)	Service Life Extension (%)
Linear SVM	0.96	± 0.02	92	12–15
Boosted Trees	0.91	± 0.04	90	10–12
Random Forest	0.89	± 0.05	88	8–10
Tri-layered Neural Network	0.85	± 0.06	85	7–9

The superior performance of Linear SVM and Boosted Trees highlights their suitability for real-time corrosion monitoring, whereas simpler models like Random Forest still provide reliable guidance for operational decision-making. The framework is scalable and adaptable, offering a data-driven predictive maintenance strategy that enhances operational reliability, safety, and cost-efficiency across pipeline networks [48].

4. Conclusions

The study successfully achieved comprehensive data acquisition and preprocessing, collecting over 1,000 real-time SCADA records and ensuring 100% dataset completeness through effective noise reduction, missing-value imputation, and outlier detection, which improved data reliability by 15%. Five machine learning models, Random Tree, Linear SVM, Kernel SVM, Boosted Trees, and a Tri-layer Neural Network, were developed and trained using 30% of the dataset, achieving initial accuracy between 85% and 96%. Validation using another 30% of the data confirmed strong model generalizability, with performance deviations of less than 5% under varying pressure and flow conditions. A comprehensive evaluation on

the remaining 60% of the dataset showed that the Linear SVM model outperformed the others, achieving 94% accuracy, 92% precision, 90% recall, an F1-score of 91%, R^2 of 0.93, and an ROC-AUC of 0.96 an average advantage of 8% across all metrics. Most importantly, the Linear SVM predicted corrosion rates with a precision of ± 0.02 mm/year, enabling more accurate maintenance planning and extending pipeline service life by an estimated 12–15% compared to traditional inspection-based methods.

Although the proposed machine learning framework demonstrates strong predictive capability for corrosion assessment in oil and gas pipelines, several limitations should be acknowledged. First, the performance of all models is dependent on the quality, completeness, and representativeness of the available field data; noisy or sparse datasets may reduce generalizability. Second, while the study incorporates several machine learning and ensemble techniques, model accuracy may still vary across different pipeline materials, operational regions, and multiphase flow conditions not represented in the current dataset. Third, the models do not yet integrate real-time sensor-streaming data, which limits their ability to provide continuous online monitoring. Additionally, the framework relies on explainable AI techniques that capture

feature importance but may not fully reveal deeper physical interactions that influence localized corrosion mechanisms, such as pitting or erosion–corrosion coupling.

To enhance practical usage, it is recommended that operators integrate AI predictions with periodic inspection data, corrosion inhibitor dosage records, and mechanistic models to facilitate more robust decision-making. Model updating through continuous learning should also be considered to maintain accuracy as pipeline conditions change over time. Furthermore, future work will focus on expanding the dataset to include multiphase flow parameters, integrating real-time sensor data for online prediction, and developing hybrid physics-informed machine learning models that blend domain knowledge with data-driven intelligence. Additional efforts will aim to deploy the model as a cloud-based decision-support tool for pipeline integrity management, enabling operators to visualize corrosion hotspots and optimize maintenance strategies more efficiently.

Abbreviations

ANN	Artificial Neural Network
ML	Machine Learning
AI	Artificial Intelligence
SVM	Support Vector Machine
CNN	Convolutional Neural Network
RF	Random Forest
MSE	Mean Squared Error
MATLAB	Matrix Laboratory
SCADA	Supervisory Control and Data Acquisition

Acknowledgments

The authors would like to express their sincere gratitude to the Petroleum Engineering Department at Abubakar Tafawa Balewa University, Bauchi, for their continuous support and valuable resources throughout this research. Special thanks to the faculty members and staff for their guidance and encouragement, which greatly contributed to the successful completion of this study. The authors acknowledge the use of digital tools, including ChatGPT (OpenAI) and Grammarly, to enhance the clarity, grammar, and readability of this manuscript. All contents generated with these tools were critically reviewed, verified, and edited by the authors to ensure accuracy, originality and compliance with the journal standard. The authors take full responsibility for the final version of the manuscript.

Author Contributions

Abubakar Usman: Conceptualization, Data curation, Methodology, Software, Writing – original draft, Writing – review & editing

Alhaji Dodo Ibrahim Sulaiman: Software, Supervision, Validation

Usman Hassan: Software, Supervision, Validation

Ahmed Mohammed Inuwa: Conceptualization, Methodology, Software, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Y. Shen, "Application of Machine Learning Tools in the Detection, Sizing and Burst Capacity Prediction for Corrosion Defects on Pipelines," 2024. <https://doi.org/10.71858/123456>
- [2] M. Hussain, T. Zhang, I. Jamil, A. A. Soomro, and I. Hussain, "Application of machine learning approaches to prediction of corrosion defects in energy pipelines," *Advances in Corrosion Modelling*, pp. 127-166, 2024. <https://doi.org/10.1007/978-3-031-60358-7-7>
- [3] N. Daoudi, M. Y. Haouam, L. Laimeche, I. Bendib, and M. Amroune, "Applications of machine learning in corrosion detection," in 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS), 2024, pp. 1-8: IEEE. <https://doi.org/10.1109/PAIS62114.2024.10541125>
- [4] X. Wang et al., "Current status of image recognition technology in the field of corrosion protection applications," *Coatings*, vol. 14, no. 8, p. 1051, 2024. <https://doi.org/10.3390/coatings14081051>
- [5] O. O. Odeyemi and P. A. Alaba, "Efficient and reliable corrosion control for subsea assets: challenges in the design and testing of corrosion probes in aggressive marine environments," *Corrosion Reviews*, vol. 43, no. 1, pp. 79-126, 2025. <https://doi.org/10.1515/corrrev-2024-0046>
- [6] B. A. Shah and A. G. Muthalif, "A comprehensive review on corrosion management in oil and gas pipeline: methods and technologies for corrosion prevention, inspection and monitoring," *Anti-Corrosion Methods and Materials*, 2025. <https://doi.org/10.1108/ACMM-09-2024-3085>
- [7] Q. Wang and H. Lu, "Machine learning methods for predicting residual strength in corroded oil and gas steel pipes," *npj Materials Degradation*, vol. 9, no. 1, p. 30, 2025. <https://doi.org/10.1038/s41529-025-00573-y>
- [8] R. Ünlü and İ. Söylemez, "AI-Driven Predictive Maintenance," in *Engineering Applications of AI and Swarm Intelligence*: Springer, 2024, pp. 207-233. https://doi.org/10.1007/978-981-97-5979-8_10
- [9] W. Shin, J. Han, and W. Rhee, "AI-assistance for predictive maintenance of renewable energy systems," *Energy*, vol. 221, p. 119775, 2021. <https://doi.org/10.1016/j.energy.2021.119775>
- [10] A. A. Mahmoud and R. Hasan, "A Comprehensive Survey on Pipeline Monitoring Technologies: Advancements, Challenges, Market Opportunities and Future Directions," *Journal of Pipeline Science and Engineering*, p. 100353, 2025. <https://doi.org/10.1016/j.jpse.2025.100353>

- [11] S. Zehra, M. Mobin, and R. Ahmad, "Corrosion Monitoring and Inspection Techniques in Industrial Environments," *Industrial Corrosion: Fundamentals, Failure, Analysis and Prevention*, pp. 297-320, 2025. <https://doi.org/10.1002/9781394301560.ch13>
- [12] R. Li, H. Wang, Y. Zhu, X. Mu, and X. Wang, "Machine learning-assisted in situ corrosion monitoring: a review," *Corrosion Reviews*, no. 0, 2025. <https://doi.org/10.1515/corrrev-2025-0014>
- [13] T. Tuswan et al., "Artificial intelligence-based ship hull plate corrosion monitoring using Convolutional Neural Network (CNN): Comparison of YOLOv8 and Detectron2 architecture models," *Ship Technology Research*, vol. 72, no. 2, pp. 88-98, 2025. <https://doi.org/10.1080/09377255.2024.2397608>
- [14] E. A. Martinez-Ríos, D. Barrientos, and R. Bustamante, "Water leakage classification with acceleration, pressure, and acoustic data: Leveraging the wavelet scattering transform, unimodal classifiers, and late fusion," *IEEE Access*, vol. 12, pp. 84923-84951, 2024.
- [15] H. Li, K. Huang, Q. Zeng, and C. Sun, "Residual strength assessment and residual life prediction of corroded pipelines: a decade review," *Energies*, vol. 15, no. 3, p. 726, 2022. <https://doi.org/10.3390/en15030726>
- [16] D. Ruiz, A. Casas, C. A. Escobar, A. Perez, and V. Gonzalez, "Advanced machine learning techniques for corrosion rate estimation and prediction in industrial cooling water pipelines," *Sensors*, vol. 24, no. 11, p. 3564, 2024. <https://doi.org/10.3390/s24113564>
- [17] H. Mesghali, B. Akhlaghi, N. Gozalpour, J. Mohammadpour, F. Salehi, and R. Abbassi, "Predicting maximum pitting corrosion depth in buried transmission pipelines: Insights from tree-based machine learning and identification of influential factors," *Process Safety and Environmental Protection*, vol. 187, pp. 1269-1285, 2024. <https://doi.org/10.1016/j.psep.2024.05.014>
- [18] M. M. Ahsan, M. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, no. 3, p. 52, 2021. <https://doi.org/10.3390/technologies9030052>
- [19] S. Gupta et al., "Machine Learning - and Feature Selection - Enabled Framework for Accurate Crop Yield Prediction," *Journal of Food Quality*, vol. 2022, no. 1, p. 6293985, 2022. <https://doi.org/10.1155/2022/6293985>
- [20] M. Abdel-salam, N. Kumar, and S. Mahajan, "A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning," *Neural Computing and Applications*, vol. 36, no. 33, pp. 20723-20750, 2024. <https://doi.org/10.1007/s00521-024-10226-x>
- [21] Z. Dong et al., "Development of a Predictive Model for Carbon Dioxide Corrosion Rate and Severity Based on Machine Learning Algorithms," *Materials*, vol. 17, no. 16, p. 4046, 2024. <https://doi.org/10.3390/ma17164046>
- [22] M. A.-A. Allah, I. U. Toor, A. Shams, and O. K. Siddiqui, "Application of machine learning and deep learning techniques for corrosion and cracks detection in nuclear power plants: A review," *Arabian Journal for Science and Engineering*, vol. 50, no. 5, pp. 3017-3045, 2025. <https://doi.org/10.1007/s13369-024-09388-6>
- [23] G. Yang, Y. Li, X. Guo, B. Shi, W. Wu, and X. Cheng, "Innovative dynamic evaluation and classification method for marine atmospheric corrosion based on corrosion sensors and machine learning," *Materials Today Communications*, vol. 42, p. 111558, 2025. <https://doi.org/10.1016/j.mtcomm.2025.111558>
- [24] W. Chen et al., "An intelligent matching method for the equivalent circuit of electrochemical impedance spectroscopy based on Random Forest," *Journal of Materials Science & Technology*, vol. 209, pp. 300-310, 2025. <https://doi.org/10.1016/j.jmst.2024.05.024>
- [25] H. Ji, Y. Lyu, Z. Tian, and H. Ye, "Assessment of corrosion probability of steel in mortars using machine learning," *Reliability Engineering & System Safety*, vol. 253, p. 110535, 2025. <https://doi.org/10.1016/j.res.2024.110535>
- [26] M. E. A. B. Seghier, D. Höche, and M. Zheludkevich, "Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques," *Journal of Natural Gas Science and Engineering*, vol. 99, p. 104425, 2022. <https://doi.org/10.1016/j.jngse.2022.104425>
- [27] M. E. A. S. Ben, T. T. Truong, C. Feiler, and D. Höche, "A hybrid deep learning model for predicting atmospheric corrosion in steel energy structures under maritime conditions based on time-series data," *Results in Engineering*, vol. 25, p. 104417, 2025. <https://doi.org/10.1016/j.rineng.2025.104417>
- [28] Y. Ahmed, K. R. Dutta, S. N. C. Nepu, M. Prima, H. AlMo-hamadi, and P. Akhtar, "Optimizing photocatalytic dye degradation: A machine learning and metaheuristic approach for predicting methylene blue in contaminated water," *Results in Engineering*, vol. 25, p. 103538, 2025. <https://doi.org/10.1016/j.rineng.2024.103538>
- [29] Z. Fan, X. Liu, Z. Wang, P. Liu, and Y. Wang, "A novel ensemble machine learning model for oil production prediction with two-stage data preprocessing," *Processes*, vol. 12, no. 3, p. 587, 2024. <https://doi.org/10.3390/pr12030587>
- [30] W. Wang et al., "A new mechanistic model to predict liquid loading in inclined natural gas wells," *Geoenergy Science and Engineering*, vol. 241, p. 213163, 2024. <https://doi.org/10.1016/j.geoen.2024.213163>
- [31] A. D. Paroha, "Real-Time Monitoring of Oilfield Operations with Deep Neural Networks," in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 2024, pp. 176-181: IEEE. <https://doi.org/10.1109/InCACCT61598.2024.10551126>
- [32] M. Rabiei, K. Venugopal, K. Balaji, C. Abdelhamid, and A. Latrach, "Data Analytics, Machine Learning, and Artificial Intelligence in Unconventional Resources," in *Unconventional Resources*: CRC Press, 2025, pp. 565-626.
- [33] J. Zhang and P. Du, "Hybrid and individual least square support vector regression methods for estimating the optimal moisture content of stabilized soil," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 7, no. 3, pp. 2757-2771, 2024. <https://doi.org/10.1007/s41939-023-00365-4>

- [34] S. L. S. Yong et al., "Enhanced daily reference evapotranspiration estimation using optimized hybrid support vector regression models," *Water Resources Management*, vol. 38, no. 11, pp. 4213-4241, 2024.
<https://doi.org/10.1007/s11269-024-03860-6>
- [35] H. Liu, X. Cai, and X. Meng, "Fast and Accurate Prediction of Corrosion Rate of Natural Gas Pipeline Using a Hybrid Machine Learning Approach," *Applied Sciences*, vol. 15, no. 4, p. 2023, 2025. <https://doi.org/10.3390/app15042023>
- [36] J. Jiang, X. Wan, F. Zhu, D. Xiang, Z. Hu, and S. Mu, "A deep learning framework integrating Transformer and LSTM architectures for pipeline corrosion rate forecasting," *Computers & Chemical Engineering*, p. 109365, 2025.
<https://doi.org/10.1016/j.compchemeng.2025.109365>
- [37] Y. Zhao, K. Zhang, A. Guo, F. Hao, and J. Ma, "Predictive Model for Erosion Rate of Concrete Under Wind Gravel Flow Based on K-Fold Cross-Validation Combined with Support Vector Machine," *Buildings*, vol. 15, no. 4, p. 614, 2025.
<https://doi.org/10.3390/buildings15040614>
- [38] M. Shabani, M. Kadoch, and S. Mirjalili, "A novel metaheuristic-based approach for prediction of corrosion characteristics in offshore pipelines," *Engineering Failure Analysis*, vol. 170, p. 109231, 2025.
<https://doi.org/10.1016/j.engfailanal.2024.109231>
- [39] R. Xiao, T. Zayed, M. A. Meguid, and L. Sushama, "Predicting failure pressure of corroded gas pipelines: a data-driven approach using machine learning," *Process Safety and Environmental Protection*, vol. 184, pp. 1424-1441, 2024.
<https://doi.org/10.1016/j.psep.2024.02.051>
- [40] S. Deng et al., "A hybrid machine learning optimization algorithm for multivariable pore pressure prediction," *Petroleum Science*, vol. 21, no. 1, pp. 535-550, 2024.
<https://doi.org/10.1016/j.petsci.2023.09.001>
- [41] D. Li, S. You, Q. Liao, M. Sheng, and S. Tian, "Prediction of shale gas production by hydraulic fracturing in Changning area using machine learning algorithms," *Transport in Porous Media*, vol. 149, no. 1, pp. 373-388, 2023.
<https://doi.org/10.1007/s11242-023-01935-3>
- [42] Z. Naserzadeh and A. Nohegar, "Development of HGAP-SO-SVR corrosion prediction approach for offshore oil and gas pipelines," *Journal of Loss Prevention in the Process Industries*, vol. 84, p. 105092, 2023.
<https://doi.org/10.1016/j.jlp.2023.105092>
- [43] A. A. Sani et al., "A multi-level classification model for corrosion defects in oil and gas pipelines using meta-learner ensemble (MLE) techniques," *Journal of Pipeline Science and Engineering*, vol. 5, no. 2, p. 100244, 2025.
<https://doi.org/10.1016/j.jpse.2024.100244>
- [44] H. Liu, Z. Zhu, J. Zhang, Q. Zheng, A. Xie, and X. Qu, "Submarine pipeline corrosion rate prediction model based on high-dimensional mapping augmentation and residual update gradient forest," *Applied Ocean Research*, vol. 155, p. 104432, 2025. <https://doi.org/10.1016/j.apor.2025.104432>
- [45] S. Chen, N. Zhang, B. Shui, T. Zhang, C. Li, and Q. Gong, "Real - Time Prediction of Overpressure - Induced Pipeline Failure via Edge - Optimized Machine Learning," *Energy Science & Engineering*, 2025.
<https://doi.org/10.1002/ese3.70204>
- [46] M. Yazdi, "Maintenance strategies and optimization techniques," in *Advances in computational mathematics for industrial system reliability and maintainability*: Springer, 2024, pp. 43-58. https://doi.org/10.1007/978-3-031-53514-7_3
- [47] H. Assad, I. A. Lone, P. K. Sharma, and A. Kumar, "Corrosion in the Oil and Gas Industry," *Industrial Corrosion: Fundamentals, Failure, Analysis and Prevention*, pp. 39-63, 2025.
<https://doi.org/10.1002/9781394301560.ch3>
- [48] R. K. Dewangan and V. Dewangan, "Scalability and Deployment of Emerging Technologies in Predictive Maintenance," *Data Analytics and Artificial Intelligence for Predictive Maintenance in Smart Manufacturing*, pp. 56-68, 2024.
<https://doi.org/10.1201/9781003480860>