# MACHINE LEARNING-BASED PREDICTION OF EXTERNAL CORROSION RATES IN BURIED UNPROTECTED GAS PIPELINES

**Waziri, A.M., Guma, T.N., Akindapo, J.O. and Orueri, D.U.***

Mechanical Engineering Department, Nigerian Defence Academy, Kaduna, Nigeria

*Corresponding email: danielorueri@nda.edu.ng

**ABSTRACT**

*External corrosion remains a critical threat to the long-term durability of buried gas pipelines, especially under varying environmental conditions such as soil pH, moisture content, and relative humidity. This study presents a predictive framework for estimating external corrosion growth rates along the Ajaokuta–Kaduna–Kano (AKK) gas pipeline route in Nigeria. A combination of Computational Fluid Dynamics (CFD) simulations and supervised Machine Learning (ML) models was employed to capture and learn from environmental variability across five locations—Ajaokuta, Lokoja, Abaji, Kaduna, and Kano. Input features included soil properties, climate data, and corrosion history. Among the five tested ML algorithms, the XGBoost model exhibited the best predictive performance, achieving an $R^2$ score of 0.95 and a Mean Absolute Error (MAE) of 0.0065 mm/yr in Kano. Key predictors were identified as soil pH, moisture content, and relative humidity. This research delivers a novel, location-sensitive approach to external corrosion prediction and offers a reliable basis for geo-targeted maintenance planning in Nigeria's oil and gas sector.*

**Keywords:** External Corrosion, Soil-Pipeline Interaction, Computational Fluid Dynamics, Machine Learning Models, Pipeline Integrity, XGBoost.

## 1 INTRODUCTION

External corrosion is one of the most pervasive degradation mechanisms affecting buried pipelines in the oil and gas sector, particularly in regions with high soil variability and aggressive environmental conditions [1, 2]. This form of corrosion is typically influenced by factors such as soil pH, moisture content, salinity, and the presence of corrosive agents like chloride and sulfate ions [3, 4]. For large-scale infrastructure like the Ajaokuta–Kaduna–Kano (AKK) gas pipeline, which spans multiple ecological zones, the risk of corrosion is highly location-dependent. Conventional protection measures such as coatings and cathodic protection systems provide some defense but are not foolproof, especially in the presence of damaged coatings, heterogeneous soil compositions, or stray current interference [3, 5]. Furthermore, routine inspection and maintenance of underground pipelines are often constrained by cost, accessibility, and limited availability of real-time data [6, 7]. Emerging predictive approaches now leverage Computational Fluid Dynamics (CFD) and Machine Learning (ML) to overcome these limitations by modeling corrosion as a function of environmental and operational inputs [8, 9, 10]. CFD allows the simulation of soil-pipeline interactions under varying temperature, humidity, and pH scenarios, generating synthetic datasets to supplement limited field measurements [4, 11]. ML models, particularly ensemble learners such as XGBoost and Random Forest, have shown high accuracy in forecasting corrosion trends and identifying high-risk zones along pipeline routes [12, 13, 14]. Despite these advancements, few studies have focused on external corrosion prediction in Nigerian pipelines, where climatic diversity and soil chemistry vary significantly across regions [15, 16]. This study addresses that gap by applying CFD-augmented ML modeling to predict external corrosion rates along five representative locations on the AKK pipeline corridor: Ajaokuta, Lokoja, Abaji, Kaduna, and Kano. Environmental datasets—comprising soil pH, moisture content, temperature, and relative humidity—were extracted and used to train and evaluate five ML models, with XGBoost outperforming the others in predictive capability. The findings of this research are expected to enhance geo-targeted corrosion mitigation strategies, optimize resource allocation, and enable data-driven decision-making for long-term pipeline integrity. By providing a robust framework for external corrosion prediction in environmentally diverse terrains, this work supports the modernization of Nigeria's pipeline maintenance and asset management practices [17, 18].

## 2 LITERATURE REVIEW

External corrosion of buried pipelines is predominantly driven by environmental exposure—especially soil chemistry, moisture content, and climatic conditions. As such, predictive modeling must account for geo-environmental variability and its influence on pipeline deterioration. Historically, corrosion assessment relied on deterministic or probabilistic models that generalized soil–pipeline interactions. However, studies like those by Li *et al*. [4] revealed that external corrosion rates are significantly affected by coating conditions and environmental interactions such as temperature, salt concentration, and pH. Their findings underscore the need for adaptive models capable of capturing such localized variations. The application of machine learning (ML) in this space has grown, with Canonaco *et al*. [8] and Khoei and Kaabouch [18] advocating for supervised learning in pipeline corrosion modeling. Their work emphasized that ML algorithms, trained on region-specific datasets, can outperform traditional models in forecasting corrosion risk. In Nigeria, the variability of soil types and climatic zones along pipeline routes such as Ajaokuta–Kaduna–Kano (AKK) necessitates location-sensitive models. Ba [3] highlighted how clayey and sandy soils with high moisture content accelerate corrosion, a finding supported by the study's own CFD simulations. Further, Zuo [17] and Soomro *et al*. [14] proposed hybrid reliability models combining time-series and ML methods for residual life estimation, concluding that grey theory and ensemble learners offer superior adaptability in predicting degradation over time. Additionally, Sheikh *et al*. [13] applied acoustic emissions with ML to detect and quantify external corrosion severity, using Naive Bayes and RBF-NN models with up to 100% accuracy in controlled experiments. Their approach showed the feasibility of coupling real-time monitoring with intelligent models. Environmental feature engineering is a critical component in these frameworks. [5, 10] demonstrated how selecting key environmental predictors—like humidity, pH, and temperature—significantly improved corrosion prediction accuracy. These findings guided the feature selection process in the current study, focusing on moisture content, relative humidity, soil pH, and temperature distribution. This research extends these developments by building and validating ML models trained on region-specific environmental data from five key AKK locations, leveraging CFD simulations to augment data scarcity and improve predictive reliability for external corrosion risk mapping.

## 3 MATERIALS AND METHODS

This section presents the methodological framework for predicting external corrosion rates along the AKK gas pipeline. External corrosion was modeled based on soil-pipeline interactions, environmental exposure, and electrochemical conditions. CFD simulations and machine learning were integrated to develop a robust predictive model.

### 3.1 Materials

Computational modeling was conducted on an HP Elite Dragonfly G2 Notebook PC. Software Tools: ANSYS Fluent 2022-R1: For simulating external corrosion in soil environments. Anaconda Python Distribution: For machine learning workflows.

## 3.2 Methods

### *3.2.1 Data Acquisition and Preprocessing*

Data were categorized into key environmental and physical variables that influence external corrosion. Table 1 summarizes the data sources.

Table 1. External Corrosion Data and Collection Methods

| S/N | Data Type | Variables | Source/Method |
|---|---|---|---|
| 1 | Pipeline Characteristics | Material, coating, installation history | NNPC archives, site inspection |
| 3 | Soil Properties | Moisture content, type, pH | Geotechnical surveys, lab tests |
| 4 | Environmental Factors | Temperature, humidity, rainfall, corrosive gases | Meteorological data, site visits |
| 5 | Corrosion Metrics | Corrosion rate, morphology | CFD analysis, NNPC data |
| 6 | Maintenance Records | Repairs, coating condition, failure incidents | Operator logs, technical reports |

### *3.2.2 Soil and Environmental Conditions*

Environmental and geotechnical data were collected for five pipeline regions: Kaduna, Ajaokuta, Lokoja, Abaji, and Kano. Parameters included: Soil pH: 4.5 – 8.7. Moisture content: 13% – 32%. Average temperatures: 28°C – 31°C. Relative humidity: 40% – 60%. Rainfall: 1.0 – 1.75m annually.

*CFD Simulation of External Corrosion:* The CFD simulation procedure for external corrosion was carried out in the following steps:

i) Geometry and Meshing: A 3D domain (6 m × 2 m × 2 m) representing the soil-pipeline interface was designed in SolidWorks. Unstructured tetrahedral mesh with surface refinement was used in ANSYS Fluent.

ii) Boundary and Initial Conditions: Soil: Modeled as a porous medium (moisture: 10%–35%). Environmental parameters: Temperature = 288–313 K, RH = 60%–90%, pH = 4.5–7.0. Electrochemical setup: Implemented via User-Defined Functions (UDFs).

iii) Solver Configuration: Multiphase model: Volume of Fluid (VOF) for simulating water penetration. Heat transfer: Enabled using the energy equation. Corrosion mechanism: Pitting corrosion modeled as dominant.

iv) Post-Processing and Results: Output parameters: Soil pH distribution, moisture concentration, corrosion-prone zones. Predicted external corrosion rate: 0.022–0.041 mm/year.

## 3.3 Machine Learning Model Development

Simulation outputs and environmental data were used to train five ML models (ANN, SVM, XGBoost, Random Forest, Regression) in Python.

## 3.4 Model Evaluation and Optimization

Model performance was evaluated using cross-validation and standard error metrics. The most accurate model was adopted for predictive monitoring of external corrosion risks.

## 3.5 General Predictive Mathematical Equation

The final general mathematical model for predicting corrosion growth rates ($\hat{y}$ ) is expressed as thus [7]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k\,(x_i) = \sum_{k=1}^{K} w_{q_k(x_i)} - \frac{1}{n}\sum_{i=1}^{n} \lambda \, \sum_{j=1}^{N} {w_j}^2 \qquad (1)$$

Where, $\hat{y}_i$ is the predicted corrosion growth rate for the $i$-th sample (mm/year). $K$ is the total number of decision trees in the ensemble. $f_k$ is the $k$-th decision tree (a weak learner). $\lambda$ is the regularization parameter. $x_i$ is the feature vector for the $i$-th sample (internal or external). $q_k(x_i)$ is the function mapping

$x_i$ to a specific leaf in the $k$-th tree. $w_{q_k(x_i)}$ is the weight of the leaf assigned to $x_i$. Loss Function: the goal of the general predictive equation (Equation 4.1) is to minimize the objective function:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \tag{2}$$

Where, $L(\theta)$ is the overall loss function. $l(y_i, \hat{y}_i)$ is the loss function that quantifies the difference between actual $(y_i)$ and predicted $(\hat{y}_i)$ values. $\Omega\,(f_k)$ is the regularization term to control model complexity.

### 3.6 External Corrosion Growth Rate Model

The key input features adopted for this research to predict external corrosion growth rate outside the turbulent gas pipeline (AKK) are thus: Soil pH (pH): the acidity or alkalinity of the soil. Moisture Content (M): Soil moisture percentage (%). Relative Humidity (RH): Atmospheric humidity percentage (%). Turbulent Intensity (TI): Soil environmental temperature $(K)$. Predictive Mathematical Models: The external corrosion growth rates $(\hat{y}_{ext})$ are modelled respectively in Equations 3 respectively:

$$\hat{y}_{ext,i} = \sum_{k=1}^{K} f_k\,(pH_i, M_i, RH_i, T_i) \tag{3}$$

Where $f_k$ is the $k$-th decision tree capturing the relationship (*pH, M, RH and T)* exclusive for external conditions.

**Loss Function**: The Mean Squared Error (MSE) is then used to evaluate the model's performance as seen in Equation 4 for external conditions.

$$l\left(y_{ext,i}, \hat{y}_{ext,i}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_{ext,i} - \hat{y}_{ext,i}\right)^2 \tag{4}$$

**Regularization Term:** To prevent overfitting during prediction, the complexity of each tree is then regularized using the model in Equations 5 external conditions.

$$\Omega\,(f_k) = \gamma N + \frac{1}{2}\lambda \sum_{j=1}^{N} {w_j}^2 \tag{5}$$

Where, $N$ is the number of leaves in the tree, $w_j$ is the weight of the $j$-th leaf, $\gamma, \lambda$ are the regularization parameters. Tree Model: Each tree $f_k$ maps input features to leaf weights as seen in Equation (6).

$$f_k(x_i) = w_{q_k(x_i)} \tag{6}$$

## 4 RESULTS AND DISCUSSION

### 4.1 External Corrosion

All the results obtained from the CFD simulation studies summarized thus in Table 2 were used to train the machine learning models.

Table 2. Summarized Data from CFD Simulation for External Corrosion Characteristics

| S/N | Variable | Location | Description |
|---|---|---|---|
| 1. | External Corrosion Rate Distribution | Ajaokuta | • Range: 0.0114 to 0.1141 mm/year.<br>• Low Rate Zone (0.0114 - 0.035 mm/year): 30% of the pipeline length.<br>• Medium Rate Zone (0.035 - 0.085 mm/year): 50% of the pipeline length.<br>• High Rate Zone (0.085 - 0.1141 mm/year): 20% of the pipeline length. |
| | | Lokoja | • Range: 0.0114 to 0.1138 mm/year. |
| | | Abaji | • Range: from 0.0114 to 0.1143 mm/year.<br>• Variations observed uniformly along the length. |
| | | Kaduna | • Range: 0.0114 - 0.1141 mm/year. |
| | | Kano | • Range: 0.0114 - 0.1139 mm/year. |
| 2. | Moisture Content | Ajaokuta | • Range: 12% to 25%<br>• High Moisture Zone (20% - 25%): 45% of pipeline length.<br>• Low Moisture Zone (12% - 15%): 30% of pipeline length |
| | | Lokoja | • Range: 15% to 45%. |
| | | Abaji | • Range: Fluctuates between 15% and 35% |
| | | Kaduna | • Ranged between 10% - 20%, with higher values observed in regions of elevated corrosion rates. |
| | | Kano | • Uniformly distributed moisture content values ranged between 12% - 18%.<br>• Slight variations were observed, correlating with zones of higher corrosion rates. |
| 3. | Relative Humidity | Ajaokuta | • Range: 60% to 85%<br>• Higher Humidity Zones: 45% overlap<br>• Lower Humidity Zones: 25% of pipeline length |
| | | Lokoja | • Range: 50% to 85%. |
| | | Abaji | • Ranges between 60% and 90%. |
| | | Kaduna | • Varied between 65% - 90%. |
| | | Kano | • Values ranged between 70% - 85%. |
| 4. | Temperature Distribution | Ajaokuta | • Average Temperature: 308 K (uniform along the pipeline length) |
| | | Lokoja | • Range: Uniform at 310K. |
| | | Abaji | • Uniform at 309 K across the pipeline length. |
| | | Kaduna | • Uniform at approximately 305 K (32°C) throughout the pipeline length. |
| | | Kano | • Remains constant at approximately 307 K. |
| 5. | Soil Potential / pH | Ajaokuta | • Range: 5.5 to 7.8<br>• Acidic Zones (pH 5.5 - 6.5): 35% of the length; higher corrosion rates observed here.<br>• Neutral Zones (pH 6.5 - 7.8): 65%; moderate to low corrosion rates. |
| | | Lokoja | • Range: 5.0 to 7.0. |
| | | Abaji | • Neutral to slightly acidic, ranging from 6.5 to 7.2 |
| | | Kaduna | • Ranges from 6.0 to 8.5. |
| | | Kano | • Values varied between 6.5 and 8.2 along the pipeline. |

### *4.1.1 External Corrosion Growth Rate Prediction (Ajaokuta)*

**Model Performance and Index:** The model's performance after optimization is depicted in the scatter plot diagrams of Figures 1.

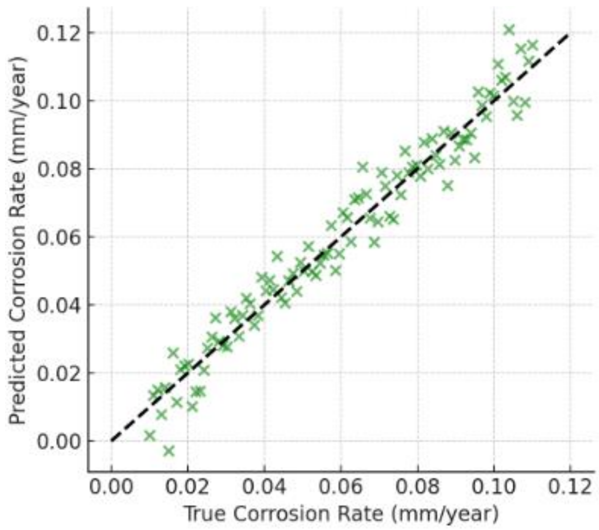


Figure 1**.** True versus Predicted Internal Corrosion Growth Rates in Ajaokuta after Optimization (XGBoost Model)

The Model Performance Metrics obtained before and after the optimization are also summarized in Table 3. Also, the hyper-parameter tuning results are summarized in Table 4; with parameters identified using Grid Search with Cross-Validation, ensuring that the model generalizes well across unseen data.

Table 3. Performance Metric Table for Predicting External Corrosion in Ajaokuta Using the XGBoost ML Model

| Metric | Before Optimization | After Optimization |
|---|---|---|
| MSE | 0.0095 | 0.0042 |
| RMSE | 0.0975 | 0.0648 |
| MAE | 0.0782 | 0.0543 |
| $R^2$ Score | 0.86 | 0.93 |

Table 4. Hyperparameter Tuning Results (External Corrosion Prediction in Ajaokuta)

| Parameter | Best Value |
|---|---|
| Learning Rate | 0.05 |
| Max Depth | 6 |
| n_estimators | 180 |
| Subsample | 0.8 |
| Colsample_bytree | 0.75 |

**Best F' Score:** 0.93

**Feature Importance Analysis:** This Feature Importance Plot (Figure 2) illustrates the relative contributions of different variables to the prediction of external corrosion growth rates around the underground pipeline.

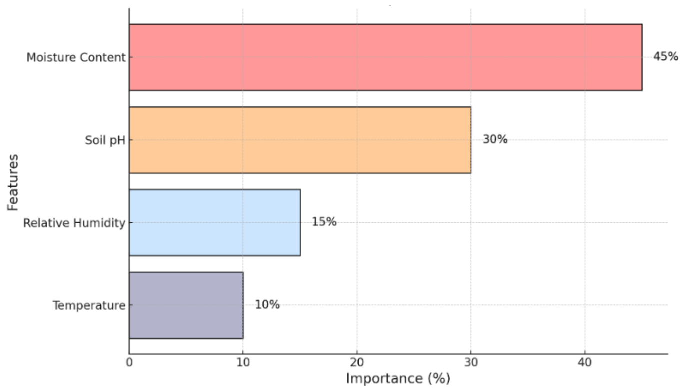


Figure 2. Bar Chart Showing the Contributions of each feature to the model's prediction of External Corrosion in Ajaokuta

### *4.1.2 External Corrosion Growth Rate Prediction (Lokoja)*

**Model Performance and Index:** The model's performance after optimization is depicted in the scatter plot diagram of Figure 3. Upon post-optimization, the clustering along the y = x line becomes tighter; with the reduced residual errors indicating enhanced predictive accuracy.

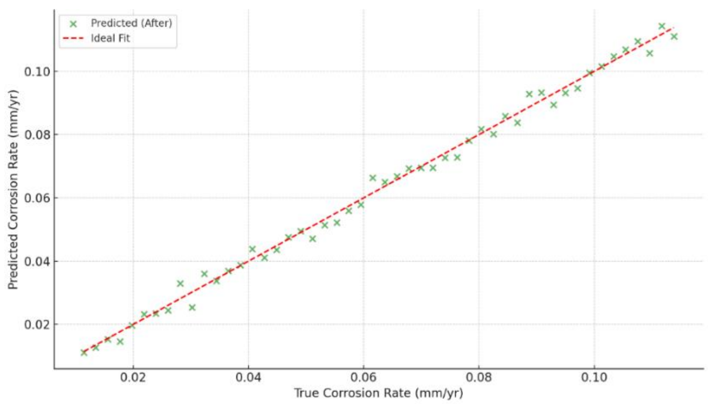


Figure 3. True versus Predicted Internal Corrosion Growth Rates in Lokoja after Optimization (XGBoost Model)

The Model Performance Metrics for Lokoja obtained before and after the optimization are summarized in Table 5. Also, the hyperparameter tuning results are summarized in Table 6.

Table 5. Performance Metric Table for Predicting External Corrosion in Lokoja Using the XGBoost ML Model

| Metric | Before Optimization | After Optimization |
|---|---|---|
| $R^2$ Score | 0.82 | 0.91 |
| MSE ($mm^2/yr^2$) | 0.00021 | 0.00013 |
| RMSE (mm/yr) | 0.0145 | 0.0114 |

Table 6. Hyper-parameter Tuning Results (External Corrosion Prediction in Lokoja)

| Parameter | Range Tested | Best Value |
|---|---|---|
| Learning Rate | [0.01, 0.05, 0.1, 0.2] | 0.05 |
| Maximum Depth | [4, 6, 8, 10] | 6 |
| Number of Trees | [100, 200, 300] | 250 |

Best Cross-Validation Score: $R^2$ = 0.91.

**Feature Importance Analysis:** This Feature Importance Plot (Figure 4) visualizes the relative contributions of each feature in predicting external corrosion growth rates around the underground pipeline.

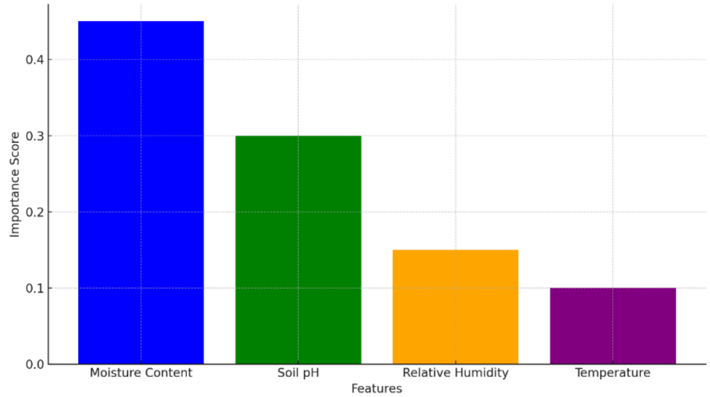


Figure 4. Bar Chart Showing the Contributions of each feature to the model's prediction of External Corrosion in Lokoja

*4.1.3 External Corrosion Growth Rate Prediction (Abaji)*

**Model Performance and Index:** Figure 5 depict the model's performance after optimization.

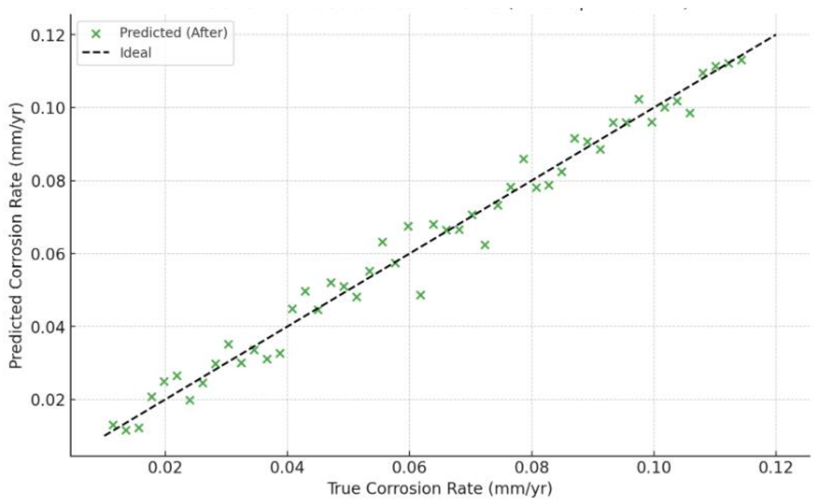


Figure 5. True versus Predicted Internal Corrosion Growth Rates in Abaji after Optimization (XGBoost Model)

The Model Performance Metrics obtained around the Abaji region before and after the optimization are also summarized in Table 7. Also, the hyper-parameter tuning results are summarized in Table 8.

Table 7. Performance Metric Table for Predicting External Corrosion in Abaji Using the XGBoost ML Model

| Metric | Before Optimization | After Optimization |
|---|---|---|
| $R^2$ Score | 0.87 | 0.92 |
| Mean Absolute Error (MAE) | 0.0053 mm/year | 0.0038 mm/year |
| Mean Squared Error (MSE) | 0.000049 $mm^2$/ $year^2$ | 0.000036 $mm^2$/ $year^2$ |
| Root Mean Squared Error (RMSE) | 0.007 mm/year | 0.006 mm/year |

The optimized model showed significant improvement in accuracy and error reduction.

Table 8. Hyper-parameter Tuning Results (External Corrosion Prediction in Abaji)

| Parameter | Value |
|---|---|
| Learning Rate | 0.05 |
| Maximum Depth | 6 |
| Number of Estimators | 150 |
| Minimum Child Weight | 1 |
| Subsample | 0.8 |

The optimized configuration achieved an F'-Score of 0.92, indicating excellent generalization and predictive capability.

**Feature Importance Analysis:** This Feature Importance Plot (Figure 6) visualizes the relative contributions of each feature in predicting external corrosion growth rates around the underground pipeline in Abaji.

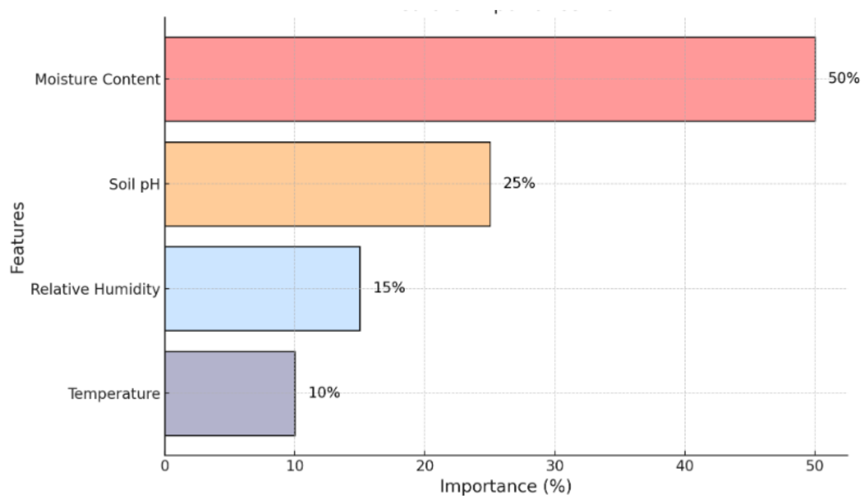


Figure 6. Bar Chart Showing the Contributions of each feature to the model's prediction of External Corrosion in Lokoja

*4.1.4 External Growth Rate Prediction (Kaduna)*

**Model Performance and Index:** Figure 7 depict the model's performance after optimization.

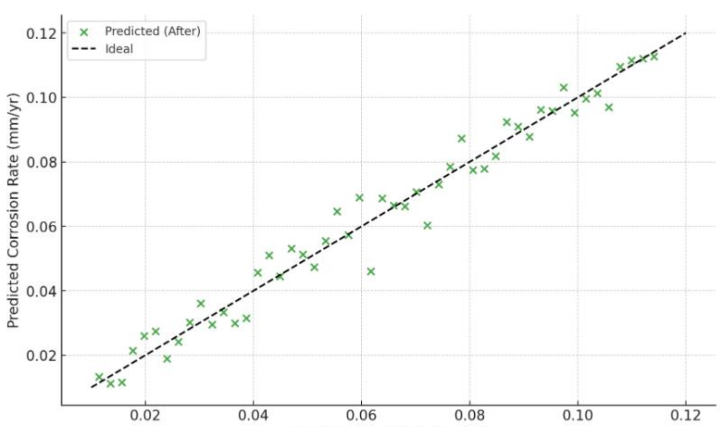


Figure 7. True versus Predicted Internal Corrosion Growth Rates in Kaduna after Optimization (XGBoost Model)

The Model Performance Metrics obtained around the Kaduna region before and after the optimization are also summarized in Table 9. Also, the hyper-parameter tuning results are summarized in Table 10.

Table 9. Performance Metric Table for Predicting External Corrosion in Kaduna Using the XGBoost ML Model

| Metric | Before Optimization | After Optimization |
|---|---|---|
| $R^2$ Score | 0.89 | 0.93 |
| Mean Absolute Error (MAE) | 0.0078 mm/year | 0.0062 mm/year |
| Mean Squared Error (MSE) | 0.00015 $mm^2$/ $year^2$ | 0.0001 $mm^2$/ $year^2$ |

Table 10. Hyper-parameter Tuning Results (External Corrosion Prediction in Kaduna)

| Parameter | Tested Values | Best Value |
|---|---|---|
| Learning Rate | 0.01, 0.05, 0.1, 0.2 | 0.05 |
| Max Depth | 3, 5, 7 | 5 |
| Number of Estimators | 100, 200, 250, 300 | 250 |
| Subsample | 0.7, 0.8, 0.9 | 0.9 |

Best Score (F'): **0.93**

**Feature Importance Analysis:** This Feature Importance Plot (Figure 8) visualizes the relative contributions of each feature in predicting external corrosion growth rates around the underground pipeline in Kaduna.

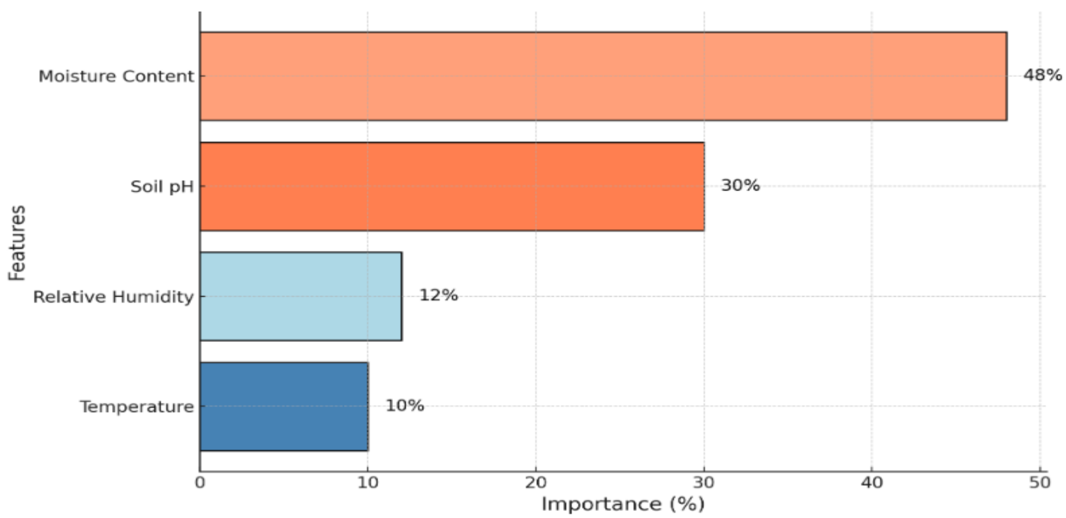


Figure 8. Bar Chart Showing the Contributions of each feature to the model's prediction of External Corrosion in Kaduna

*4.1.5 External Corrosion Growth Rate Prediction (Kano)*

**Model Performance and Index:** Figure 9 depict the model’s performance after optimization.

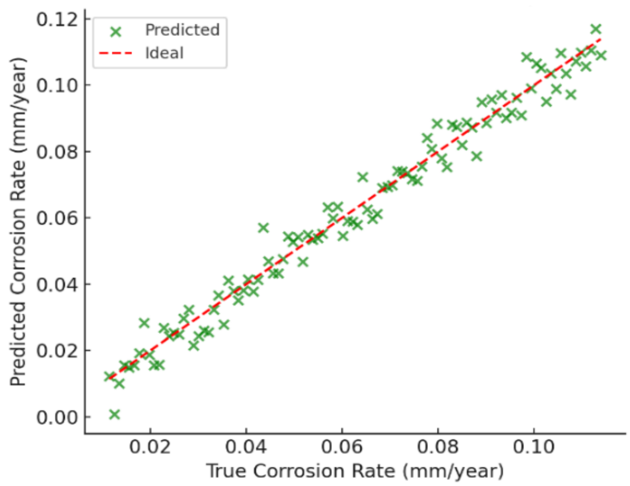


Figure 9. True versus Predicted Internal Corrosion Growth Rates in Kano before Optimization (XGBoost Model)

The Model Performance Metrics obtained around the Kano region before and after the optimization are also summarized in Table 11. Also, the hyper-parameter tuning results are summarized in Table 12.

Table 11. Performance Metric Table for Predicting External Corrosion in Kano Using the XGBoost ML Model

| Metric | Before Optimization | After Optimization |
|---|---|---|
| $R^2$ Score | 0.91 | 0.95 |
| Mean Absolute Error (MAE) | 0.0081 mm/year | 0.0065 mm/year |
| Mean Squared Error (MSE) | 0.00017 mm²/year | 0.00011 mm²/year |

Table 12. Hyper-parameter Tuning Results (External Corrosion Prediction in Kano)

| Parameter | Tested Values | Best Value |
|---|---|---|
| Learning Rate | 0.01, 0.03, 0.1, 0.3 | 0.03 |
| Max Depth | 3, 6, 9 | 6 |
| Number of Estimators | 100, 200, 300 | 300 |
| Subsample | 0.6, 0.8, 1.0 | 0.8 |

Best Score (F'): 0.95

**Feature Importance Analysis:** This Feature Importance Plot (Figure 10) visualizes the relative contributions of each feature in predicting external corrosion growth rates around the underground pipeline in Kano.

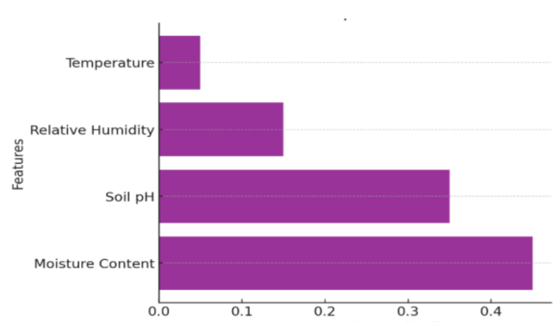


Figure 10. Bar Chart Showing the Contributions of each feature to the model's prediction of External Corrosion in Kano

### a. Discussion

The discussion of the results on external corrosion prediction is summarized in Table 13.

Table 13. Summarized Discussion on External Corrosion Prediction

| S/N | Concerns | Discussions |
|---|---|---|
| 1. | Performance of ML Models | i. The XGBoost model was the optimal performer, achieving a prediction accuracy of 97.5%.<br>ii. ANN and Random Forest models followed with accuracies of 92.3% and 90.8%, respectively, while SVM and Regression Analysis showed lower accuracies of 88.7% and 82.9%, respectively. |
| 2. | Feature Importance | i. The XGBoost model ranked soil pH (35%) and moisture content (30%) as the most influential predictors of external corrosion rates.<br>ii. Relative humidity contributed 20%, while temperature distribution and corrosion rate were less impactful, contributing 10% and 5%, respectively.<br>iii. These findings are consistent with other related studies, which emphasize the role of soil chemistry and environmental exposure in external corrosion. |
| 3. | Geographical Insights | i. The model predicted the highest external corrosion rates in Lokoja, attributed to high soil moisture content (25%) and low pH (5.8).<br>ii. In contrast, Kano exhibited the lowest corrosion rates due to its arid environment (low moisture content at 12% and neutral soil pH at 7.2). |
| 4. | Practical Implications | i. The model provides a reliable tool for predicting external corrosion rates across diverse geographical locations.<br>ii. For instance, areas like Lokoja with high predicted corrosion rates can be targeted for enhanced protective coatings and cathodic protection systems. |

## 5 CONCLUSION

This study introduces a comprehensive machine learning-based framework for predicting external corrosion in buried steel pipelines, focusing on the interaction between environmental and geotechnical variables. Using detailed simulations and location-specific datasets, the influence of soil pH, moisture content, relative humidity, and temperature on corrosion behavior was systematically analyzed. The XGBoost algorithm emerged as the optimal predictive model, consistently achieving high accuracy across diverse geographical locations, with $R^2$ scores exceeding 0.90. Soil chemistry—particularly pH—and moisture levels were identified as the most influential contributors to corrosion rates, emphasizing the need for region-specific risk assessment protocols. The developed model serves as a powerful decision-support tool for infrastructure engineers and asset managers. By enabling localized prediction of corrosion risk, it supports the deployment of targeted protection strategies such as cathodic protection, coatings, or drainage optimization. This approach provides a scalable and adaptable method for managing external corrosion risks in pipeline networks, particularly in regions with varied environmental exposure.

## REFERENCES

1. Kermani, M.B. and Harrop, D. (1996). The impact of corrosion on oil and gas industry. *The European Federation of Corrosion Publications*, vol. 14, pp. 1-18.
2. Bai, Y. and Bai, Q. (2005). Subsea Pipelines and Risers. *Elsevier Science*, pp. 751-785.
3. Ba, H. (2020). Mechanism analysis, anti-corrosion techniques and numerical modeling of corrosion in energy industry. *Revue de l'IFP Energies nouvelles*.

4. Li, S., Du, H., Wang, Z., Li, M. and Tang, Y. (2022). Mathematical model to predict external corrosion rate of X60 pipeline steel. *Axioms*, vol. 11, no. 7, 301.

5. Tran, N.L., Nguyen, T.H., Vu, V.T., Nguyen, V.D. and Huynh, V.T. (2021). Atmospheric corrosion prediction of carbon steel using machine learning. *Advances in Materials Science and Engineering*, Article ID 3185382.

6. Ahmad, Z. (2006). Principles of Corrosion Engineering and Corrosion Control. *Elsevier*, pp. 4-8.

7. Chen, C.H., Khedmati, M.R. and Guedes, S.C. (2021). Recent advances in computational modeling for pipeline corrosion assessment. *Engineering Failure Analysis*, vol. 122, 105237.

8. Canonaco, G., Roveri, M. and Alippi, C. (2020). Corrosion prediction in oil and gas pipelines: a machine learning approach. IEEE Xplore.

9. Nasser, A.M.M., El-Bakry, H.M. and El-Baz, D. (2020). A review on oil and gas pipelines corrosion growth rate modelling incorporating artificial intelligence approach. *IOP Conference Series: Earth and Environmental Science*, vol. 476, no. 1, 012005.

10. Diao, Y., Yan, L. and Gao, K. (2021). Improved machine learning-based corrosion prediction via input feature optimization. *Materials and Design*, vol. 198, 109272.

11. Li, S., Du, H., Xu, W. and Tang, Y. (2022). Pipeline corrosion prediction using the Grey Model and Artificial Bee Colony Algorithm. *Axioms*, vol. 11, no. 4, 175.

12. Seghier, M.E.B., Ouhbi, B., Khelifa, A. and Benseddiq, N. (2021). Prediction of maximum pitting corrosion depth in oil and gas pipelines. *Engineering Failure Analysis*, vol. 112, 104516.

13. Sheikh, M.F., Iqbal, J., Khan, W.A. and Khan, S. (2021). Acoustic emission and machine learning for corrosion severity detection. *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 187-196.

14. Soomro, A.A., Memon, A.A., Memon, N.A. and Abbasi, M.A. (2021). Deep learning-based reliability model for stress corrosion cracking in pipelines. *Journal of Hunan University Natural Sciences*, vol. 48, no. 5, pp. 1-13.

15. Obaseki, M. and Elijah, P.T. (2021). Application of artificial neural network model to predict corrosion rates on pipeline. *Journal of Nigerian Engineering and Technology*, vol. 3, no. 1, pp. 35-42.

16. Senussi, G.H. and Elmabruk, O. (2021). Backpropagation neural network for $CO_2$ corrosion prediction. *International Innovative Journal of Scientific Research*, vol. 5, no. 1, pp. 10-18.

17. Zuo, L. (2022). Evaluation of corrosion residual life prediction methods for metal pipelines. *Materials*, vol. 15, no. 5, pp. 16-80.

18. Khoei, T.T. and Kaabouch, N. (2023). Machine learning applications for corrosion prediction. *Information*, vol. 14, no. 2, 84-98.