Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

NEXT-GEN CORROSION ANALYTICS: INTELLIGENT FORECASTING FOR OIL PIPELINE INTEGRITY

Abhishek Mohanty [a], Rabi Shankar Panda [b,*], Chinmay Pradhan [c], Susmita Sahoo [d], Soumya Mishra [e] and Sanjeet Kumar Subudhi [f]

[a, b*, c, d] *Department of Computer Science & Engineering, C.V. Raman Global University, Bhubaneswar, 752054, India.*
[e] *Department of Electronics & Communication Engineering, C.V. Raman Global University, Bhubaneswar, 752054, India.*
[f] *Department of Electrical Engineering, C.V. Raman Global University, Bhubaneswar, 752054, India.*

ABSTRACT

Effective monitoring and precise forecasting of internal CO_2 corrosion are paramount for safeguarding the structural integrity of oil pipelines. This research employs an array of advanced machine learning and deep learning methodologies to predict corrosion progression, leveraging a comprehensive dataset encapsulating critical environmental and operational factors. Various regression techniques, including Support Vector Regression (SVR), Decision Trees, Random Forest, K-Nearest Neighbours (KNN), and Polynomial Regression, were meticulously evaluated. Among linear models, the Random Forest Regressor exhibited superior predictive accuracy with an R^2 of 0.9936 and an MSE of 0.0080. Notably, third-degree polynomial regression emerged as the most effective nonlinear approach (R^2 = 0.9988). Regularized regression models underscored Ridge Regression as the optimal choice, achieving an R^2 of 0.9989. Within deep learning paradigms, the Feed-Forward Neural Network (FNN) and ResNet models demonstrated remarkable efficacy, attaining a R^2 score of 0.9928. Furthermore, ensemble learning, particularly a hybridized CNN-CNN Forecaster model, exhibited commendable predictive precision. The study emphasizes the necessity of model selection for accurate corrosion forecasting and facilitating proactive maintenance.

Keywords: *AI-driven corrosion prediction, Machine Learning, Deep Learning, CO_2 corrosion forecasting, Ridge Regression, Feed-Forward Neural Network (FNN)*

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

1. INTRODUCTION

An inevitable process, corrosion happens when things, notably metals, gradually erode as a result of electrochemical or chemical interactions with their environment [1]. It is a process that impacts a variety of materials and constructions; additionally, corrosion may affect a material's solidity and functionality in addition to changing its appearance [2].

According to [3], 37 manufacturers transported 216 million m³ (1.4 billion barrels) of oil through 21,636 km of pipelines. Meanwhile, 67 companies transported 152 billion m³ (5.4 trillion cubic feet) of natural gas via 55,982 km of pipelines, with 11 companies handling both oil and gas. A 2016 NACE estimate pegged the cost of pipeline steel corrosion in the oil and gas industry at USD 2.5 trillion, or 3.4% of global GDP, excluding safety and environmental expenses, highlighting its critical impact. [4].

The lifespan of equipment is shortened by corrosion, which results in prohibitive upkeep, unexpected shutdowns, and financial losses. To guarantee productivity and cost-effectiveness, preventive maintenance and continuous evaluation are crucial [5]. Non-destructive testing (NDT) of structures is increasingly utilizing imaging techniques, which improve visual inspections and conserve money and time. More efficient image-analysis techniques must be created in order to modify image data into an arrangement that can be used for damage assessment [6].

However, the failure impacts on the nearby pipeline were not explained by the analysis for that specific pipeline failure situation. Therefore, each pipeline must be analysed independently [7]. Since practical manufacturing is rare except in hazardous conditions, detecting corrosion before leaks occur is essential [8]. Since pipelines are the primary means of transferring crude and product oils, accurate identification and prediction of pipeline leaks is crucial to the pipeline's safe functioning.

1.1. Background

Internet of Things (IoT) and Machine Learning (ML) based corrosion identification and prediction system which makes use of sensors involving pH, thickness, and GPS. It utilizes a classifier based on Q-learning and semi-supervised learning for validation [9]. However, it needs Deep Learning (DL) implementation for improved performance, and its accuracy is dependent on the quality of the sensor information.

Imran et al., examines AI applications in steel corrosion prediction and detection, focusing developments in ML and DL. Although AI enhances accuracy and effectiveness, it has drawbacks, such as low-quality information and difficult-to-understand models [10]. Future studies should improve collecting information, create interpretable models, and verify AI in practical settings.

Data is transferred to the cloud for real-time analysis utilizing IoT-based sensor technologies to monitor sustainability metrics and identify corrosion in reinforced concrete structures [11]. The strategy depends on sensor precision, reliability of data transfer, and integration difficulties in extensive infrastructure.

A cathodic protection (CP) system helps minimize corrosion in steel buildings that are surrounded by air. It delivers constant protection by using a fiber sheet and sacrificial anode [12]. However, environmental conditions and the presence of residual corrosion affect how effective it is.

2. METHODOLOGY

This section presents an extensive account of the research methodology used, encompassing both sophisticated DL frameworks for corrosion rate prediction and classic ML regression techniques. There were two main stages to the experimentation:

1. Machine Learning-Based Regression Analysis
2. Deep Learning-Based Corrosion Rate Prediction

The overall workflow of the methodology is depicted in Fig.1. To identify the best predictive model, each strategy was executed into practice, refined, and contrasted using legitimate evaluation metrics. To identify the best predictive model, each strategy was executed into practice, refined, and contrasted using legitimate evaluation metrics.

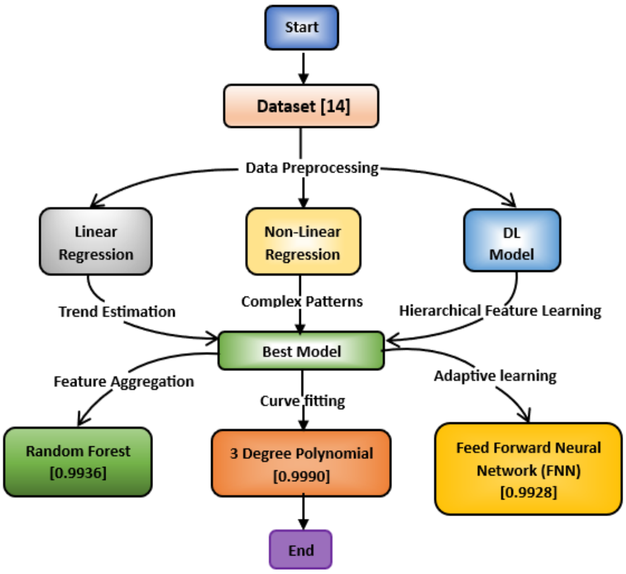


Fig.1. Topology of the proposed prediction methodology.

*Corresponding author email: rabishankarpanda6@gmail.com

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025

ISSN: 1355-5243

SCOPUS

2.1. Dataset Description

The dataset includes oil pipeline corrosion rates under extreme conditions like high temperatures, acidic fluids, and limited inhibitor efficiency. Generated using the NORSOK M506 model with Monte Carlo simulation, it incorporates factors such as temperature, flow velocity, CO_2 pressure, internal pressure, inhibitor efficiency, shear stress, pH, and corrosion rate [13]. Due to limited experimental data, this synthetic dataset aids in predictive modelling. Feature relationships were analysed using a correlation heatmap (Fig.2) to identify dependencies before model development.

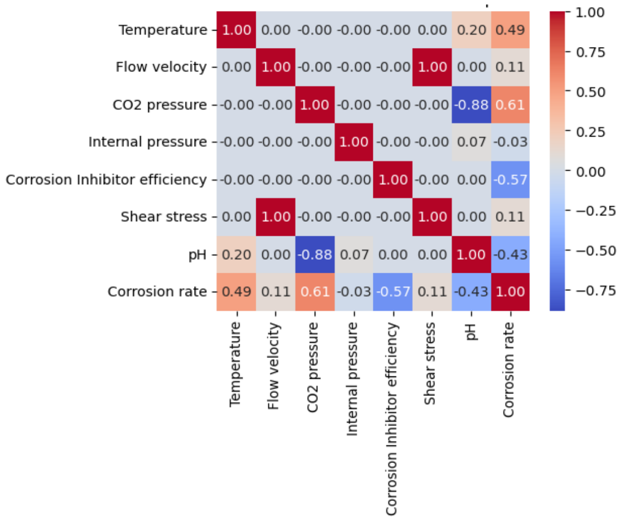


Fig.2. Feature Correlation Heatmap.

2.2. Machine Learning-Based Regression Analysis

2.2.1. Linear Regression (1st Degree)

To assess the core linkages between corrosion rate and pipeline characteristics (like material composition, exposure factors, and environmental conditions), Linear Regression has been utilized as a baseline model [10]. The model conforms to the following formula:

$$y = \beta_0 + \beta_1 x + \epsilon \qquad (1)$$

where, y is the corrosion rate, x is the independent variable (pipeline characteristic), and $\beta 0$ and $\beta 1$ are the regression coefficients, ε is the residual error term. The Least Squares Method has been used for training the model so as to reduce residual errors, assure ideal parameter estimation.

2.2.2 Non-Linear Regression Models (2nd, 3rd, and 4th Degree)

Nonlinear regression models of degree 2, 3, and 4 has been established so as to incorporate any possible non-linear correlations between the corrosion rate and impacting factors [14]. The generic equation for polynomials is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \epsilon \qquad (2)$$

where, intricate relationships between variables are conveyed by higher-degree terms. To implement this, Scikit-learn's Polynomial Features module has been applied for feature engineering. Overfitting concerns were mitigated by employing 5-Fold Cross-Validation (CV), particularly for fourth-degree polynomials. Model performance was compared using Mean Squared Error (MSE) and Mean Absolute Error (MAE).

2.2.3 Ridge and Lasso Regression

Regularized Regression Techniques has been applied to prevent overfitting in polynomial regression of higher-degree models [15]:

i. **Ridge Regression:** Prevents high coefficient values via L2 penalty.

$$\min_\beta \sum (y_i - \hat{y}_i)^2 + \lambda \sum \beta_j^2 \qquad (3)$$

ii. **Lasso Regression:** By using an L1 penalty to reduce unnecessary coefficients to zero, Lasso Regression enables feature selection viable.

$$\text{smin}_\beta \sum (y_i - \hat{y}_i)^2 + \lambda \sum |\beta_j| \qquad (4)$$

Grid Search CV has been used to optimize the regularization parameter (λ), and model generalization performance has been compared against conventional polynomial regression.

2.3 Deep Learning-Based Corrosion Rate Prediction

2.3.1 Feedforward Neural Network (FNN)

A FNN has been implemented to model complex, nonlinear interactions within the dataset [10]. The architecture consists of an input layer incorporating relevant pipeline characteristics, three fully connected hidden layers with ReLU activation for non-linearity, and an output layer with a linear activation function to predict corrosion rates. Training has been conducted using the Adam optimizer minimizing MSE loss function with a batch size of 16, 1000 epochs, and a 0.2 dropout rate to mitigate overfitting.

2.3.2 Residual Neural Network (ResNet) Model

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
ISSN: 1355-5243
SCOPUS

Given the advantages of deep residual learning, a ResNet-based architecture has been adopted to address the vanishing gradient problem. ResNet introduces skip connections, allowing information to bypass certain layers and improve learning efficiency [16]. The residual function is represented as:

$$y = x + F(x, W) \tag{5}$$

where, W represents transformation's weights, x is the input, and F (x, W) is the residual function. The architecture incorporates multiple residual blocks with batch normalization, ReLU activation, and has been trained using the Adam optimizer.

2.3.3 CNN-CNN Forecaster Ensemble Model

To improve prediction accuracy, an ensemble approach is proposed, combining several Convolutional Neural Network (CNN) models [10].

i. **CNN-Based Feature Extraction** – Multiple CNN architectures have been trained to capture hierarchical corrosion-related features.

ii. **Ensemble Learning** – The outputs from different CNN models has been aggregated using a weighted averaging technique to produce the final prediction.

2.3.4 Generative Adversarial Networks (GAN)

For forecasting corrosion rates and creating synthetic data, a GAN-based architecture has been used [17]. Two model consists of:

- Generator (G): Produces artificial corrosion rate information.
- Discriminator (D): Distinguishes between synthetic and genuine data.

During training, the generator attempts to fool the discriminator, while the discriminator learns to distinguish between genuine and synthetic data. The model has been optimized using Binary Cross-Entropy loss and Adam optimizer, with 1000 training epochs ensuring realistic data generation.

3. RESULTS AND ANALYSIS

The analysis of the model's efficiency is a crucial aspect of this research, with an emphasis on important regression metrics like MAE, MSE, Root Mean Squared Error (RMSE), R-Squared (R2). The objective is to ascertain the model that is more effective in providing precise predictions. The research compares sophisticated DL models, such as ensemble techniques and Generative Adversarial Networks (GANs), with conventional ML regression models.
It's crucial to comprehend the target variable's distribution prior to analysing model-specific outcomes. The distribution of corrosion rates throughout the dataset is shown in Fig. 3, which sets the tone for the subsequent model analysis.

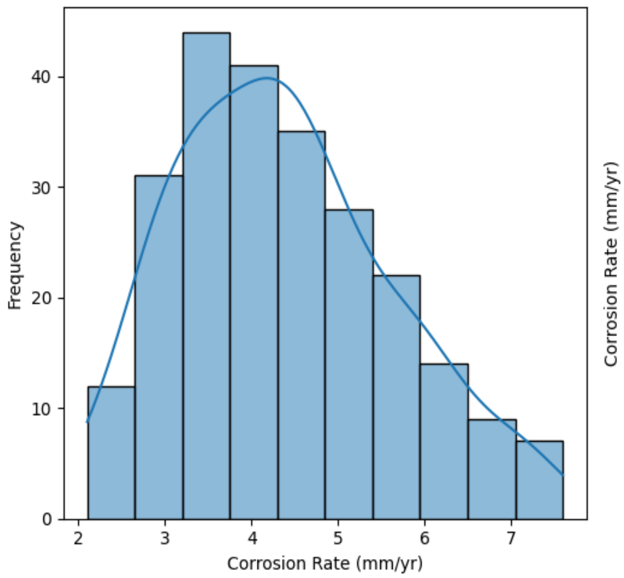
Model	**MAE**	**MSE**	**RMSE**	**R^2 Score**
SVR	0.1212	0.0275	0.1658	0.9776
Decision Tree	0.0755	0.0120	0.1097	0.9902
RandomForest	0.0737	0.0078	**0.0885**	**0.9936**
KNN	0.2955	0.1388	0.3725	0.8869

Fig. 3. Distribution of Corrosion Rates.

3.1 Machine Learning Regression Models (Linear & Non-Linear)

3.1.1 Linear Regression Models

SVR, Decision Tree, Random Forest, and K-Nearest Neighbours (KNN) has been used to evaluate the



performance of conventional regression models. Table 1 provides a summary of the findings.

Table 1: Performance Comparison of Conventional Regression Models

with the lowest RMSE of 0.0885 and the highest R2 score of 0.9936, the Random Forest model surpassed others, indicating that it can effectively manage non-linearity.

Corrosion Management
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
ISSN: 1355-5243
SCOPUS

3.1.2 Non-Linear Regression Models

As shown in Table 2, polynomial regression models with degrees 1 to 4 were evaluated. Degree 3 offered the best match.

Table 2. Evaluation of Polynomial Regression Models Across Different Degrees

Polynomial Degree	Train R^2	Test R^2	Train MSE	Test MSE	CrossValidation R^2
1	0.9621	0.9323	0.0633	0.0830	0.9571 ± 0.0099
2	0.9988	0.9975	0.0020	0.0031	0.9982 ± 0.0004
3	**0.9996**	**0.9988**	**0.0007**	**0.0014**	**0.9990 ± 0.0004**
4	0.9998	0.9977	0.0004	0.0028	0.9954 ± 0.0012

With the lowest test MSE of 0.0014 and a test R2 of 0.9988, polynomial degree 3 performed best. The Lasso and Ridge regression models were also used; Ridge regression had the best R2 score, at 0.9989.

3.2 Deep Learning Model Performance

DL models, such as a CNN ensemble, FNN, ResNet, and GAN has been evaluated, in terms of predicted accuracy. The ResNet model fared noticeably better than the other models. A thorough summary of each model's obtained performance measures is given in Table 3.

Table 3. Performance Metrics of DL Models for Corrosion Rate Prediction

Model	Metric	Value
Feed Forward Neural Network (FNN)	R^2	0.9928
	MSE	0.0088
	RMSE	0.0940
ResNet	Test MAE	0.0461
CNN Ensemble	Ensemble MAE	10.9580
GAN	Test MAPE	89.37%
	Test MAE	0.4820

4. CONCLUSIONS

This study analysed a variety of modelling techniques for predictive analysis, such as DL models, regularized regression, non-linear regression, and linear regression. With an R2 of 0.9936, the Random Forest Regressor was determined to be the top-performing linear regression model after extensive testing. A third-degree model in polynomial regression exhibited exceptional accuracy, with R2 values above 0.998 in every assessment. With an R2 of 0.9989, Ridge Regression showed the best generalization among regularized models. In DL, the FNN demonstrated great predictive performance (R2 = 0.9928).
The results show that DL architectures like ResNet perform better in difficult circumstances, achieving a Test MAE of 0.0461. while classic ML models still produce reliable answers. This study emphasizes how crucial it is to determine models according to the traits of the data and the specifications of the application. Future research can focus on testing different degrees and types of corrosion on well-performing models to assess their robustness and generalizability across diverse real-world corrosion scenarios.

5. REFERENCES

[1] Alamri, A. H. (2020). Localized corrosion and mitigation approach of steel materials used in oil and gas pipelines – An overview. Engineering Failure Analysis, 116, Article 104735. http://dx.doi.org/10.1016/j.engfailanal.2020.104735.
[2] Pourazizi, R., Mohtadi-Bonab, M., & Szpunar, J. (2020). Investigation of different failure modes in oil and natural gas pipeline steels. Engineering Failure Analysis, 109, Article 104400. http://dx.doi.org/10.1016/j.engfailanal.2020.104400.
[3] P. Becklumb and M. Zakzouk, Bill C-46: An Act to Amend the National Energy Board Act and the Canada Oil and Gas Operations Act, 2015.
[4] Fondevik, S. K., Stahl, A., Transeth, A. A., & Knudsen, O. O. (2020). Image segmentation of corrosion damages in industrial inspections. In 2020 IEEE 32nd international conference on tools with artificial intelligence (pp. 787–792). Baltimore, MD, USA: http://dx.doi.org/10.1109/ICTAI50040.2020.00125.
[5] Vasagar, V., Hassan, M. K., Abdullah, A. M., Karre, A. V., Chen, B., Kim, K., Al-Qahtani, N., & Cai, T. (2024). Non-destructive techniques for corrosion detection: A review. *Corrosion Engineering Science and Technology*, *59*(1), 56–85. https://doi.org/10.1177/1478422x241229621

[6] Kabir, S., Rivard, P., & Ballivy, G. (2008). Neural-network-based damage classification of bridge infrastructure using texture analysis. *Canadian Journal of Civil Engineering*, *35*(3), 258–267. https://doi.org/10.1139/l07-105

[7] N. Abdurahman, Y. Rosli, N. Azhari, and B. Hayder, "Pipeline transportation of viscous crudes as concentrated oil-in-water emulsions," Journal of Petroleum Science and Engineering, vol. 90, pp. 139-144, 2012

[8] A. Mazzoldi, D. Picard, P. G. Sriram, and C. M. Oldenburg, "Simulation-based estimates of safety distances for pipeline transportation of carbon dioxide," Greenhouse Gases: Science and Technology, vol. 3, pp. 66-83, 2013.

[9] Parjane, V. A., Arjariya, T., & Gangwar, M. (2023). Corrosion Detection and Prediction for Underwater pipelines using IoT and Machine Learning Techniques. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(2s), 293–300. https://ijisae.org/index.php/IJISAE/article/view/2626

[10] Imran, Md Mahadi & Jamaludin, Shahrizan & Faisal, Ahmad & Ayob, Mohamad & Chowdhury, Ruhi & Khan, Mohammad & Hasan, Ibnul & Ullah, Aman & Rezaul, M. (2024). Application of Artificial Intelligence in Corrosion Research, Part II: Corrosion of Steel.

[11] Technik und Forschung im Betonbau AG, & Schiegg, Y. (2022). A new IoT Corrosion Monitoring-System for concrete structures. *E-Journal of Nondestructive Testing*, *27*(9). https://doi.org/10.58286/27339

[12] Yang, M., Kainuma, S., Ishihara, S., Kaneko, A., & Yamauchi, T. (2020). Atmospheric corrosion protection method for corroded steel members using sacrificial anode of Al-based alloy. *Construction and Building Materials*, *234*(117405), 117405. https://doi.org/10.1016/j.conbuildmat.2019.117405

[13] Khakzad, N. (2021). *Simulation data for CO2 corrosion rate of oil pipeline* [Data set]. Technische Universiteit Delft.

[14] Varbai, B., Wéber, R., Farkas, B., Danyi, P., Krójer, A., Locskai, R., Bohács, G., & Hős, C. (2024). Application of regression models on the prediction of corrosion degradation of a crude oil distillation unit. *Advances in Materials Science*, *24*(1), 72–85. https://doi.org/10.2478/adms-2024-0005

[15] Wang, H., & Chen, Z. (2023). Assessment and prediction of corrosion rate of marine railway bridges based on ridge regression model. *2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)*, 631–636.

[16] Oyedeji, O. A., Khan, S., & Erkoyuncu, J. A. (2024). Application of CNN for multiple phase corrosion identification and region detection. *Applied Soft Computing*, *164*(112008), 112008. https://doi.org/10.1016/j.asoc.2024.112008

ACKNOWLEDGEMENT

The authors acknowledge 'National Seminar on Corrosion and its Prevention- Oil & Gas Industry' CPOG-2025, C.V. Raman Global University in collaboration with AMPP India Chapter.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

FUNDING

The authors received no financial support for the research, authorship and/or publication of this article.