CAPÍTULO 8

MACHINE LEARNING MODEL FOR PREDICTING CO_2 CORROSION RATE IN BRINES WITH DIVALENT IONS IN THE PRESENCE OF CARBON QUANTUM DOTS (CQDS)

Viviana Ortiz-Pérez
Grupo de Investigación Fenómenos de Superficie Michael-Polanyi, Facultad de Minas, Universidad Nacional de Colombia, Medellín

Lady J. Giraldo
Grupo de Investigación Fenómenos de Superficie Michael-Polanyi, Facultad de Minas, Universidad Nacional de Colombia, Medellín

Camilo A. Franco
Grupo de Investigación Fenómenos de Superficie Michael-Polanyi, Facultad de Minas, Universidad Nacional de Colombia, Medellín

Farid B. Cortés
Grupo de Investigación Fenómenos de Superficie Michael-Polanyi, Facultad de Minas, Universidad Nacional de Colombia, Medellín

ABSTRACT: The use of carbonated water (CW) in carbon capture, utilization, and storage (CCUS) processes is an effective strategy to address security and enviromental sustaninability challenges. However, the presence of minerals, organic acids, inorganic ions such as calcium (Ca^{2+}), and dissolved CO_2 in formation brines enhances electron exchange, accelerating corrosion in pipelines and facilities. This poses a substantial risk to equipment integrity and reduces the effectiveness of CW injection in enhanced oil recovery (EOR) processes and CO_2 storage. To mitigate these effects, environmentally friendly corrosion inhibitors, such as carbon quantum dots (CQDs), have emerged as promising alternatives due to their high dispersibility, low toxicity, and stable chemical properties. Nevertheless, accurately predicting corrosion rates remain a major challenge, as traditional models often fail to capture the complex interactions among multiple factors. In this study, the inhibitory mechanism of CQDs and the corrosive effect of divalent ions (Ca^{2+}) in CO_2-saturated production brines were evaluated through gravimetric measurements. Additionally, four predictive models were developed using machine learning algorithms: Random Forest (RF), K-Nearest Neighbors (KNN), Extreme Grading Boosting (XGBoost) and Support Vector machine Regressor (SVR), to enhance predictive accuracy and operational efficiency

in CO_2 corrosion management. Among these models, SVR demonstrated the best predictive performance, achieving an R^2 of 0.9281 on the test set, indicating strong generalization capabilities. In contrast, XGBoost exhibited overfitting, performing exceptionally well on the training set but showing a noticeable drop in accuracy on the test set. RF and KNN showed moderate predictive capacity but struggled to generalize effectively. These results highlight the need for further refinement of predictive models, including improved feature selection, hyperparameter tuning, and the exploration of additional experimental conditions such as temperature, pH variations, and different CQD concentrations.

KEYWORDS: CO_2 corrosion, Calcium carbonate, Carbon quantum dots, Machine learning.

INTRODUCTION

In the Oil & Gas industry, the presence of carbon dioxide (CO_2) plays a crucial role in enhanced oil recovery (EOR) processes by reducing crude oil viscosity and allowing for greater extraction efficiency, as well as in carbon capture and storage (CCS) processes where it can be injected into depleted reservoirs or saline aquifers for geological storage to mitigate environmental impacts [1-3]. However, when injected or naturally occurring CO_2 dissolves in water, it hydrates to produce carbonic acid (H_2CO_3), changing the acidity of the formation water and promoting electrochemical reactions between the aqueous phase and the steel surface, leading to iron dissolution [4-7]. CO_2 corrosion of carbon steel represents serious problems in the sector by causing pipeline perforation and cracking, which can result in catastrophic accidents and massive economic losses [8]. Despite its susceptibility and low corrosion resistance in CO_2-containing environments, carbon steel is still widely used in oil and gas production and transportation due to its high tensile strength, low cost, and wide availability [9,10].

Corrosion in carbonated brines is a complex process influenced by parameters such as CO_2 content in the gas phase, temperature, pH, and ionic composition of the solution [11, 12]. Formation brine can vary from a simple solution to a more complex mixture containing various types of carbonaceous species and divalent cations dissolved at high concentrations, such as Ca^{2+} (calcium ion), along with the predominant presence of monovalent salts like NaCl [13, 14]. Typically, the corrosion product deposited on the steel surface in an aqueous medium of sodium chloride (NaCl) saturated with CO_2 consists of iron carbonate ($FeCO_3$) scales [5]. The formation of siderite scales acts as a protective surface film, blocking active dissolution sites and creating a barrier to the diffusion of electrochemically active species, significantly reducing the corrosion rate [15]. However, the presence of Ca^{2+} ions can react with

CO_2, facilitating the precipitation of carbonate-type corrosion products ($CaCO_3$ and $Fe_xCa_yCO_3$), altering the protective capacity of $FeCO_3$ layers, and affecting the stability of already precipitated corrosion products [16, 17], making CO_2 corrosion even more complex. Authors such as Ding et al. evaluated the effect of these ions in simulated brine with different Ca^{2+} concentrations, finding that the corrosion rate increases with the rise in calcium ion concentration. Additionally, they attributed that as Ca^{2+} concentration increases, the corrosion scales become looser, and the grain size becomes larger, decreasing the resistance and adhesion strength of the scales to the matrix [18]. Later, Shamsa et al. studied the role of Ca^{2+} ions in Ca/Fe products in CO_2-corrosive environments at temperatures of 80 and 150°C, finding that at 80°C, the addition of Ca^{2+} ions increased the corrosion rate and produced a more porous $Fe_xCa_yCO_3$ layer, whereas at 150°C, the mixed carbonate layer was as protective as pure $FeCO_3$, reducing overall corrosion and minimizing pitting formation [19].

Thus, the use of corrosion inhibitors is crucial in the industry to prolong the operational lifespan of field facilities by mitigating CO_2 corrosion [20]. Currently, numerous CO_2 corrosion inhibitors derived from imidazoline, Schiff bases, and quaternary ammonium salts have been reported [21-23]. However, these organic compounds exhibit high toxicity, poor biodegradability, and complex synthesis in the petrochemical industry, impacting the environment and limiting their application [24]. Therefore, the development of carbon quantum dots (CQDs) is considered one of the most promising candidates as eco-friendly corrosion inhibitors due to their outstanding properties, such as high water solubility, good biocompatibility, low toxicity, and stable chemical properties, making these fluorescent nanomaterials applicable in the field of corrosion. The inclusion of Carbon Quantum Dots in corrosion research has gained traction in recent years. Cen et al. synthesized nitrogen and sulfur co-doped Carbon Dots as corrosion inhibitors in 3.5wt% NaCl solutions saturated with CO_2 [25], finding that the inhibitory effect of CDs is associated with the chemisorption of functional groups on the surface of N, S-CDs with iron, contributing to the formation of the protective film. More recently, authors such as Wu et al. synthesized Carbon Dots from sodium salicylate and urea to evaluate them as effective corrosion inhibitors for N80 steel in 1M HCl and 3.5wt% NaCl solutions saturated with CO_2 [26]. Their results showed that N-CDs gradually adhered to the steel surface through physical and chemical adsorption to form a corrosion-inhibiting film that isolates corrosive ions from contacting the carbon steel.

Over the last decade, with the growth of computing technology and the advancement of artificial intelligence, machine learning has been applied to various corrosion-related problems, such as corrosion detection through automated image analysis [27], prediction of corrosion defect growth in pipelines [28], material

inspection [29], and estimation of corrosion rate in marine environments [30]. Aghaaminiha et al. [31] evaluated supervised machine learning models to predict carbon steel corrosion rates over time in the presence of corrosion inhibitors. Their dataset considered the addition of corrosion inhibitors at different concentrations and dosing schemes. The authors tested four machine learning algorithms: Random Forest (RF), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Support Vector Regression (SVR), finding that the Random Forest model accurately predicted the temporal evolution of corrosion and its sensitivity to environmental conditions such as temperature, pH, CO_2 concentration, and inhibitor type. Additionally, authors like Dong et al. [32] developed a predictive model to estimate CO_2 corrosion rate and severity based on machine learning algorithms. They evaluated six machine learning models: Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting Decision Tree (GBDT), Support Vector Machine (SVM), XGBoost, and LightGBM. Among the most significant results, they found that SVM had the lowest performance in both training and test sets, while Random Forest (RF) achieved the best results, with R^2 values of 0.92 for the training set and 0.88 for the test set.

Despite advances in CO_2 corrosion prediction and the development of efficient inhibitors, no studies have modeled corrosion rates in the simultaneous presence of Ca^{2+} ions and Carbon Quantum Dots (CQDs) as inhibitors. The presence of Ca^{2+} in carbonated brines can significantly alter the stability and effectiveness of corrosion products, modifying the dynamics of protective layer formation and affecting inhibitor performance. However, to date, predictive models have not comprehensively considered these effects, limiting their applicability in real environments. Therefore, machine learning techniques serve as a highly valuable tool by offering nonlinear modeling capabilities, superior predictive performance, and the ability to comprehensively consider multiple influential factors. In this study, the implementation of a machine learning-based model to predict corrosion rates over a wide range of experimental conditions will enhance the understanding of the combined effect of Ca^{2+} ions and CQDs under different salinity scenarios, immersion times, and inhibitor concentrations. This will contribute to the development of more effective strategies for corrosion mitigation in oil and gas production environments, where Ca^{2+} presence in brines is common, and its impact on corrosion product stability has not been fully characterized.

METHODOLOGY

Description of the experimental data

The experimental data used was obtained from a series of tests on the corrosion inhibition of carbon steel in CO_2-saturated aqueous solutions. The corrosion rate was

measured using weight loss tests. Carbon steel samples were immersed in prepared CO_2-saturated solutions containing NaCl and $CaCl_2$ at different concentrations (5000, 10000, and 30000 mg/L), both in the absence and presence of Carbon Quantum Dots (CQDs), at 70°C and under static conditions for different immersion times. At the end of the immersion period, the corrosion products on the steel surface were removed using a cleaning solution (5% HCl), rinsed with distilled water, and dried with compressed air. All samples were weighed before the experiment and after cleaning, ensuring the removal of corrosion products to determine the average corrosion rate. The corrosion rate (CR) was calculated using Equation 1:

$$CR = \frac{87600\ \Delta W}{\rho S t} \quad (1)$$

Where represents the average weight loss (g) of the samples, is the density of the carbon steel samples (7.79 g cm^{-3}), is the exposed area (cm^2), and represents inmersion time (h). The experimental matrix design is presented in **Table 1.**

Description	Range	Unit	Type
Type of Brine	{NaCl, $CaCl_2$}	-	Categorical
Time	[24 – 120]	hour	Numerical
Salinity	[4500 – 30500]	mg L^{-1}	Numerical
Carbon Quantum Dots concentration	[0 – 100]	ppm	Numerical
Temperature	70	**°C**	Numerical
pH	[3.3 – 4]	-	Numerical

Table 1. Operational Input variables (features) considered to model corrosion rate

Source: produced by the author

Overview of machine learning algorithms

Machine learning techniques refers to a a set of algorithms that learn to perform tasks, such as predicting experimental outcomes, by being trained on data from previously conducted experiments. In particular, supervised learning involves training a model using labeled data, where the algorithm learns to map input data to their corresponding labels, enabling accurate predictions based on previously observed patterns [33,34]. In this study, common machine learning algoritms in corrosion such as Supported Vector machines Regression (SVR), Random Forest (RF), K Nearest Neighbors (KNN) and Extreme Gradient Boosting (XGBoost) were applied to the processed data. The data were randomly sampled, with 80% used as the training set and 20% for the testing set to evaluate the model's performance.

Random Forest (RF) is an ensemble learning algorithm built upon the principles of decision trees. A decision tree splits a dataset into progressively smaller subsets based on input variables, aiming to reduce variance in outcome values at each step. This hierarchical structure consists of branches representing decision rules and leaves containing outcome predictions. The splitting process continues until a predefined condition, such as a maximum number of splits or a variance threshold, is met [35]. RF enhances the predictive power of decision trees by constructing multiple trees using bootstrap sampling of the training data. Each tree is trained on a different subset of the data, and the final prediction is obtained through majority voting (for classification) or averaging (for regression). This approach reduces overfitting and improves model generalization. The effectiveness of RF lies in its ability to combine multiple decision trees, leveraging their individual strengths while minimizing weaknesses through ensemble learning techniques [36].

Support Vector Regression (SVR) is a supervised learning algorithm designed to handle non-linear relationships between input variables and target labels. Unlike traditional regression models, which assume a linear relationship, SVR transforms the input data into a higher-dimensional space where the relationship can be more effectively captured as a linear function. To achieve this transformation, SVR employs kernel functions, which map the original data into a higher-dimensional feature space without the need for explicit computation in that space, thereby reducing computational complexity. In this transformed space, SVR fits a hyperplane that best represents the data while maintaining a margin of tolerance, allowing the model to generalize well to new inputs. This ability to capture complex patterns makes SVR a powerful tool for regression tasks where traditional linear models fall short [37].

K-Nearest Neighbors (KNN) is a simple yet effective algorithm that predicts the outcome of a data point based on its **K** nearest neighbors in the training set. The nearest neighbors are identified using a distance metric, such as Euclidean distance for numerical variables or Hamming distance for categorical variables. The final prediction is obtained by averaging the labels of the **K** closest points (for regression) or by majority voting (for classification). KNN is non-parametric and relies on the assumption that similar data points have similar outcomes [38-39].

Lastly, XGBoost is a high-performance boosting algorithm designed for speed and accuracy. It builds decision trees sequentially, with each new tree learning to correct the residuals of the previous ones. To enhance model generalization, XGBoost incorporates a regularization term in the cost function, balancing predictive power and complexity. The final prediction is obtained by summing the contributions from all trees, making it a robust and efficient approach for various machine learning tasks [40].

Performance evaluation metrics for machine learning models

To evaluate the performance of the predictive models, this study employs the coefficient of determination (R^2) and root mean squared error (RMSE) as key metrics. R^2 measures the goodness of fit, quantifying the proportion of variance in the target variable explained by the independent variables. It is calculated as shown in Equation 2:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (2)$$

where is the actual value, is the predicted value, and is the mean of the actual values. R^2 ranges from 0 to 1, with higher values indicating better predictive accuracy.

The root mean squared error (RSME) evaluates the average magnitude of prediction errors by considering squared differences between predicted and actual values, followed by a square root transformation, presented in Equation 3:

$$RSME = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (3)$$

where is the number of samples, is the actual value, and is the predicted value. RMSE penalizes larger errors more than smaller ones, making it useful for assessing model precision.

RESULTS AND DISCUSSION

Weight loss measurements

The weight loss tests of carbon steel samples in CO_2-saturated solutions at 70°C at different brine concentration for NaCl and $CaCl_2$ solutions were investigated. **Figure 1** shows the weight loss results, in which the corrosion rate in the presence of $CaCl_2$ is consistently higher than that observed in NaCl solutions.

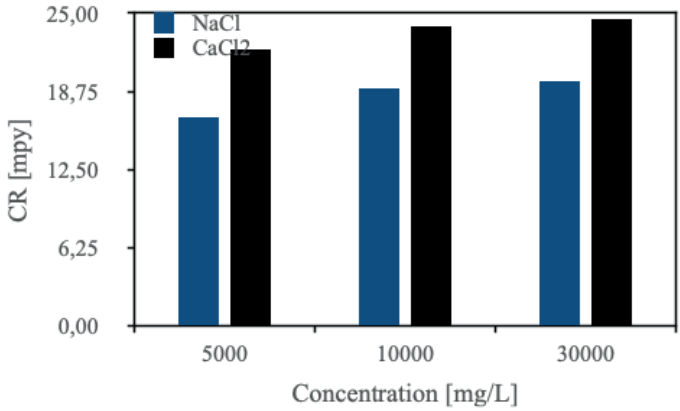


Figure 1. Corrosion rates of carbon steel at different salinity after 24 h inmersion time

Source: produced by the author

This results, suggests that the presence of Ca^{2+} ions enhances the corrosion rate, potentially due to their influence to alter the stability of protective oxide layers and affecting the solubility of corrosion products, thereby increasing the overall corrosion rate.

Aditionally, the results presented in the **Figure 2** illustrate the effect of 100 ppm CQD on the corrosion rate (CR) of carbon steel in NaCl and $CaCl_2$ brines at 70°C after 24 hours. A notable reduction in CR is observed in both solutions upon CQD addition, with a more pronounced inhibition effect in $CaCl_2$ solutions. The reduction percentages in CR indicate that CQDs effectively mitigate corrosion, with a maximum inhibition efficiency of **46.2%** in $CaCl_2$ and **45.3%** in NaCl at the lowest concentration (5000 mg/L).

These findings suggest that CQDs may act as corrosion inhibitors by interacting with the steel surface, possibly forming a protective layer that limits metal dissolution. Furthermore, the stronger inhibition observed in $CaCl_2$ solutions could indicate a chelating effect, where CQDs coordinate with Ca^{2+} ions, altering their impact on corrosion dynamics. This potential chelation mechanism may contribute to the stabilization of corrosion products or reduce their solubility, thereby enhancing the inhibition effect.

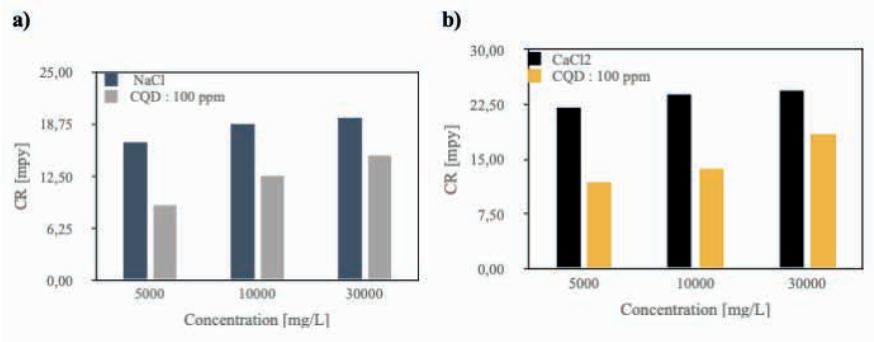


Figure 2. Corrosion rate of carbon steel in presence and absence of carbon quantum dots (**a**) NaCl and (**b**) CaCl2

Source: produced by the author

Data preprocessing

Before developing machine learning models, it is essential to carefully organize and analyze the data using various analytical techniques. This process helps uncover key characteristics and hidden patterns within the dataset, providing valuable insights that are crucial for building effective models. The dataset consists of 108 sample points, each containing six features: temperature, inmersion time, pH, salinity, inhibitor concentration and brine type. A summary of the data is presented in **Table 2**.

	Inhibitor	Time	pH	Salinity	CR
Mean	50.56	72.00	3.82	15018.43	10.95
Std	49.23	39.37	0.17	10892.79	5.73
Min	0.00	24.00	3.31	4510.00	3.20
25%	0.00	24.00	3.73	5119.50	6.28
50%	49.50	72.00	3.85	10000.00	9.33
75%	100.00	120.00	3.97	29997.75	14.28
Max	100.00	120.00	4.05	30435.00	25.42

Table 2. Data summary

Source: produced by the author

Table 2 provides a statistical analysis of the key influencing factors in the dataset. The data exhibit considerable dispersion, particularly in salinity, inhibitor concentration, and corrosion rate. The average corrosion rate is 10.95 mpy, with a minimum of 3.20 and a maximum of 25.42, indicating that higher corrosion rates

occur under specific conditions and may represent extreme cases. The distribution of corrosion rates suggests that most values fall within a moderate range, with a standard deviation of 5.73.

The inhibitor concentration varies widely, indicating a clear distinction between samples with and without inhibitors. Similarly, salinity ranges from 4,510 to 30,435 mg/L, with an average of 15,018 mg/L and a high standard deviation suggesting significant variations in experimental conditions. The pH remains relatively stable, with values concentrated around a mean of 3.82 and a narrow standard deviation of 0.17, ensuring a consistently acidic environment.

These statistical insights highlight key trends in the dataset, allowing for a better understanding of how different factors contribute to corrosion rates. The observed variability suggests that corrosion behavior is influenced by multiple interacting variables, underscoring the need for predictive modeling to identify dominant effects and optimize corrosion control strategies.

Correlation analysis

The correlation matrix provides insights into the relationships between various factors influencing corrosion. As shown in **Figure 1**, the correlation coefficients between the studied variables reveal key interactions affecting the corrosion rate (CR).

The inhibitor concentration exhibits a moderate negative correlation with CR (-0.47), indicating that higher inhibitor levels effectively reduce the corrosion rate. Similarly, immersion time shows a strong negative correlation with CR (-0.76), suggesting that prolonged exposure contributes to corrosion mitigation, potentially due to the formation of protective layers. The pH also presents a negative correlation with CR (-0.49), implying that higher pH values (less acidic conditions) are associated with lower corrosion rates. Conversely, salinity has a weak correlation with CR (0.10), suggesting a minimal direct impact on the corrosion process within the tested conditions, despites the results obtained in the weight loss tests. Overall, the correlation coefficients indicate that the influencing factors operate relatively independently, and feature reduction is not necessary. These findings highlight the importance of inhibitor concentration, exposure time, and pH control in corrosion prevention strategies

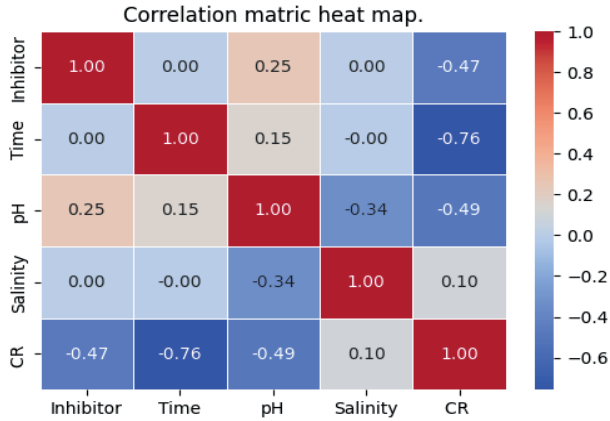
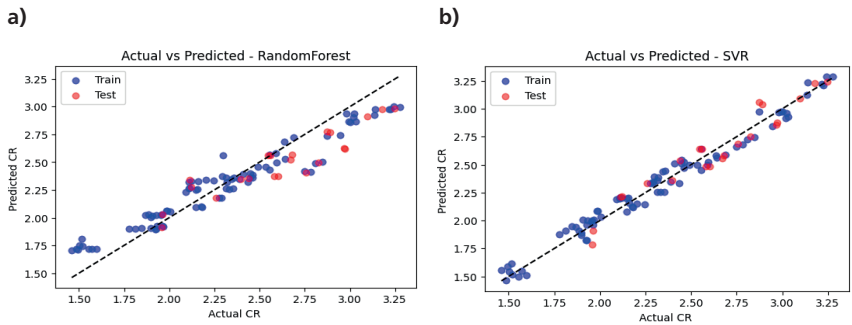


Figure 3. Correlation matrix heat map

Source: produced by the author

ML models for CO_2 corrosion rate

In this study, widely used algorithms relevants in corrosion field, such as RF, KNN, SVR, and XGBoost, were employed to predict the CO_2 corrosion rate in presence of divalent ions and carbon quantum dots as corrosion inhibitors. To thoroughly evaluate the performance of these models, RMSE and R^2 were chosen as the primary assessment metrics. **Figure 2** illustrates the predictive performance of each algorithm in modeling the corrosion rate. By analyzing these results, we can compare the strengths and limitations of different approaches, gaining deeper insights into their effectiveness in corrosion rate prediction.



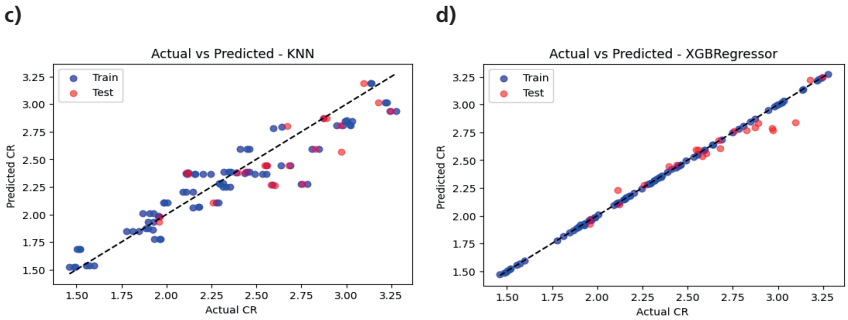


Figure 4. Fit graps of corrosion rates with different machine learning algortihms **(a)** RF, **(b)** SVR, **(c)** KNN and **(d)** XGBoost. Dashed lines represents the agreement between actual and predicted corrosion rates

Source: produced by the author

Figure 2 presents the fitting curves of corrosion rate predictions for different machine learning algorithms. The left axis represents the actual corrosion rate, while the right axis corresponds to the predicted values generated by each model. The X = Y line serves as a reference, representing an ideal scenario where the predicted values perfectly match the actual corrosion rates. In an optimal case, all data points would be closely aligned with this reference line, indicating high predictive accuracy. The distribution of predictions across both the training and test sets suggests a reasonable approximation to actual values, reflecting a certain level of model consistency.

A comparative analysis reveals that the fitting performance of Random Forest (RF) model in **Figure 2a** exhibits a moderate alignment between predicted and actual values, though some dispersion is observed, particularly in the test set. In contrast, the SVR model in **Figure 2b** demonstrates a stronger fitting performance, with predictions closely following the X = Y reference line, suggesting higher accuracy. The KNN model, depicted in **Figure 2c**, achieves reasonable performance, but noticeable deviations appear in both training and test sets, indicating potential overfitting or sensitivity to data distribution. Among all models, XGBoost in **Figure 2d** appears to exhibits the best fitting performance, with predictions almost perfectly aligned along the X = Y line. Despite these observations, the evaluation metrics are necessary to calculate the performance assessment of these models.

In **Figure 5**, it is observed that the SVR model demonstrates the best predictive performance among the evaluated models. It achieves a high R^2 value of 0.9725 on the training set and 0.9281 on the test set, indicating strong generalization capabilities while maintaining a relatively low RMSE (0.9210 for training and 1.4802 for testing). This suggests that SVR effectively captures the underlying patterns in the data, making it the most reliable model for corrosion rate prediction.

The XGBoost model also exhibits strong predictive performance, however, its near-perfect fit on the training set, indicating overfitting, as the model may be memorizing patterns rather than generalizing them. Despite this, its test RMSE (1.5994) is relatively low, indicating that it still performs well on unseen data but might require further tuning to reduce overfitting.

In contrast, RF and KNN show weaker generalization, with significantly lower R^2 values on the test set (0.6820 and 0.6677, respectively) and higher RMSE values (3.1136 and 3.1830). These results indicate that RF and KNN models struggle to maintain predictive accuracy on new data, suggesting a potential lack of complexity in capturing the data relationships effectively.

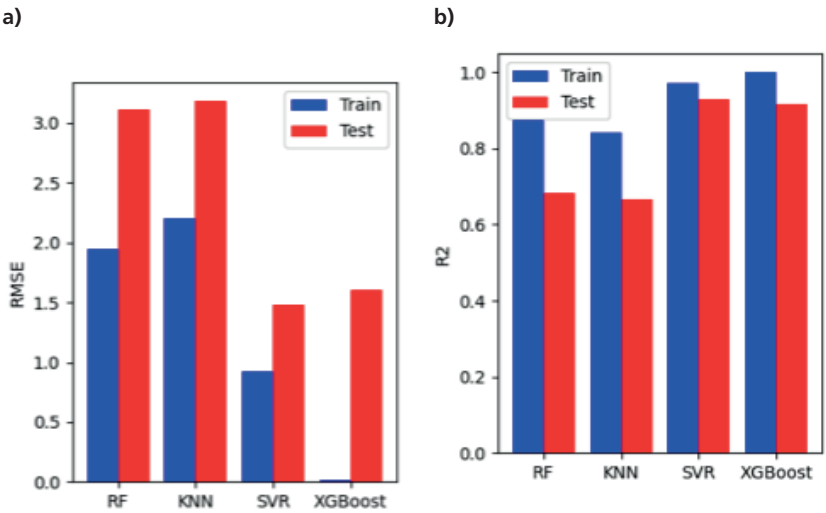


Figure 5. RSME and R^2 charts for different models

Source: produced by the author

CONCLUSIONS AND/OR FINAL CONSIDERATIONS

This study evaluated the corrosion behavior of carbon steel in CO_2-saturated brines with varying concentrations of NaCl and $CaCl_2$, and the effect of CQD as a corrosion inhibitor in this scenario. The presence of Ca^{2+} ions was consistently associated with higher corrosion rates compared to NaCl solutions at equivalent concentrations. This observation suggests that Ca^{2+} influences the stability of protective oxide layers and alters the solubility of corrosion products, accelerating material degradation. The addition of CQD at 100 mg L^{-1} demonstrated significant inhibition effects in both NaCl and $CaCl_2$ solutions, with a more pronounced impact

in $CaCl_2$ solutions. This supports the hypothesis that CQD not only acts as a corrosion inhibitor but also as a chelating agent for Ca^{2+} ions, mitigating their influence on steel corrosion.

Among the machine learning models tested, SVR emerged as the best-performing algorithm, achieving the highest R^2 values and lower RMSE for both training and test sets. This suggests that SVR effectively captured the underlying patterns governing corrosion rates. However, some degree of overfitting was still observed, indicating that further refinement is necessary to enhance generalization capabilities. On the other hand, XGBoost exhibited extreme overfitting, delivering near-perfect predictions on the training set while showing a significant drop in accuracy on the test set. While XGBoost successfully captured intricate relationships within the data, its poor generalization underscores the importance of employing more robust validation techniques and model tuning strategies.

Despite the insights gained, the current study emphasizes the need for further exploration of experimental conditions to improve model accuracy and robustness. Expanding the range of environmental variables, such as temperature, pH variations, and different CQD concentrations, could provide a more comprehensive dataset for model training. Moreover, optimizing feature selection and hyperparameter tuning will be essential to reduce overfitting and enhance predictive performance. By addressing these challenges, future research can move toward more reliable and generalizable models for corrosion rate estimation in CO_2 environments.

BIBLIOGRAPHIC REFERENCES

[1] J. Choi, J.N. Seyyed, A. Hosseini, S.J. Clift, S.D. Hovorka, CO2recycling accounting and EOR operation scheduling to assist in storage capacity assessment at a U.S. gulf coast depleted reservoir, Int. J. Greenhouse Gas Control 18 (2013) 474–484.

[2] S. Sim, I.S. Cole, Y.-S. Choi, N. Birbilis, A review of the protection strategies against internal corrosion for the safe transport of supercritical CO2via steel pipelines for CCS purposes, Int. J. Greenhouse Gas Control 29 (2014) 185–199.

[3] Navabzadeh Esmaeely, S., Choi, Y. S., Young, D., & Nešić, S. (2013). Effect of calcium on the formation and protectiveness of iron carbonate layer in CO2 corrosion. Corrosion, 69(9), 912-920.

[4] R.H. Hausler, H.P. Gaddart, Advances in CO2 Corrosion Vol. 2 NACE International, Houston TX, 1986.

[5] M. Rogowska, J. Gudme, A. Rubin, K. Pantleon, R. Ambat, Effect of Fe ion con centration on corrosion of carbon steel in CO 2 environment, Corros. Eng. Sci. Technol. 51 (2016) 25–36, https://doi.org/10.1179/1743278215Y.0000000029.

[6] S. Nešić, Key issues related to modelling of internal corrosion of oil and gas pipe lines- A review, Corros. Sci. 49 (2007) 4308–4338, https://doi.org/10.1016/j. corsci.2007.06.006.

[7] A. Kahyarian, M. Singer, S. Nesic, Modeling of uniform CO2 corrosion of mild steel in gas transportation systems: a review, J. Nat. Gas Sci. Eng. 29 (Suppl. C) (2016) 530–549.

[8] H. Mansoori, R. Mirzaee, F. Esmaeilzadeh, A. Vojood, A.S. Dowrani, Pitting corrosion failure analysis of a wet gas pipeline, Eng. Fail. Anal. 82 (2017) 16–25.

[9] Hua, Y., Barker, R., & Neville, A. (2015). Comparison of corrosion behaviour for X-65 carbon steel in supercritical CO2-saturated water and water-saturated/unsaturated supercritical CO2. The Journal of Supercritical Fluids, 97, 224-237.

[10] Kermani, M. B., and A. Morshed. "Carbon dioxide corrosion in oil and gas productiona compendium." Corrosion 59.08 (2003).

[11] Forero, A., M. Núñez, and I. Bott, *Analysis of the corrosion scales formed on API 5L X70 and X80 steel pipe in the presence of CO2.* Materials Research, 2014. 17: p. 461-471.

[12] Fang, H., B. Brown, and S. Nešić, *Sodium chloride concentration effects on general CO2 corrosion mechanisms.* Corrosion, 2013. 69(3): p. 297-302.

[13] Elgaddafi, R., R. Ahmed, and S. Shah, *Corrosion of carbon steel in CO2 saturated brine at elevated temperatures.* Journal of Petroleum Science and Engineering, 2021. 196: p. 107638.

[14] Mansoori, H., et al., *Case study: production benefits from increasing C-values.* Oil & gas journal, 2013. 111(6): p. 64-73.

[15] Rizzo, R., et al., *Corrosion of carbon steel under CO2 conditions: Effect of CaCO3 precipitation on the stability of the FeCO3 protective layer.* Corrosion Science, 2020. 162: p. 108214.

[16] Esmaeely, S.N., et al., *Effect of incorporation of calcium into iron carbonate protective layers in CO2 corrosion of mild steel.* Corrosion, 2017. 73(3): p. 238-246.

[17] Tavares, L.M., et al., *Effect of calcium carbonate on low carbon steel corrosion behavior in saline CO2 high pressure environments.* Applied Surface Science, 2015. 359: p. 143-152.

[18] Ding, Chen, Ke-wei Gao, and Chang-feng Chen. "Effect of Ca2+ on CO2 corrosion properties of X65 pipeline steel." *International Journal of Minerals, Metallurgy and Materials* 16.6 (2009): 661-666.

[19] Shamsa, Amir, et al. "The role of Ca2+ ions on Ca/Fe carbonate products on X65 carbon steel in CO2 corrosion environments at 80 and 150° C." *Corrosion Science* 156 (2019): 58-7

[20] Nor, A. M., Suhor, M., Mohamed, M., Singer, M., & Nesic, S. (2011). Corrosion of carbon steel in high CO2 environment: flow effect. NACE CORROSION,

[21] A. Charles, K. Sivaraj, S. Thanikaikarasan, Synthesis, characterization and corrosion studies of Schiff bases derived from pyrrole-2-carbaldehyde, Mater. Today: Proc. 33 (2020) 3135–3138

[22] A. Shamsa , R. Barker , Y. Hua , E. Barmatov, T.L. Hughes , A. Neville , Performance evaluation of an imidazoline corrosion inhibitor in a CO 2 -saturated environment with emphasis on localised corrosion, Corros. Sci. 176 (2020) 108916

[23] Tang Y, Zhang F, Hu S, Cao Z, Wu Z, Jing W. Novel benzimidazole derivatives as corrosion inhibitors of mild steel in the acidic media. Part I: gravimetric, electrochemical, SEM and XPS studies. Corros Sci 2013;74:271–82 .

[24] Luo X, Pan X, Yuan S, Du S, Zhang C, Liu Y. Corrosion inhibition of mild steel in simulated seawater solution by a green eco-friendly mixture of glucomannan (GL) and bisquaternary ammonium salt (BQAS). Corros Sci 2017;125:139–51

[25] Cen, Hongyu, Zhenyu Chen, and Xingpeng Guo. "N, S co-doped carbon dots as effective corrosion inhibitor for carbon steel in CO2-saturated 3.5% NaCl solution." *Journal of the Taiwan Institute of Chemical Engineers* 99 (2019): 224-238

[26] Wu, Xiaodan, et al. "Novel carbon dots as effective corrosion inhibitor for N80 steel in 1 M HCl and CO2-saturated 3.5 wt% NaCl solutions." *Journal of Molecular Structure* 1250 (2022): 131897.

[27] W.T. Nash, C.J. Powell, T. Drummond, N. Birbilis, Automated corrosion detection using crowdsourced training for deep learning, Corrosion 76 (2019) 135–141, https://doi.org/10.5006/3397.

[28] C.I. Ossai, A data-driven machine learning approach for corrosion risk assessment—a comparative study, Big Data Cogn. Comput. 3 (2019) 28, https:// doi.org/10.3390/bdcc3020028.

[29] G. Sanchez, W. Aperador, A. Cer´ on, Corrosion grade classification: a machine learning approach, Indian Chem. Eng. 62 (2020) 277–286, https://doi.org/10.1080/00194506.2019.1675539.

[30] L. Yan, Y. Diao, Z. Lang, K. Gao, Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach, Sci. Technol. Adv. Mater. 21 (2020) 359–370, https://doi.org/10.1080/ 14686996.2020.1746196.

[31] Aghaaminiha, Mohammadreza, et al. "Machine learning modeling of time-dependent corrosion rates of carbon steel in presence of corrosion inhibitors." *Corrosion Science* 193 (2021): 109904

[32] Dong, Zhenzhen, et al. "Development of a Predictive Model for Carbon Dioxide Corrosion Rate and Severity Based on Machine Learning Algorithms." *Materials* 17.16 (2024): 4046.

[33] T.M. Mitchell, Does machine learning really work?, 11–11, AI Mag. 18 (1997), https://doi.org/10.1609/aimag.v18i3.1303.

[34] A. G´ eron, Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts. Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, Inc, 2019.

[35] L. Breiman, Random forests, Mach. Learn 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[36] Paul, A.; Mukherjee, D.P.; Das, P.; Gangopadhyay, A.; Chintha, A.R.; Kundu, S. Improved Random Forest for Classification. IEEE Trans. Image Process. 2018, 27, 4012–4024.

[37] S.R. Gunn, Support vector machines for classification and regression, ISIS Tech. Rep. 14 (1998) 5–16.

[38] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1967) 21–27, https://doi.org/10.1109/TIT.1967.1053964.

[39] S. Xia, Z. Xiong, Y. Luo, L. Dong, G. Zhang, Location difference of multiple distances-based k-nearest neighbors algorithm, Knowl. -Based Syst. 90 (2015) 99–110, https://doi.org/10.1016/j.knosys.2015.09.028.

[40] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Fracisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.

ACKNOWLEDGMENTS

The authors acknowledge MINCIENCIAS for the financial support provided through contract No. 112721-282-2023 for the development of this research, as well as the National University of Colombia - Sede Medellín.