UNIVERSIDAD
NACIONAL
DE COLOMBIA

MODELO DE APRENDIZAJE DE MÁQUINAS PARA LA PREDICCIÓN DE LA TASA DE CORROSIÓN POR CO_2 EN SALMUERAS CON IONES DIVALENTES Y PRESENCIA DE CARBON QUANTUM DOTS (CQDs)

Leidy Viviana Ortiz Pérez

Universidad Nacional de Colombia
Facultad de Minas
Medellín, Colombia
2025

MODELO DE APRENDIZAJE DE MÁQUINAS PARA LA PREDICCIÓN DE LA TASA DE CORROSIÓN POR CO_2 EN SALMUERAS CON IONES DIVALENTES Y PRESENCIA DE CARBON QUANTUM DOTS (CQDs)

Leidy Viviana Ortiz Pérez

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en Ingeniería - Analítica

Director (a):

Ph.D. Farid Bernardo Cortés Correa

Codirector (a):

Ph.D. Maria Constanza Torres Madronero

Ph.D. Camilo Andrés Franco Ariza

Grupo de Investigación Fenómenos de superficie – Michael Polanyi

Universidad Nacional de Colombia

Facultad de Minas

Medellín, Colombia

2025

A mi madre, familia y Buddy *por ser mi inspiración y mi mayor motivación. Agradezco su paciencia y apoyo, que hicieron posible este sueño.*

Agradecimientos

Agradezco a mis directores de tesis, los profesores Farid B. Cortés, Camilo A. Franco y María Constanza Torres. Su invaluable sabiduría, paciencia y el conocimiento que compartieron fueron fundamentales para la concepción y el desarrollo de este trabajo. Aprecio enormemente la guía y el apoyo constante que me brindaron a lo largo de este enriquecedor proceso.

Agradezco especialmente a Lady J. Giraldo por sus ideas, ayuda y retroalimentación desde el inicio de este proyecto. Ha sido un honor y placer aprender y crecer profesionalmente a su lado.

Agradezco a mi amigos que me apoyaron durante este tiempo, en especial a Daniel Gómez, por su constante compañía, consejos y apoyo incondicional en los momentos más desafiantes.

Finalmente, este proyecto de investigación fue posible gracias a la financiación otorgada por el Fondo Francisco José de Caldas, MINCIENCIAS y la Agencia Nacional de Hidrocarburos (ANH), a través del contrato No. 112721-282-2023 (Proyecto 1118-1035-9300) con la Universidad Nacional de Colombia – Sede Medellín y PAREX RESOURCES COLOMBIA AG SUCURSAL. Su apoyo fue crucial para la realización de este estudio de maestría.

Resumen

Modelo de aprendizaje de máquinas para la predicción de la tasa de corrosión por CO_2 en salmueras con iones divalentes y presencia de Carbon Quantum Dots (CQDs)

La corrosión del acero al carbono en ambientes salinos saturados con CO_2 constituye un reto crítico para la confiabilidad de las infraestructuras en energía y procesos de Captura, Utilización y Almacenamiento de Carbono (CCUS). Las salmueras de formación con NaCl, $CaCl_2$ y CO_2 disuelto aceleran la degradación electroquímica, generando altas tasas de corrosión y riesgos operativos. Como alternativa de mitigación, los Carbon Quantum Dots (CQDs) se perfilan como inhibidores verdes prometedores por su dispersabilidad, estabilidad química y bajo impacto ambiental. Sin embargo, predecir con precisión estas tasas sigue siendo un desafío, ya que los modelos tradicionales no logran representar las complejas interacciones entre múltiples variables. En esta tesis se evaluó experimentalmente el efecto inhibidor de los CQDs mediante ensayos gravimétricos bajo diferentes condiciones de salinidad y tiempos de inmersión, y se desarrollaron modelos de aprendizaje de máquinas para predecir la tasa de corrosión en estos ambientes. Se evaluaron cinco algoritmos: Random Forest (RF), XGBoost, Support Vector Regression (SVR), K-Near Neighbors (KNN) y Multilayer Perceptron (MLP). El mejor desempeño predictivo se obtuvo con RF y XGBoost, alcanzando R^2 superiores a 0.91, RMSE de 1.1 mpy y MAE de 0.8 mpy. Estos resultados demuestran el potencial del aprendizaje de máquinas como herramienta poderosa y confiable para predecir el comportamiento corrosivo en salmueras complejas ricas en CO_2, apoyando la implementación de estrategias de mitigación de la corrosión más seguras, eficientes y sostenibles en aplicaciones industriales.

Palabras clave: Aprendizaje de máquinas, CO_2 corrosión, Carbon quantum dots.

Abstract

Machine Learning Model for Predicting CO_2 Corrosion Rate in Brines with Divalent Ions and the Presence of Carbon Quantum Dots (CQDs)

The corrosion of carbon steel in CO_2-saturated saline environments represents a critical challenge for the reliability of infrastructures in energy production and Carbon Capture, Utilization, and Storage (CCUS) processes. Formation brines containing NaCl, $CaCl_2$, and dissolved CO_2 accelerate electrochemical degradation, leading to high corrosion rates and significant operational risks. As a mitigation alternative, Carbon Quantum Dots (CQDs) have emerged as promising green inhibitors due to their dispersibility, chemical stability, and low environmental impact. However, accurately predicting corrosion rates remains a major challenge, since traditional models fail to capture the complex interactions among multiple variables. In this thesis, the inhibitory effect of CQDs was experimentally evaluated through gravimetric tests under different salinity conditions and immersion times, and machine learning models were developed to predict the corrosion rate in these environments. Five algorithms were evaluated: Random Forest (RF), XGBoost, Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). The best predictive performance was obtained with RF and XGBoost, achieving R^2 values above 0.91, RMSE of 1.1 mpy, and MAE of 0.8 mpy. These results demonstrate the potential of machine learning as a powerful and reliable tool to predict corrosive behavior in complex CO_2-rich brines, supporting the implementation of safer, more efficient, and sustainable corrosion mitigation strategies in industrial applications.

Keywords: Machine learning, CO_2 corrosion, Carbon quantum dots.

Contenido

Pág.

Resumen ..... **IX**

Lista de figuras ..... **XIV**

Lista de tablas ..... **XV**

Lista de Símbolos y abreviaturas ..... **XVI**

Introducción ..... Error! Bookmark not defined. **17**

1. Planteamiento del problema **21**
1.1 Justificación 22
1.2 Objetivo General 22
1.2.1 Objetivos Específicos 23

2. Marco teórico **24**
2.1 Aprendizaje de máquinas 24
2.1.1 Modelos de aprendizaje supervisado 24
2.1.1.1 Random Forest (RF) 25
2.1.1.2 Extreme Gradient Boosting (XGBoost) 26
2.1.1.3 Support Vector Regression (SVR) 26
2.1.1.4 K-Nearest Neighbors (KNN) 26
2.1.1.5 Multi-layer Perceptron (MLP) 27

3. Estado del arte **28**

4. Metodología **31**
4.1 Evaluación de corrosión 31
4.2 Conformación del conjunto de datos experimentales 33
4.2.1 Preprocesamiento y análisis exploratorio 34
4.3 Modelos de aprendizaje de maquinas para la predicción de corrosión en salmueras con iones divalentes y CQDs 35
4.3.1 Métricas de desempeño 36

5. Resultados **37**
5.1 Medidas gravimétricas 37
5.1.1 Efecto de iones: Na^+ vs Ca^{2+} 37
5.1.2 Efecto de CQDs 39
5.1.3 Modelado de tasas de corrosión con SCDM 44
5.2 Análisis descriptivo del conjunto de datos 48

5.2.1 Análisis de Correlación 50
5.3 Desempeño de modelos de aprendizaje de máquinas implementados para la predicción de tasas de corrosión..... 52
5.3.1 Predicción de la tasa de corrosión en condiciones iniciales 52
5.3.2 Predicción de la tasa de corrosión en el tiempo 58

Conclusiones y recomendaciones 63

Bibiliografía.....65

Lista de figuras

Pág.

Figura 1. Modelo general de aprendizaje supervisado 24
Figura 2. Arquitectura de MLP 27
Figura 3. Efecto del tipo de ion (Na^+ vs Ca^{2+}) en la tasa de corrosión de acero al carbono 37
Figura 4. Influencia del tipo de ion en la fuerza iónica a diferentes salinidades 38
Figura 5. Tasas de corrosión de acero al carbón en soluciones con presencia y ausencia de CQDs **a)** NaCl y **b)** $CaCl_2$ 40
Figura 6. Tasa de corrosión y eficiencia de inhibición de acero al carbono en soluciones con presencia de varias concentraciones de CQDs **a)** NaCl y **b)** $CaCl_2$ 41
Figura 7. Tasa de corrosión y eficiencia de inhibición de acero al carbón a diferente temperatura durante 24h **a)** NaCl y **b)** $CaCl_2$ 42
Figura 8. Tasa de corrosión del acero al carbono y eficiencia de inhibición luego de varios tiempos de inmersión **a)** NaCl y **b)** $CaCl_2$ 43
Figura 9. SCMD en ausencia de CQDs **a)** Contornos de la superficie de respuesta y **b)** Superficie de respuesta estimada 44
Figura 10. SCMD con presencia de CQDs **a)** Contornos de la superficie de respuesta y **b)** Superficie de respuesta estimada 46
Figura 12. Distribución de la tasa de corrosión frente a variables experimentales **a)** Tipo de ion **b)** Temperatura **c)** Tiempo de inmserión **d)** Salinidad **e)** Concentración de CQD y **f)** Fuerza iónica 50
Figura 13. Matriz de correlación 51
Figura 14. Comparación de los valores reales y predichos con diferentes modelos de aprendizaje supervisado evaluados a 24h **a)** RF **b)** XGBoost **c)** KNN **d)** SVR y **e)** MLP . 53
Figura 15. Importancia de caracteristicas para RF implementado en la predicción de tasas de corrosion a 24h 54
Figura 16. Métricas de desempeño para los modelos evaluados en la predicción de tasas de corrosión a 24h **a)** R^2 **b)** RMSE y **c)** MAE 56
Figura 17. Comparación de los valores reales y predichos con diferentes modelos de aprendizaje supervisado evaluados en el tiempo **a)** RF **b)** XGBoost **c)** KNN **d)** SVR y **e)** MLP 59
Figura 18. Importancia de caracteristicas para RF implementado en la predicción de tasas de corrosion en el tiempo 59
Figura 19. Métricas de desempeño para los modelos evaluados en la predicción de tasas de corrosión en el tiempo **a)** R^2 **b)** RMSE y **c)** MAE 60

Lista de tablas

Pág.

Tabla 1. Diseño de mezclas Simplex-Centroide 33
Tabla 2. Variables operaciones consideradas para la predicción de tasas de corrosión 33
Tabla 3. Búsqueda de Hiperparámetros para los modelos implementados en la predicción de tasas de corrosión 35
Tabla 4. Métricas implementadas en los modelos de predicción 36
Tabla 5. Coeficientes estimados para modelo de CR en ausencia de CQD 45
Tabla 6. Predicción de CR con SCDM en ausencia de CQDs 45
Tabla 7. Coeficientes estimados para modelo de CR en presencia de CQD 47
Tabla 8. Predicción de CR con SCDM en presencia de CQDs 47
Tabla 9. Análisis descriptivo de las variables experimentales 48
Tabla 10. Hiperparámetros resultantes de la optimización de los modelos implementados para la predicción de tasas de corrosion a 24h 57
Tabla 11. Hiperparámetros resultantes de la optimización de los modelos implementados para la predicción de tasas de corrosion en el tiempo 61

Lista de Símbolos y abreviaturas

Abreviatura	Término
CO_2	Dióxido de carbono
CQDs	Carbon Quantum Dots
Ca^{2+}	Ion de calcio
$FeCO_3$	Siderita
$CaCO_3$	Calcita
RF	Random Forest
KNN	K-Nearest Neighbors
SVR	Support Vector Regression
XGBoost	Extreme Gradient Boosting
MLP	Multi-layer Perceptron

Introducción

El cambio climático se ha convertido en una de las mayores amenazas contemporáneas, impulsada principalmente por el aumento de las emisiones de gases de efecto invernadero (GEI), como el dióxido de carbono (CO_2), el metano (CH_4) y el óxido nitroso (N_2O). Entre ellos, el dióxido de carbono es reconocido como el principal GEI. Sus niveles atmosféricos continúan aumentando a nivel mundial, pasando de 280 ppm en la era preindustrial a un nivel récord de 414 ppm en 2020. Este aumento del 32 % se atribuye en gran parte a la combustión de combustibles fósiles para la generación de energía y las operaciones industriales [1], [2], [3], [4]. Según el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC), el CO_2 representa aproximadamente el 60% del forzamiento radiativo total, lo que intensifica el efecto invernadero y acelera el calentamiento global, lo que plantea desafíos a las sociedades y economías de todo el mundo [5], [6], [7], [8], [9], [10]. Por lo tanto, esta aceleración del cambio climático ha intensificado la necesidad de medidas drásticas para reducir las emisiones de dióxido de carbono y limitar el aumento de la temperatura global por debajo de 2 °C para finales de siglo, tal y como se establece en el Acuerdo de París sobre el Clima (PCA) [11], [12].

Para mitigar y prevenir un mayor daño ambiental, es necesario reducir la concentración de CO_2 a 450 ppm para 2100, de acuerdo con la Agencia Internacional de Energía (IEA, 2021). [13]. Para abordar este desafío, las tecnologías de Captura, Utilización y Almacenamiento de Carbono (CCUS) han surgido como soluciones prometedoras para reducir las emisiones de CO_2. Estas tecnologías están diseñadas para capturar el CO_2 directamente de sus fuentes de emisión (como fábricas de cemento, centrales eléctricas, industrias químicas y refinerías) y evitar que se libere a la atmósfera. El CO_2 capturado puede ser transportado para diversos usos, como la producción de químicos y la recuperación mejorada de petróleo, o bien puede ser almacenado en depósitos geológicos profundos. Se proyecta que para 2050, las tecnologías CCUS contribuyan hasta un 17 % de las reducciones globales de emisiones, asegurando así una transición hacia un futuro sostenible y con bajas emisiones de carbono [14][15] [16].

Para la industria del petróleo y gas, las tecnologías CCUS le permiten avanzar hacia un desarrollo eficiente y sostenible, al mismo tiempo que aseguran un suministro confiable de energía fósil [17]. En este sentido, la recuperación mejorada de petróleo con CO_2 (CO_2-EOR) acoplada con CCUS (CCUS-EOR) es actualmente la tecnología CCUS más viable económicamente ya que permite mayores tasas de recuperación, lo que contribuye a la seguridad energética. Según la IEA, la industria del petróleo y gas participa en casi el 90% de todos los proyectos CCUS operativos. Esto impulsa la reducción de las emisiones de CO_2 y el almacenamiento geológico a gran escala, al tiempo que logra una transformación del sector energético hacia un modelo con bajas emisiones de carbono [18].

Entre las tecnologías de CO_2-EOR, la inyección de agua carbonatada (CWI) y la inyección de agua carbonatada mejorada (ECW) se consideran técnicas prometedoras, ya que pueden mejorar significativamente la recuperación de petróleo entre un 37% y un 65% [19]. Sus mecanismos principales incluyen la mejora de la disolución del CO_2, la prolongación de su retención en la fase acuosa, la reducción de la viscosidad del crudo, la disminución de la tensión interfacial y la alteración de la humectabilidad [20][21]. Sin embargo, la presencia de CO_2 tiene una influencia crucial en la corrosión de las tuberías e instalaciones relacionadas que se utilizan en la producción y explotación de pozos de petróleo y gas. Cuando el CO_2 (ya sea inyectado o presente de forma natural en la geología local) se disuelve en agua, se hidrata para formar ácido carbónico (H_2CO_3), lo que disminuye el pH del agua de la formación y provoca reacciones entre la fase acuosa y la superficie del acero, lo que lleva a la disolución del hierro [22][23]. La corrosión por CO_2 en el acero al carbono representa serios desafíos para la industria, debido que puede causar la perforación y el agrietamiento de los sistemas de tuberías, lo que podría provocar accidentes catastróficos y enormes pérdidas económicas [24]. A pesar de su susceptibilidad y escasa resistencia a la corrosión en ambientes ricos en CO_2, el acero al carbono sigue siendo ampliamente utilizado en la producción y el transporte de petróleo y gas, debido a su alta resistencia a la tracción, bajo costo y amplia disponibilidad [25][26].

Las películas de productos de corrosión que se forman en la superficie del acero al carbono en ambientes con CO_2 generalmente se cree que mitigan el ataque corrosivo al impedir la difusión de los agentes corrosivos. Sin embargo, la presencia de numerosos iones inorgánicos en el agua de producción (especialmente Ca^{2+}) puede cambiar las propiedades de corrosión y precipitación, formando películas de productos de corrosión complejas que hacen que la corrosión por CO_2 sea más difícil de predecir [27]. Estas limitaciones resaltan la importancia de desarrollar estrategias para mitigar la corrosión en ambientes ricos en CO_2 y complejos en iones.

En este sentido, el uso de inhibidores de corrosión es crucial para prolongar la vida útil de los equipos en el campo, al mitigar la corrosión por CO_2 [28]. Se han reportado numerosos inhibidores de corrosión por CO_2, entre los que se incluyen los derivados de imidazolina, las bases de Schiff y las sales de amonio cuaternario [29][30][31]. Sin embargo, estos compuestos orgánicos suelen exhibir una alta toxicidad, una biodegradabilidad deficiente y procedimientos de síntesis complejos en la industria petroquímica. Esto genera un impacto ambiental negativo y limitan su aplicación [32]. Por lo tanto, el desarrollo de puntos cuánticos de carbono (CQDs) se considera una alternativa prometedora como inhibidores de corrosión ecológicos, debido a sus excelentes propiedades, como alta solubilidad en agua, buena biocompatibilidad, baja toxicidad y características químicas estables. Estos nanomateriales fluorescentes tienen aplicaciones potenciales en el control de la corrosión, la fotocatálisis, la electrocatálisis y la administración de fármacos [33][34][35].

Hasta ahora, los puntos cuánticos de carbono (CQD) han mostrado notables efectos de inhibición para varios metales en soluciones corrosivas de CO_2. Sin embargo, los mecanismos de corrosión que se producen en la interfaz entre el metal y el electrolito en presencia de CO_2 aún están bajo investigación, debido a la innumerable cantidad de variables que están involucradas [19,20]. Por lo tanto, predecir las tasas de corrosión en ambientes complejos de salmuera y CO_2, así como la eficiencia de inhibición de estos nanomateriales, sigue siendo una tarea desafiante. Las estrategias confiables de prevención de la corrosión y la predicción precisa de las tasas de corrosión son esenciales para la estimación futura del estado de las tuberías y para garantizar un mantenimiento eficiente y la operación segura de los equipos [36][37].

Los modelos tradicionales para predecir la corrosión por CO_2 incluyen modelos empíricos, semi-empíricos y mecanicistas. Los modelos de observación se basan completamente en datos de laboratorio y de campo, por lo que se han propuesto correlaciones empíricas a partir de la regresión no lineal de los datos experimental. Los modelos mecanicistas utilizan principios físicos básicos; sin embargo, estos modelos están relacionados con las dependencias de las correlaciones empíricas [38][39][40][41]. Si bien estos modelos han contribuido a la comprensión de los procesos de corrosión, presentan limitaciones significativas en el manejo de relaciones no lineales complejas, la adaptabilidad, la dependencia de mecanismos físicos y el rendimiento predictivo cuando se consideran múltiples factores influyentes [37].

En contraste, los recientes avances en inteligencia artificial y tecnología de la información han posibilitado la aplicación de técnicas de aprendizaje automático para resolver problemas relacionados con la corrosión. Los algoritmos de aprendizaje automático (ML) ofrecen sólidas capacidades de modelado no lineal, procesamiento de datos a gran escala y adaptabilidad, sin requerir mecanismos físicos explícitos, lo que se traduce en un rendimiento predictivo superior. En la última década, el aprendizaje automático se ha aplicado a la industria para el análisis automatizado de imágenes para la detección de corrosión**,** la predicción del crecimiento de defectos de corrosión en tuberías [42], la inspección de materiales [43], la estimación de la tasa de corrosión en ambientes marinos [44], y la predicción de espectros de impedancia electroquímica. Estas ventajas superan de manera efectiva las deficiencias de los modelos tradicionales, lo que mejora tanto la precisión como la aplicabilidad de la predicción de la corrosión por CO_2 [45].

En este trabajo, se emplea el aprendizaje automático para predecir las tasas de corrosión en ambientes salinos complejos, incorporando el efecto de los iones Ca^{2+}, que no se había considerado en modelos predictivos anteriores. Adicionalmente, se evalúa la eficiencia de inhibición de los puntos cuánticos de carbono bajo estas condiciones. La novedad de este estudio reside en la integración de resultados experimentales con enfoques avanzados de aprendizaje automático. Esto permite capturar la influencia combinada de los iones Ca^{2+} y los puntos cuánticos de carbono, mejorando así la predicción de las tasa de corrosión y contribuyendo al desarrollo de estrategias de mitigación más precisas y sostenibles.

1. Planteamiento del problema

La corrosión por CO_2 en salmueras carbonatadas es un proceso complejo influenciado por factores como el contenido de CO_2 gaseoso, la temperatura, el pH y la composición iónica de la solución [46][47]. Las salmueras de formación pueden variar desde soluciones simples hasta mezclas complejas que contienen diversas especies de carbonato y iones divalentes como el Ca^{2+} en altas concentraciones, junto con sales monovalentes dominantes como el NaCl [48][49]. Típicamente, la capa de productos de corrosión que se forma en las superficies de acero en medios acuosos saturados con CO_2 y NaCl es carbonato de hierro ($FeCO_3$) [50].

La formación de capas de siderita actúa como una película protectora en la superficie, bloqueando los puntos de disolución activa y creando una barrera para la difusión de especies electroquímicamente activas. Con esto, se logra una reducción significativa de la tasa de corrosión [51]. Sin embargo, la presencia de iones Ca^{2+} puede reaccionar con el CO_2 y facilitar la precipitación de productos de corrosión de tipo carbonato ($CaCO_3$ y $Fe_x Ca_yCO_3$). Esto altera la capacidad protectora de las capas de carbonato de hierro ($FeCO_3$) y afecta la estabilidad de los productos de corrosión ya formados [52] [45], haciendo la corrosión por CO_2 aún más complicada.

Curiosamente, a pesar de la presencia común de iones de Ca^{2+} en las salmueras de formación, no se han realizado estudios sobre nanomateriales para la inhibición de la corrosión por CO_2 en estas condiciones. En este contexto, la inclusión de puntos cuánticos de carbono (CQD) podría actuar como un inhibidor de la corrosión y, debido a sus propiedades superficiales, formar interacciones quelantes con iones divalentes como el Ca2+ en salmueras carbonatadas. Esto influiría en la formación de subproductos de tipo carbonato que afectan la corrosión por CO_2. Aunque se han desarrollado e implementado modelos para predecir la corrosión en ausencia de inhibidores, se ha logrado un éxito limitado al incorporar los efectos de los inhibidores de corrosión

Esto se debe a que la eficacia de los inhibidores depende de una amplia gama de factores operativos, incluidos el pH, la temperatura, la química de la salmuera y las condiciones de flujo [53][54][55]. Por lo tanto, las técnicas de aprendizaje de máquinas ofrecen

herramientas de gran valor debido a su capacidad para manejar modelos no lineales, a su rendimiento predictivo superior y a su habilidad para considerar de manera integral múltiples factores influyentes

1.1 Justificación

A pesar de los avances en la predicción de la corrosión por CO_2 y el desarrollo de inhibidores eficaces, hasta la fecha no existen estudios que hayan modelado la tasa de corrosión en la presencia simultánea de iones de Ca^{2+} y puntos cuánticos de carbono (CQD) como inhibidores. La presencia de Ca^{2+} en salmueras carbonatadas puede alterar significativamente la estabilidad y eficacia de los productos de corrosión, modificar la dinámica de la formación de la capa protectora y afectar el rendimiento del inhibidor. Sin embargo, los modelos predictivos aún no han integrado estos efectos, lo que limita su aplicabilidad a entornos del mundo real.

En este trabajo, se implementan modelos de aprendizaje de máquinas para predecir las tasas de corrosión bajo diversas condiciones experimentales. Estos modelos abordan de manera explícita el desafío de capturar los efectos de los iones divalentes y de los CQD como inhibidores. Al evaluar el papel de los iones Ca^{2+} en la química de la salmuera y el desempeño inhibidor de los puntos cuánticos de carbono (CQD), este estudio contribuye a una comprensión más realista de los procesos de corrosión y apoya el desarrollo de estrategias de mitigación efectivas en entornos de producción de petróleo y gas.

1.2 Objetivo General

Desarrollar un modelo de aprendizaje de máquinas para la predicción de la tasa de corrosión del acero al carbono en salmueras carbonatadas en presencia de iones de calcio (Ca^{2+}) y puntos cuánticos de carbono (CQD), considerando el efecto de variables como fuerza iónica, la salinidad, el tiempo de inmersión, la temperatura y la concentración del inhibidor.

1.2.1 Objetivos Específicos

- **Evaluar** el efecto de los iones de calcio divalentes (Ca^{2+}), así como de la fuerza iónica, la salinidad, el pH, la temperatura y el tiempo de inmersión en la tasa de corrosión.
- **Evaluar** la eficiencia de inhibición de los puntos cuánticos de carbono (CQD) en salmueras carbonatadas bajo diferentes condiciones experimentales.
- **Seleccionar** las variables relevantes que influyen en las tasas de corrosión para la construcción de modelos de aprendizaje automático, garantizando su representatividad y calidad.
- **Comparar** modelos de aprendizaje automático para la predicción de tasas de corrosión en ambientes con CO_2.
- **Optimizar** los modelos de aprendizaje automático supervisado y evaluar sus métricas de rendimiento.

2. Marco teórico

2.1 Aprendizaje de máquinas

Las técnicas de aprendizaje automático se refieren a un conjunto de algoritmos que aprenden a realizar tareas, como predecir resultados experimentales, al ser entrenados con datos de experimentos realizados previamente. En particular, el aprendizaje supervisado implica entrenar un modelo utilizando datos etiquetados. En este proceso, el algoritmo aprende a mapear los datos de entrada con sus etiquetas correspondientes, lo que permite realizar predicciones precisas basándose en patrones observados anteriormente [54][55].

2.1.1 Modelos de aprendizaje supervisado

En el aprendizaje supervisado (SL), el algoritmo tiene la capacidad de comparar la salida aprendida con la salida real y, a partir de esa comparación, ajustar las salidas futuras. Por lo tanto, para seleccionar la variable de salida, el algoritmo debe aprender las propiedades importantes dentro de cada punto de datos del conjunto y, al mismo tiempo, ser capaz de predecir las variables de salida.

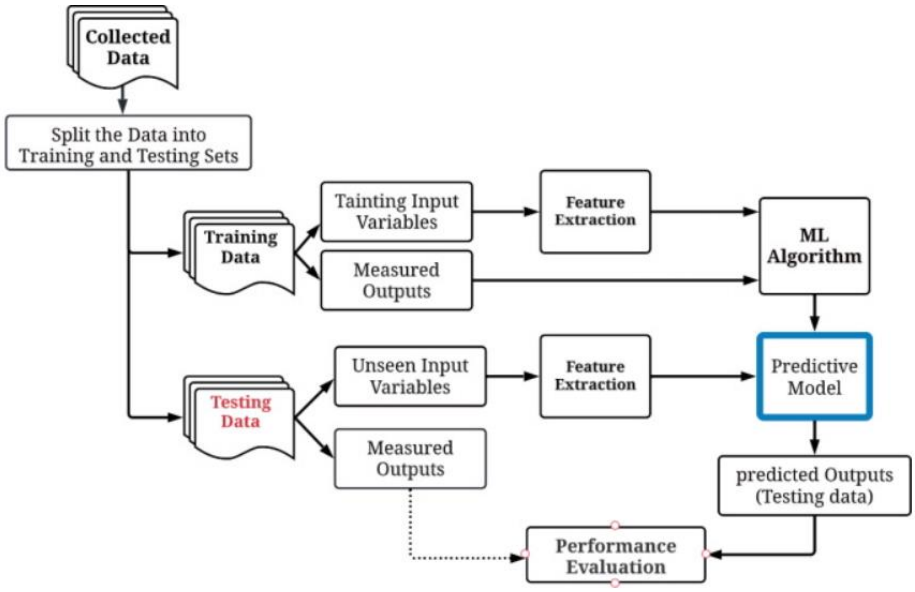


Figura 1. Modelo general de aprendizaje supervisado [56]

Las técnicas de aprendizaje supervisado (SL) se pueden clasificar en dos algoritmos principales, dependiendo del tipo de variable de salida:

- Algoritmo de regresión: para variables continuas.
- Algoritmo de clasificación: para variables discretas.

Considerando la naturaleza continua de la variable independiente (tasa de corrosión), en este estudio modelos de aprendizaje de máquinas supervisado comunes en estudios de corrosión como Supported Vector machines Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN) y Multi-layer Perceptron (MLP) se aplicarán a la data preprocesada.

2.1.1.1 Random Forest (RF)

Random Forest es un algoritmo de aprendizaje en conjunto (o "ensemble learning") construido sobre los principios de los árboles de decisión. Un árbol de decisión divide un conjunto de datos en subconjuntos progresivamente más pequeños basándose en variables de entrada, con el objetivo de reducir la varianza en los valores de los resultados en cada paso. Esta estructura jerárquica consiste en ramas que representan reglas de decisión y hojas que contienen las predicciones de los resultados. El proceso de división continúa hasta que se cumpla una condición predefinida, como un número máximo de divisiones o un umbral de varianza [57].

RF mejora el poder predictivo de los árboles de decisión al construir múltiples árboles usando un muestreo de Bootstrap sampling de los datos de entrenamiento. Cada árbol se entrena en un subconjunto diferente de los datos, y la predicción final se obtiene a través de un voto mayoritario (para clasificación) o un promedio (para regresión). Este enfoque reduce el sobreajuste y mejora la generalización del modelo. La efectividad de RF reside en su capacidad para combinar múltiples árboles de decisión, aprovechando sus fortalezas individuales mientras se minimizan las debilidades a través de técnicas de aprendizaje en conjunto [58].

2.1.1.2 Extreme Gradient Boosting (XGBoost)

XGBoost es un algoritmo de *boosting* basado en árboles de regresión [59]. La idea central de XGBoost es agregar árboles continuamente para la división de características, donde cada árbol añadido esencialmente aprende una nueva función para ajustar los residuos de la predicción anterior. Una vez que se completa el entrenamiento y se obtienen *k* árboles, la puntuación de una muestra se predice basándose en los nodos hoja específicos de las características en cada árbol. Finalmente, al sumar las puntuaciones de todos los árboles, se obtiene la predicción para la muestra.

2.1.1.3 Support Vector Regression (SVR)

SVR está diseñado para manejar relaciones no lineales entre variables de entrada y etiquetas de destino. A diferencia de los modelos de regresión tradicionales, que asumen una relación lineal, SVR transforma los datos de entrada en un espacio de mayor dimensión donde la relación se puede capturar de manera más efectiva como una función lineal. Para lograr esta transformación, SVR emplea funciones *kernel*, que mapean los datos originales a un espacio de características de mayor dimensión sin necesidad de un cálculo explícito en ese espacio, lo que reduce la complejidad computacional.

En este espacio transformado, SVR ajusta un hiperplano que mejor representa los datos mientras mantiene un margen de tolerancia, lo que permite que el modelo se generalice bien a nuevas entradas. Esta capacidad para capturar patrones complejos hace de SVR una herramienta poderosa para tareas de regresión donde los modelos lineales tradicionales se quedan cortos [60].

2.1.1.4 K-Nearest Neighbors (KNN)

KNN es un algoritmo sencillo pero efectivo que predice el resultado de un punto de datos basándose en sus *K* vecinos más cercanos en el conjunto de entrenamiento. Los vecinos más cercanos se identifican utilizando una métrica de distancia, como la distancia euclidiana para variables numéricas o la distancia de Hamming para variables categóricas. La predicción final se obtiene promediando las etiquetas de los K puntos más cercanos (para regresión) o mediante voto mayoritario (para clasificación). KNN es no paramétrico y

se basa en la suposición de que los puntos de datos similares tienen resultados similares [61][62].

2.1.1.5 Multi-layer Perceptron (MLP)

El Perceptrón Multicapa (MLP) es un tipo de red neuronal artificial utilizada en el aprendizaje automático tanto para tareas de clasificación como de regresión. Se le llama "multicapa" porque tiene múltiples capas de nodos (neuronas), lo que le permite aprender y modelar relaciones complejas en los datos. En la **Figura 2** se muestra la estructura del MLP, incluye una capa de entrada, una capa oculta y una capa de salida.

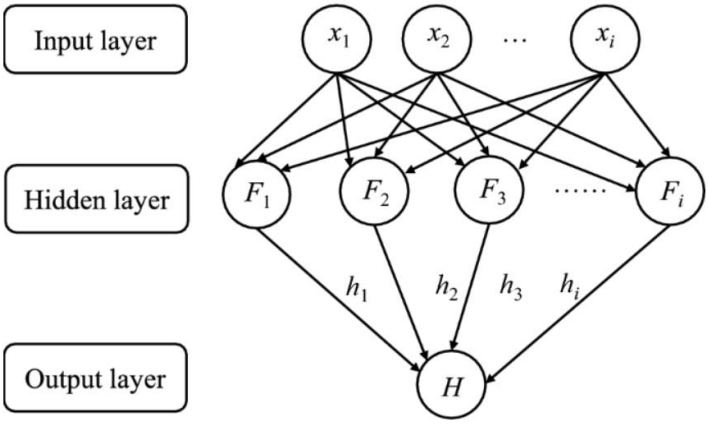


Figura 2. Arquitectura de MLP [63]

Cada variable de entrada (x_i) se conecta a un nodo en la siguiente capa (es decir, la primera capa oculta), y a cada conexión se le asigna un peso (ω_{ij}). El resultado obtenido (F_i) se introduce en la primera capa oculta al ponderar y sumar x_i. Cada capa transforma la entrada de alguna manera, lo que permite a la red aprender patrones intrincados. Finalmente, la suma de los resultados después de la ponderación se transforma a través de una función de activación para obtener la salida de la primera capa oculta (h_i), como se expresa en la Ecuación (1):

$$h_i = \frac{1}{1+ e^{-F_i}} \qquad (1)$$

donde F_i es el resultado de la suma ponderada de x_i, y h_i es la salida de la primera capa oculta.

3. Estado del arte

La revisión sistemática de literatura fue llevada a cabo para establecer los avances y estudios enfocados a la predicción de tasas de corrosión por CO_2 en salmuera con iones divalentes, presencia de CQDs, e implementación de algoritmos de aprendizaje supervisado.

El efecto de los iones de calcio en las propiedades de corrosión por CO_2 ha sido evaluado por varios autores. Entre ellos, Ding *et al*. evaluaron el impacto de estos iones en salmuera simulada con diferentes concentraciones de Ca^{2+}, encontrando que la tasa de corrosión aumenta con concentraciones más altas de iones de calcio. También informaron que a medida que aumenta la concentración de Ca^{2+}, las capas de corrosión se vuelven más sueltas y el tamaño de grano más grande, lo que reduce la resistencia y la adhesión de las capas a la matriz de acero [64].

Esmaeely *et al*. investigaron el efecto del calcio en la corrosión por CO_2 en soluciones al 1% wt NaCl a 80°C y pH 6.6. Evaluaron diferentes concentraciones de Ca^{2+} (10, 100, 1,000 y 10,000 ppm), y entre los resultados encontrados mostraron que a concentraciones bajas de Ca^{2+} (10 ppm y 100 ppm), la tasa de corrosión disminuyó con el tiempo como resultado de la formación de carbonato de hierro protector ($FeCO_3$) y/o carbonato mixto ($Fe_xCa_yCO_3$), sin embargo, la presencia de altas concentraciones de Ca^{2+} (1,000 ppm y 10,000 ppm) resultó en el cambio del producto de corrosión de $FeCO_3$ protector a carbonato de calcio ($CaCO_3$) no protector aumentando las tasas de corrosión con el tiempo [65].

Luego, Shamsa *et al*. estudiaron el papel de los iones de Ca^{2+} en los productos de corrosión Ca/Fe en entornos corrosivos por CO_2 a 80° y 150°C. Encontraron que, a 80°C, la adición de Ca^{2+} aumentó la tasa de corrosión y resultó en una capa de $Fe_xCa_yCO_3$ más porosa, mientras que, a 150°C la capa de carbonato mixto fue tan protectora como el $FeCO_3$ puro, reduciendo la corrosión general y por picaduras [66].

Por su parte, Rizzo *et al*, investigaron el efecto de Ca^{2+} en la estabilidad y protección de capas de $FeCO_3$ ya precipitadas. Las pruebas se realizaron a presión atmosférica en una solución al 1%wt NaCl saturada con CO_2 a 80°C. Se estudiaron tres concentraciones diferentes de Ca^{2+} (100, 1,000 y 10,000 ppm). Los resultados sugirieron que la precipitación

de $CaCO_3$ llevó a una solución subsaturada con respecto al $FeCO_3$, seguida de la disolución de la capa protectora [32].

Finalmente, Ren *et al.*[67] investigaron la influencia del Ca^{2+} en los productos de corrosión en soluciones de NaCl al 3.5% en peso saturadas con CO_2. Descubrieron que el aumento de la concentración de Ca^{2+} conducía a la formación de estructuras de $Fe_xCa_yCO_3$ menos cristalinas y más amorfas, las cuales eran menos protectoras y resultaban en tasas de corrosión más altas. Además, una mayor cantidad de iones Ca^{2+} reemplazó las posiciones de Fe^{2+} en la estructura cristalina, lo que demostró claramente el efecto del calcio en los mecanismos de corrosión por CO_2.

Por su parte, la inclusión de CQDs en el campo de la corrosión ha ganado atención durante los últimos años. Li *et al.* sintetizaron puntos cuánticos de carbono dopados con nitrógeno (N-CQDs), y evaluaron el mecanismo inhibitorio de los CQDs sobre acero N80 en una solución de NaCl al 3% en peso saturada con CO_2 mediante pérdida de peso y medidas electroquímicas. Los resultados muestran que la corrosión del acero al carbono N80 disminuye significativamente, alcanzando eficiencias de inhibición de 84% con la adición de 600 mg L^{-1} de N-CQDs, lo cual se atribuye a la formación de una película de adsorción que previene el contacto del hierro con los iones de cloruro y actúa como un inhibidor de corrosión efectivo [68].

Cen *et al.* sintetizaron puntos de carbono co-dopados con nitrógeno y azufre (N,S-CDs) como inhibidores de corrosión en soluciones de NaCl al 3.5% en peso saturadas con CO_2 Descubrieron que el efecto inhibidor estaba relacionado con la quimisorción de los grupos funcionales en la superficie de los N,S-CDs sobre el hierro, lo que contribuye a la formación de una película protectora, logrando eficiencias de inhibición del 93% a una concentración de 50 mg L^{-1} de CQDs [69].

Más recientemente, Wu *et al.* sintetizaron CQDs a partir de salicilato de sodio y urea para evaluar su efectividad como inhibidores de corrosión para acero N80 en soluciones a 70°C de HCl 1M y en soluciones al 3.5 wt% NaCl saturadas con CO_2. Los resultados de su estudio mostraron que los N-CDs se adhirieron gradualmente a la superficie del acero a través de adsorción física y química para formar una película que inhibe la corrosión, la cual aisló los iones corrosivos del contacto con el acero al carbono. Alcanzaron eficiencias de inhibición de 94.6% a concentraciones de 200 mg L^{-1} de N-CQDs[46].

Además, el uso de algoritmos de aprendizaje automático para predecir tasas de corrosión en ambientes con CO_2 ha sido explorado.

Aghaaminiha *et al*. evaluaron modelos de aprendizaje automático supervisado para predecir la tasa de corrosión del acero al carbono a lo largo del tiempo en presencia de inhibidores de corrosión. Su conjunto de datos incluyó concentraciones variables y esquemas de dosificación de inhibidores. Probaron cuatro algoritmos: Random Forest (RF), K-Nearest Neighbors (KNN), Redes Neuronales Artificiales (ANN) y Regresión de Vectores de Soporte (SVR).

Encontraron que el modelo de Random Forest predijo con precisión la evolución de la corrosión y su sensibilidad a las condiciones ambientales como la temperatura, el pH, la concentración de CO_2 y el tipo de inhibidor. RF mostró el mejor rendimiento en la predicción del perfil de corrosión temporal completo, con un error cuadrático medio (MSE) que osciló entre 0.005 y 0.093 [70].

Similarmente, Dong *et al*. desarrollaron un modelo predictivo para estimar la tasa y la severidad de la corrosión por CO_2 utilizando algoritmos de aprendizaje automático. Su conjunto de datos consistió en 248 muestras con 10 características, incluyendo el material, el contenido de cromo, el contenido de agua, el contenido de oxígeno, el contenido de SO_2 y NO_2, la concentración de H2S, la presión, la temperatura y el tiempo. Evaluaron seis modelos: RF, KNN, Gradient Boosting Decision Tree (GBDT), Support Vector Machine (SVM), XGBoost y LightGBM.

SVM tuvo el rendimiento más débil, mientras que RF logró los mejores resultados, con valores de R^2 de 0.92 (entrenamiento) y 0.88 (prueba). Al clasificar la severidad de la corrosión, el RF también superó a los demás, alcanzando una precisión del 99% y una puntuación F1 de 0.99. Este estudio destaca el gran potencial de los algoritmos de aprendizaje automático, especialmente Random Forest, para predecir y clasificar la corrosión por CO_2 [45].

4. Metodología

En este capítulo se aborda la metodología propuesta para la evaluación de las tasas de corrosión bajo diferentes condiciones experimentales, así como la conformación del conjunto de datos para la implementación de los diferentes modelos de aprendizaje de máquinas vistos anteriormente y, por último, las métricas de evaluación del desempeño de los modelos predictivos.

4.1 Evaluación de corrosión

Para la evaluación de la inhibición de corrosión del acero al carbono en soluciones saturadas de CO_2, inicialmente se prepararon soluciones de NaCl y $CaCl_2$ a diferentes salinidades (5000 - 30000 mg L^{-1}), para evaluar el efecto del tipo de ion dentro de la tasa de corrosión. A cada solución preparada, CO_2 fue burbujeado continuamente durante 30 minutos para asegurar la saturación antes de cada experimento.

Luego, para establecer el efecto inhibitorio de los CQDs, se prepararon adicionalmente soluciones con presencia de CQDs, a diferentes concentraciones y se llevaron a la temperatura designada durante diferentes tiempos de inmersión. Al final del tiempo de inmersión, los productos de corrosión fueron retirados de la superficie del acero usando una solución de 5% HCl, luego fueron lavados con agua destilada, y secados con aire.

Todas las muestras fueron pesadas antes del experimento y luego de la limpieza, para asegurar la remoción de los productos de corrosión y así determinar las tasas de corrosión de los sistemas evaluados. Las tasas de corrosión (CR), fueron calculados usando la Ecuación (2)

$$CR = \frac{87600\,\Delta W}{\rho S t} \qquad (2)$$

Dónde, ΔW representa la pérdida de peso promedio de las muestras [g], ρ es la densidad de las muestras de acero al carbono (7.79 g cm^{-3}), S es el área superficial expuesta (cm^2), and t representa el tiempo de inmersión (h).

Un diseño de experimentos de mezclas Simplex-centroide fue desarrollado para modelar las tasas de corrosión. El diseño de mezclas simplex-centroide (SCMD) fue corrido usando el software STATGRAPHICS Centurion XVI (StartPoint Technologies Inc., Addison, TX, USA), variando la composición másica de las sales.

Las proporciones de cada componente de la soluciones preparadas debe satisfacer las siguientes restricciones de la Ecuación (3):

$$\sum_{i=1}^{\sigma} x_i = x_1 + x_2 + x_3 + \cdots + x_q = 1 \qquad x_i \geq 0 \qquad (3)$$

Dónde, el parámetro σ se refiere al número de componentes variando en la mezcla y x_i es la proporción de cada componente. Para este estudio, cada diseño de mezcla se tiene $\sigma = 3$, x_1 = NaCl, x_2= $CaCl_2$ y x_3 = H_2O. De acuerdo a la composición máxima de cada solución expresada anteriormente, los límites de cada compuesto son:

$0.97 \leq H_2O \leq 1.00$ (4)

$0 \leq CaCl_2 \leq 0.03$ (5)

$0 \leq NaCl \leq 0.03$ (6)

Para el ajuste de los datos experimentales a modelos cúbicos, las ecuaciones de regresión del modelo son:

$$\alpha_m = \beta_1 x_1' + \beta_2 x_2' + \beta_3 x_3' + \beta_{12} x_1' x_2' + \beta_{13} x_1' x_3' + \beta_{123} x_1' x_2' x_3' \qquad (7)$$

$$x_i' = \frac{x_i - L_i}{1 - L} \qquad (8)$$

Dónde α_m es la tasa de corrosión, y β_i, β_{ij}, β_{ijk} son los coeficientes de los términos lineales, componentes binarios y terciarios de la mezcla, respectivamente. En la Ecuación (8), L_i es el límite inferior de cada componente, *L* es la suma de los limites inferiores, y x_i' es el pseudo-componente de x_i, el cual es usado por las restricciones de las Ecuaciones (4)-(6).

En la **Tabla 1** se presenta los puntos experimentales a evaluar en SCDM.

Tabla 1. Diseño de mezclas Simplex-Centroide

Test	x_1	x_2	x_3
1	0.015	0	0.985
2	0	0	1
3	0	0.015	0.985
4	0.01	0.01	0.98
5	0	0.03	0.97
6	0.015	0.015	0.97
7	0.03	0	0.97

4.2 Conformación del conjunto de datos experimentales

Para la conformación del dataset usado en la implementación de modelos de aprendizaje de máquinas orientados a la predicción de tasas de corrosión, se utilizaron los valores calculados a partir de las pérdidas de peso obtenidas en los puntos experimentales generados mediante SCDM, junto con otras valoraciones complementarias.

En la **Tabla 2** se presenta la matriz experimental, donde se detallan los parámetros y los rangos dentro de los cuales se realizan los experimentos gravimétricos. Una vez estructurado el dataset, este fue cargado y sometido a un preprocesamiento de datos empleando Python 3.12.2 y librerías de libre distribución como *pandas* para la gestión de datos, *numpy* para operaciones numéricas, *skcit-learn* para el procesamiento y modelado, y *seaborn* para la exploración y visualización.

Tabla 2. Variables operaciones consideradas para la predicción de tasas de corrosión

Descripción	Rango	Unidades	Tipo
Brine type	{NaCl, NaCl+$CaCl_2$ $CaCl_2$}	-	Categorica
Time	[24-120]	hora	Numerica
Salinity	[5000 – 30000]	mg L^{-1}	Numerica
CQD	[0 – 500]	mg L^{-1}	Numerica
Temperature	[40 - 70]	°C	Numerica
pH	[3.3 – 4.5]	-	Numerica
Fuerza iónica	[0.221 – 0.811]	M	Numerica

4.2.1 Preprocesamiento y análisis exploratorio

Una vez consolidado el conjunto de datos experimental, se realizó un preprocesamiento de los datos con el fin de garantizar la calidad antes de aplicar los algoritmos de aprendizaje automático. Este proceso incluyó la identificación y depuración de valores atípicos o inconsistentes, así como el tratamiento de datos faltantes.

Para las variables categóricas se aplicó la técnica *one-hot encoding*, lo que permitió representarlas de forma numérica sin introducir un orden artificial entre sus categorías. En el caso de las variables numéricas, se implementaron técnicas de escalamiento con *MinMax Scaler* o *Standard Scaler*, seleccionadas de acuerdo con la naturaleza de cada variable y los requerimientos del modelo, con el fin de evitar sesgos ocasionados por diferencias en las escalas de medida.

Posteriormente, el conjunto de datos fue dividido en un 80% para entrenamiento y un 20% para prueba, lo que proporcionó una primera estimación objetiva del desempeño de los modelos sobre datos no utilizados en la fase de ajuste. Con el fin de fortalecer la evaluación, se empleó adicionalmente el método de validación cruzada *k-fold*, el cual permite obtener un estimador más robusto de la capacidad predictiva del modelo al promediar el rendimiento obtenido en diferentes particiones del conjunto de datos, reduciendo así la varianza asociada a una única división y mitigando el riesgo de sobreajuste.

Se lleva a cabo un análisis exploratorio de datos (EDA) para identificar las distribuciones y relaciones entre las variables. Se realiza un análisis descriptivo estadístico con medias, medianas y desviación estándar para interpretar tendencias de los datos. Así mismo se construye la matriz de correlación de Pearson [71], que permite evaluar la correlación entre dos variables *X* y *Y*, como se presenta en la Ecuación (9):

$$r\,(X,Y) = \frac{Cov\,(X,Y)}{\sqrt{Var(X)}\,.\sqrt{Var(Y)}} \qquad (9)$$

Dónde, $Cov\,(X,Y)$ representa la covarianza de *X* y *Y*, $Var(X)$ indica la varianza de *X*, e $Var(Y)$ indica la varianza de *Y*. El coeficiente de correlación calculado varía entre −1 y 1. Si el coeficiente de correlación entre dos factores que influyen en la corrosión es 0, significa que las dos variables son relativamente independientes y no presentan correlación.

Cuando el coeficiente se encuentra entre 0 y 1, indica una correlación positiva, mientras que si se encuentra entre −1- y 0, indica una correlación negativa. A mayor valor absoluto del coeficiente de correlación, más fuerte será la relación entre los parámetros de corrosión.

4.3 Modelos de aprendizaje de máquinas para la predicción de corrosión en salmueras con iones divalentes y CQDs

En esta sección se presenta la selección de modelos implementados, las métricas de error evaluadas y la selección de los hiperparámetros para la validación de los modelos.

Para predecir los valores de corrosión en función de las condiciones propuestas se implementaron diferentes modelos de aprendizaje. Los modelos predictivos fueron RF, SVR, KNN, XGBoost y MLP. Cada uno de los modelos implementados fueron sometidos a una optimización y se lleva a cabo una búsqueda de hiperparámetros utilizando el método *GridSearchCV* de la librería *sklearn*. Los parámetros y rangos de variación se presentan en la **Tabla 3**:

Tabla 3. Búsqueda de Hiperparámetros para los modelos implementados en la predicción de tasas de corrosión

Modelo	Hiperparámetros	Valores
RF	criterion	{absolute_error, squared_error}
	max_depth	[10, 20, None]
	max_features	{1, sqrt, log2}
	max_samples	[0.7 - 0.9]
	min_samples_leaf	[1 – 4]
	min_samples_split	[2 – 10]
	n_estimators	[100 - 300]
XGBoost	colsample_bytree	[0.7 - 0.9]
	learning_rate	[0.01 - 0.3]
	max_depth	[3 – 9]
	min_child_weight	[1 – 5]
	n_estimators	[100 - 300]
	subsample	[0.7 - 0.9]
KNN	algorithm	{auto, ball_tree, kd_tree}
	leaf_size	[10 - 50]
	metric	{euclidean, manhattan, chebyshev}
	n_neighbors	[2 – 21]
	weights	{uniform, distance}

SVR	C	[0.1 – 100]
	epsilon	[0.01 - 0.2]
	gamma	{scale, auto}
	kernel	{sigmoid, linear, poly, rbf}
MLP	activation	{relu, tanh}
	alpha	[0.0001 – 0.1]
	hidden_layer_sizes	(50,), (100,0), (50,50), (100,50), (100,100)
	learning_rate	{constant, adaptative}
	max_iter	[1000 – 10000]
	solver	{sgd, adam}

4.3.1 Métricas de desempeño

Para evaluar el desempeño de los modelos predictivos, este estudio emplea el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE) como métricas principales, dado que representan la desviación de los valores predichos con los valores reales del conjunto de evaluación. La **Tabla 4** presenta las métricas, su descripción y ecuación usadas.

Tabla 4. Métricas implementadas en los modelos de predicción

Métrica	Descripción	Ecuación
R^2	Mide la bondad de ajuste, cuantificando la proporción de la varianza de la variable objetivo que es explicada por las variables independientes y_i es el valor real, f_i es el valor predicho y $\bar{y}$ es la media de los valores reales	$1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$ (10)
RMSE	Evalúa la desviación estándar de los errores de predicción n es el número de muestras, y_i es el valor real, $\hat{y}_i$ es el valor predicho	$\sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$ (11)
MAE	Mide el promedio de la magnitud de los errores en un conjunto de predicciones, sin tener en cuenta la dirección del error. y_i es el valor real, $\hat{y}_i$ es el valor predicho n es el número de muestras	$\frac{1}{n} \sum \lvert y_i - \hat{y}_i \rvert$ (12)

5. Resultados

En este capítulo se presentan los resultados asociados a la evaluación de corrosión en salmueras carbonatadas y la predicción de las tasas de corrosion a través de los modelos de aprendizaje de máquinas. En la primera sección se presentan los resultados asociados a la evaluación de las tasas de corrosión teniendo en cuenta el efecto de Ca^{2+} y la presencia o ausencia de CQDs bajo diferentes condiciones experimentales. Seguido de esto, se muestra el análisis exploratorio del conjunto de datos y el análisis de correlación para seleccionar las variables relevantes en la implementación de los modelos de aprendizaje automático. Por último, se presentan los resultados de las predicciones de las tasa de corrosion, la optimización y evaluación de desempeño de los modelos seleccionados.

5.1 Medidas gravimétricas

5.1.1 Efecto de iones: Na^{+} vs Ca^{2+}

Se evaluó la pérdida de peso de muestras de acero al carbono en soluciones saturadas con CO_2 a 70 °C, empleando diferentes concentraciones de salmuera de NaCl y $CaCl_2$. Los resultados obtenidos, presentados en la **Figura 3**, evidencian que la tasa de corrosión en presencia de $CaCl_2$ es consistentemente superior en comparación con soluciones de NaCl.

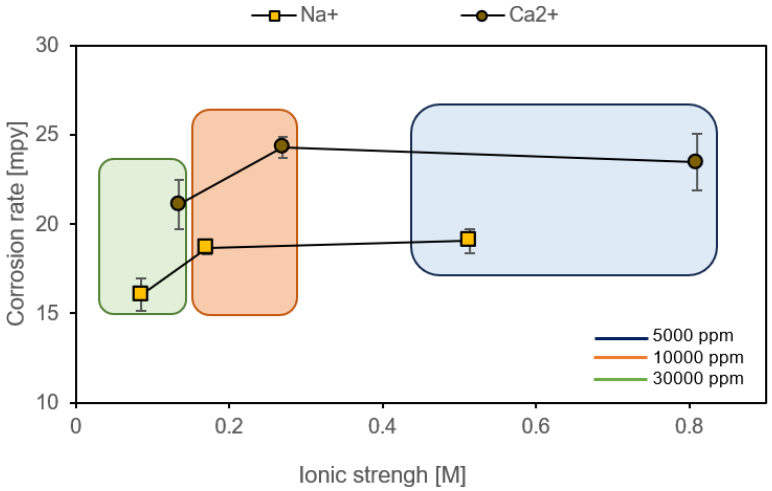


Figura 3. Efecto del tipo de ion (Na^{+} vs Ca^{2+}) en la tasa de corrosión de acero al carbono

Estos resultados pueden analizarse a partir de la contribución del tipo de ion a la fuerza iónica. A igual salinidad, los cationes divalentes Ca^{2+} aportan significativamente más a la fuerza iónica que los monovalentes Na^{+} como se muestra en la **Figura 4**. Un incremento en la fuerza iónica reduce la solubilidad del CO_2 en la fase acuosa; en presencia de sales monovalentes como el NaCl, la solubilidad del CO_2 es mayor que en presencia de sales divalentes como $CaCl_2$ [72], lo que disminuiría la disponibilidad de ácido carbónico, reduciendo la agresividad del medio.

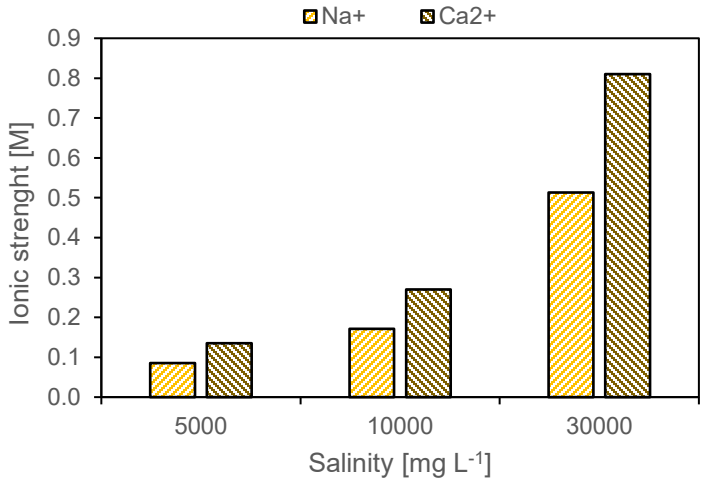


Figura 4. Influencia del tipo de ion en la fuerza iónica a diferentes salinidades

Sin embargo, los resultados indican que el efecto dominante no es por la solubilidad del CO_2, sino por la interacción específica del ion Ca^{2+} con los productos de corrosión. En medios acuosos de NaCl, la disolución anódica del acero consume hidrogeno (H^{+}) y libera Fe^{2+} a la solución, elevando el pH e incrementando la concentración de Fe^{2+} en la interfase electrolito-metal, esto lleva a una sobresaturación con respecto a $FeCO_3$ [73]. La precipitación de $FeCO_3$ genera capas protectoras densas y adherentes que limitan el transporte de especies corrosivas hacia la superficie del acero, bloqueando los sitios activos lo que resulta en una disminución de las tasas de corrosion [65][74].

En contraste, la presencia de iones Ca^{2+} no solo facilita la precipitación de $CaCO_3$, sino que también promueven la formación de carbonatos mixtos ($Fe_xCa_yCO_3$) cuando Fe^{2+} también está presente en la salmuera de producción [75], dado que Ca^{2+} puede sustituir a Fe^{2+} en la estructura de $FeCO_3$, debido a que la siderita y calcita comparten la misma estructura cristalina.

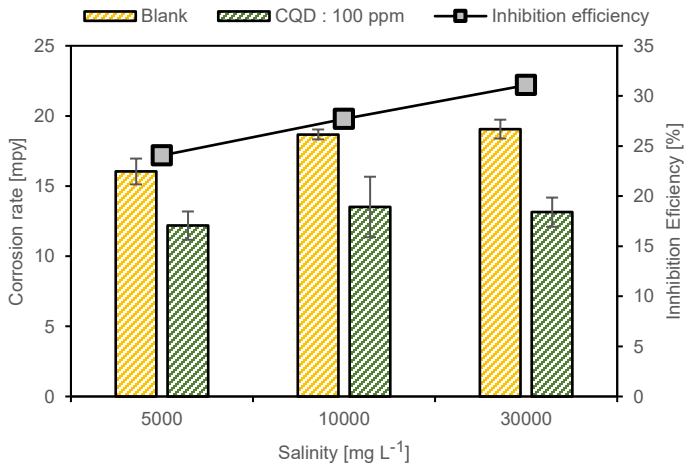
Esta sustitución da resultado a productos de corrosion con propiedades morfológicas y química completamente diferentes [74], dando lugar a capas protectoras más porosas, con granos de mayor tamaño y menor fuerza de adhesión, como consecuencia, las películas formadas en presencia de Ca^{2+} son menos protectoras y permiten la difusión de especies corrosivas, lo que se traduce en mayores tasas de corrosion a las observadas en soluciones con Na^{+} [64].

5.1.2 Efecto de CQDs

En la **Figura 5** se presenta las pruebas de pérdida de peso de acero al carbon en soluciones de NaCl y $CaCl_2$ a 70°C con concentración de 100 mg L^{-1} de CQD.

Tras 24 horas de inmersión se observó una reducción notable en la tasa de corrosión (CR) en ambas soluciones tras la adición de CQDs, con un efecto inhibidor más pronunciado en las soluciones con $CaCl_2$. Los resultados muestran que los CQDs mitigan eficazmente la corrosión, alcanzando una eficiencia máxima de inhibición de 45.9% en $CaCl_2$ y de 24.1% en NaCl a la menor concentración de salmuera (5000 mg/L). A la mayor salinidad evaluadas, las eficiencias de inhibición fueron de 34.5% y 31.1% para $CaCl_2$ y NaCl, respectivamente.

a)



b)

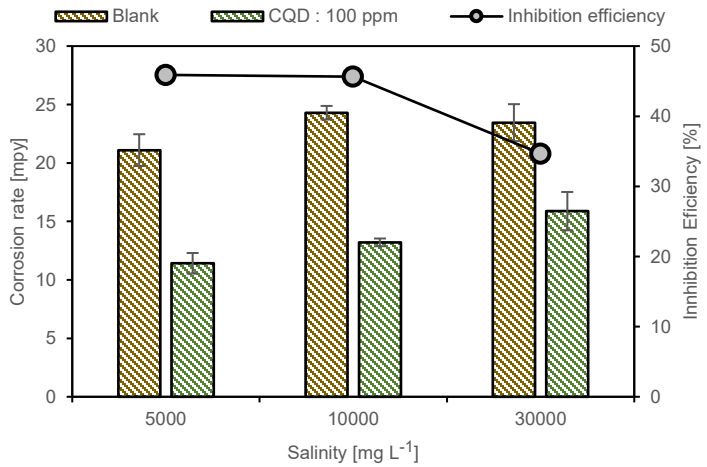


Figura 5. Tasas de corrosión de acero al carbón en soluciones con presencia y ausencia de CQDs **a)** NaCl y **b)** $CaCl_2$

Estos resultados sugieren que los CQDs pueden actuar como inhibidores de corrosión al interactuar con la superficie del acero, posiblemente mediante adsorción física y química formando de esta forma una capa protectora que limita la disolución del metal [76]. La inhibición más fuerte observada en soluciones de $CaCl_2$ podría estar relacionada con un efecto de quelación, en el cual los CQDs coordinan con los iones Ca^{2+}, modificando su influencia en la dinámica de la corrosión y mitigando su efecto sobre los productos de corrosión. Si bien no se han reportado explícitamente la interacción de CQDs con Ca^{2+} en salmueras carbonatadas, los grupos funcionales de los CQDs presentan alta afinidad por cationes metálicos, por lo que este posible mecanismo de quelación, más la adsorción superficial sobre el acero podría reforzar su efecto inhibidor [77][78].

Por otro lado, en la **Figura 6** se muestra la evolución de la tasa de corrosión y la eficiencia de inhibición en función de la concentración de CQDs para soluciones al 3% wt de NaCl y $CaCl_2$ a 70°C luego de 24 horas. En ambos sistemas, la tasa de corrosión disminuye con el incremento de la concentración de CQDs, alcanzando eficiencias de 36.5% y 48.7% para NaCl y $CaCl_2$ respectivamente, a la máxima concentración evaluada (500 mg L^{-1}).

Este comportamiento se puede asociar con la capacidad de los CQDs para adsorberse sobre la superficie metálica mediante interacciones de sus grupos funcionales que

contienen un par de electrones solitario (como N-H, N-O, N-C=O, etc.) combinados con el orbital *d* vacío de Fe^{2+} favoreciendo la formación de películas protectoras más compactas con el aumento en la concentración [79][69][80].

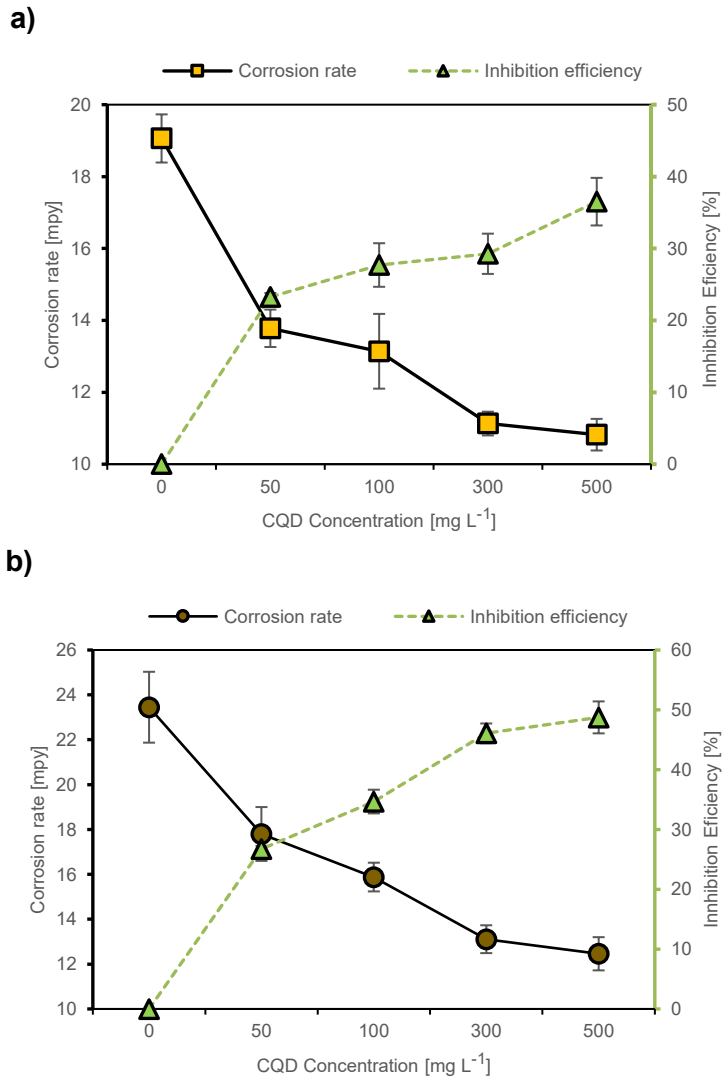


Figura 6. Tasa de corrosión y eficiencia de inhibición de acero al carbono en soluciones con presencia de varias concentraciones de CQDs **a)** NaCl y **b)** $CaCl_2$

La **Figura 7** muestra los resultados de tasa de corrosión y eficiencia de inhibición luego de 24 horas para las soluciones saturadas con CO_2 a diferentes temperaturas. En ambos medios se observa un ligero incremento en la CR al pasar de 40 a 70°C, lo cual se puede atribuir al aumento en la cinética de las reacciones electroquímicas y la aceleración de los procesos de disolución del acero a mayor temperatura.

Se puede observar que con la adición de CQDs, el grado de corrosion se inhibió cerca del 40 a 44% a la menor temperatura, no obstante, cuando la temperatura alcanzó los 70°C, la tasa de corrosión en presencia de CQDs también incrementó ligeramente y la eficiencia de inhibición disminuyó. Esto podría atribuirse a que con el incremento de la temperatura se reduce la capacidad de adsorción del CQDs sobre la superficie de acero al carbono [79]. A temperatura de 70°C, la eficiencia de inhibición en las soluciones de NaCl y $CaCl_2$ alcanzaron el 31.1% y 34.7%, respectivamente, lo que demuestra que los CQDs presentan cierta resistencia a la temperatura.

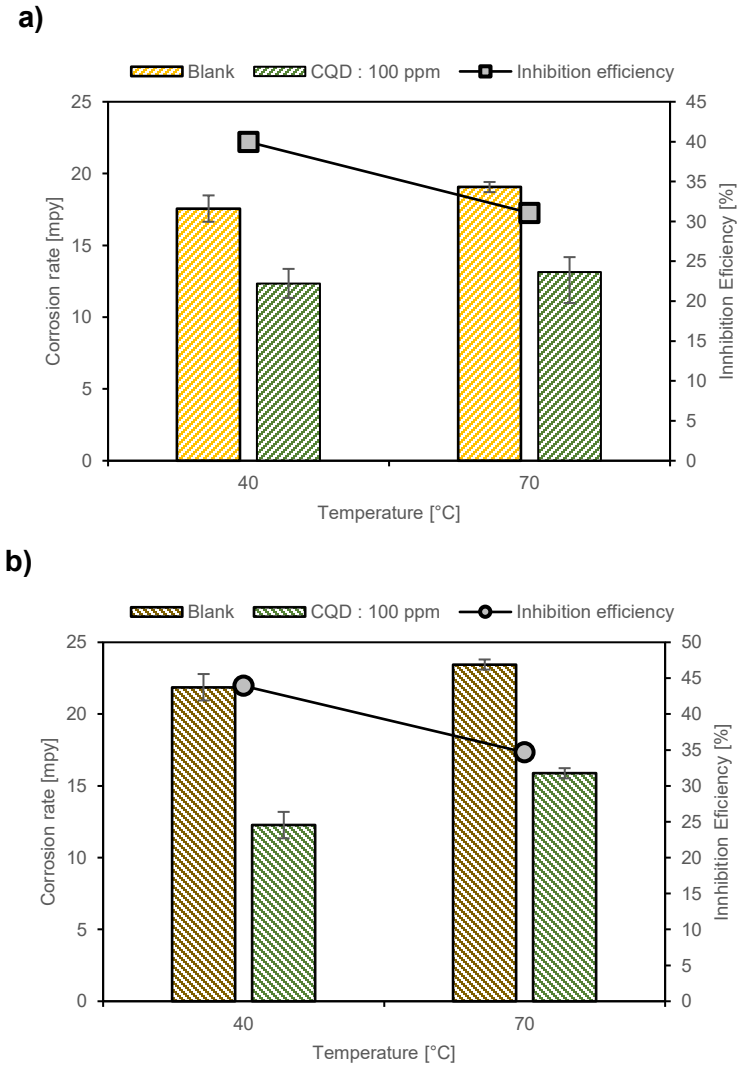


Figura 7. Tasa de corrosión y eficiencia de inhibición de acero al carbón a diferente temperatura durante 24h **a)** NaCl y **b)** $CaCl_2$

Finalmente, en la **Figura 8** se presenta la variación de las tasas de corrosión y eficiencia de inhibición en presencia y ausencia de 100 mg L^{-1} de CQDs luego de varios tiempos de inmersión a temperatura de 70°C.

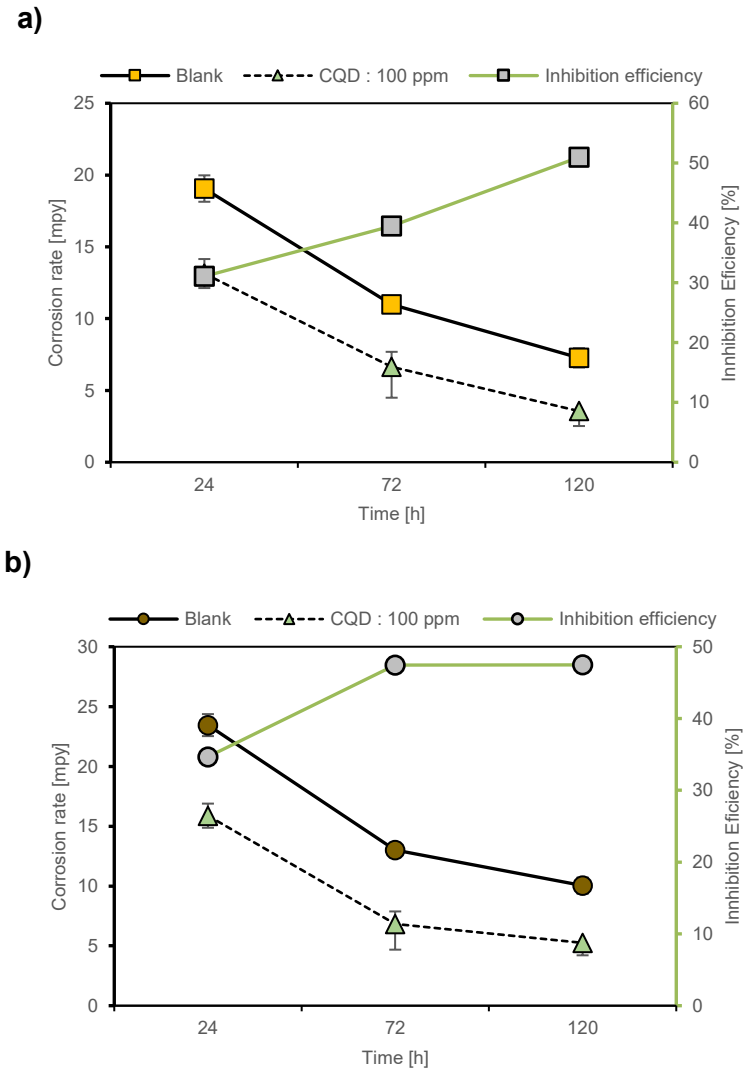


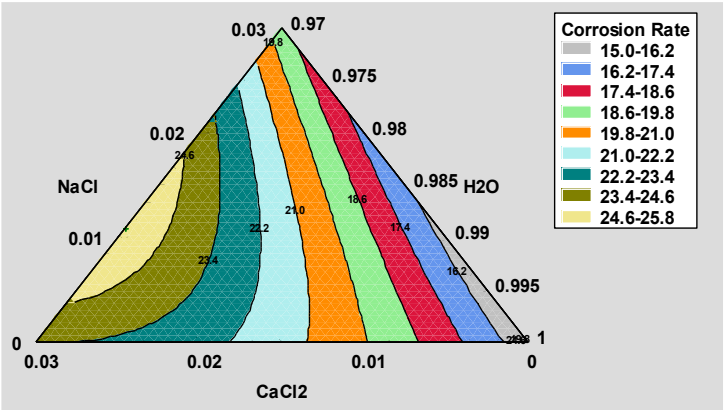
Figura 8. Tasa de corrosión del acero al carbono y eficiencia de inhibición luego de varios tiempos de inmersión **a)** NaCl y **b)** $CaCl_2$

Los resultados muestran que durante las primeras 24 horas se presentan las mayores tasas de corrosión para ambos sistemas, esto puede atribuirse a la disolución del acero por la ausencia de una capa protectora que impida la difusión de las especies corrosivas a la superficie del metal [66]. A medida que el tiempo de inmersión aumenta, una reducción en la tasa de corrosión ocurre, atribuido al desarrollo de películas protectoras y la adición de CQDs.

5.1.3 Modelado de tasas de corrosión con SCDM

En la **Figura 9** se presenta los contornos y superficie de respuesta de la tasa de corrosión en función de las proporciones relativas de NaCl, $CaCl_2$ y H_2O. Se observa un incremento progresivo de las tasas de corrosión a medida que aumenta la fracción de sales en las mezclas, siendo más pronunciado cuando se incrementan simultáneamente las concentraciones de los iones Na^+ y Ca^{2+}. Este resultado se evidencia en la superficie tridimensional observada en la **Figura 9b**, se evidencia una tendencia, en la que la presencia conjunta de los dos iones potencia el efecto corrosivo respecto a la contribución inicial de cada sal. En particular, se aprecia una curva ascendente en la región donde cada componente salino alcanzan valores cercanos al 0.03% de la fracción, indicando un comportamiento sinérgico de los iones monovalentes y divalentes en el aumento de la tasa de corrosión.

a)



b)

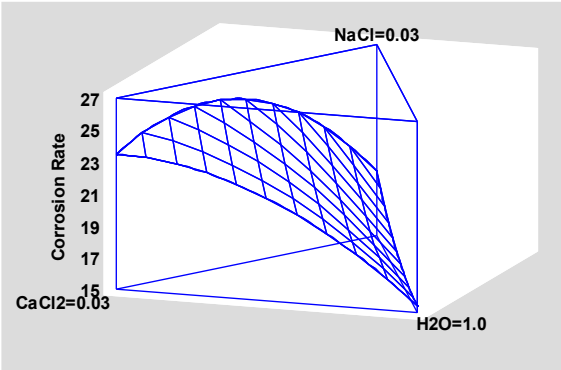


Figura 9. SCMD en ausencia de CQDs **a)** Contornos de la superficie de respuesta y **b)** Superficie de respuesta estimada

El análisis de la superficie de respuesta en ausencia de CQDs permitió ajustar un modelo cúbico especial para la predicción de la tasa de corrosión en función de las proporciones de NaCl, $CaCl_2$ y H_2O. El modelo obtenido se expresa en la Ecuación (13):

$$CR = \beta_1 NaCl' + \beta_2 CaCl_2' + \beta_3 H_2O' + \beta_{12} NaCl'\, CaCl_2' + \beta_{13} NaCl' H_2O' + \beta_{123} NaCl'\, CaCl_2' H_2O' \quad (13)$$

Dónde los coeficientes estimados se presentan en la **Tabla 5**, y en la **Tabla 6** se presentan la predicción y validación de las tasas de corrosión calculadas a partir del modelo cúbico.

Tabla 5. Coeficientes estimados para modelo de CR en ausencia de CQD

Coeficiente	Valor
NaCl	19. 06
$CaCl_2$	23.45
H_2O	15.45
NaCl * $CaCl_2$	15.9
NaCl * H_2O	-3.46
$CaCl_2$ * H_2O	7.92
NaCl * $CaCl_2$ * H_2O	0.21

Tabla 6. Predicción de CR con SCDM en ausencia de CQDs

Test	NaCl	$CaCl_2$	H_2O	CR	CR predicho	%Error
1	0.015	0	0.985	16.39	16.390	0.01
2	0	0	1	15.45	15.450	0.01
3	0	0.015	0.985	21.43	21.430	0.02
4	0.01	0.01	0.98	21.59	21.352	1.09
5	0	0.03	0.97	23.45	23.450	0.00
6	0.015	0.015	0.97	25.23	25.230	0.01
7	0.03	0	0.97	19.06	19.060	0.03
8	0.005	0.005	0.99	17.67	18.298	3.55
9	0.005	0.02	0.975	28.01	23.916	14.62
10	0.02	0.005	0.975	14.50	20.770	43.24

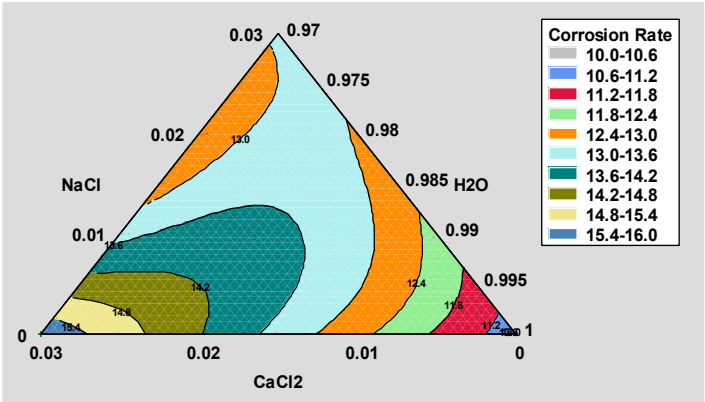
Los puntos del test 8,9,10 fueron usados para validar el modelo obtenido, a partir de estos resultados se calcularon las métricas de desempeño. Obteniendo así valores de

- R^2 = 0.43
- RMSE = 4.34
- MAE = 3.66

De igual forma, se evaluó un diseño de mezclas en presencia de CQDs. En la **Figura 10** se presenta los contornos y superficie de respuesta de la tasa de corrosión en función de las proporciones relativas de NaCl, $CaCl_2$ y H_2O. Se observa una disminución significativa en los valores de la tasa de corrosión en comparación con el escenario sin inhibidor.

Aunque la contribución de las sales (NaCl y $CaCl_2$) continúa siendo determinante en el comportamiento del sistema, el efecto sinérgico que antes potenciaba la corrosión se ve claramente atenuado por la presencia de CQDs. En particular, en las regiones con mayor fracción salina, la corrosión alcanza valores considerablemente menores respecto al sistema en blanco, lo que confirma la eficiencia inhibidora de los CQDs. Estos resultados demuestran que la inclusión de los CQDs modifica sustancialmente la respuesta del sistema, reduciendo no solo los valores absolutos de la tasa de corrosión, sino también la sensibilidad del sistema frente al aumento de la concentración salina.

a)



b)

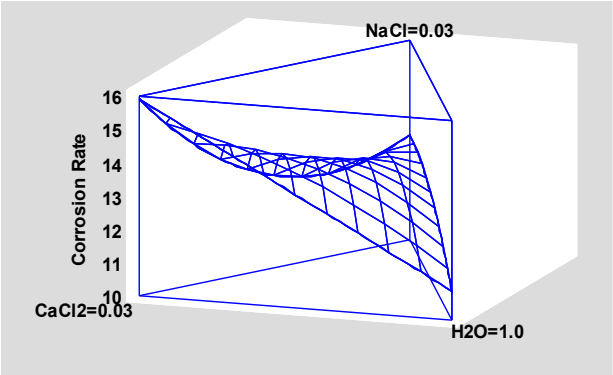


Figura 10. SCMD con presencia de CQDs **a)** Contornos de la superficie de respuesta y **b)** Superficie de respuesta estimada

El análisis de la superficie de respuesta en ausencia de CQDs permitió ajustar un modelo cúbico especial para la predicción de la tasa de corrosión en función de las proporciones de NaCl, $CaCl_2$ y H_2O. El modelo obtenido se expresa en la Ecuación (14):

$$CR_{CQD} = \beta_1 NaCl' + \beta_2 CaCl_2' + \beta_3 H_2O' + \beta_{12} NaCl'\, CaCl_2' + \beta_{13} NaCl' H_2O' + \beta_{123} NaCl'\, CaCl_2' H_2O' \quad (14)$$

Dónde los coeficientes estimados se presentan en la **Tabla 7**, y en la **Tabla 8** se presentan la predicción y validación de las tasas de corrosión calculadas a partir del modelo cúbico.

Tabla 7. Coeficientes estimados para modelo de CR en presencia de CQD

Coeficiente	Valor
NaCl	13.14
$CaCl_2$	15.88
H_2O	10.87
NaCl * $CaCl_2$	-7.04
NaCl * H_2O	2.5
$CaCl_2$ * H_2O	0.1
NaCl * $CaCl_2$ * H_2O	24.48

Tabla 8. Predicción de CR con SCDM en presencia de CQDs

Test	NaCl	$CaCl_2$	H_2O	CR	CR predicho	%Error
1	0.015	0	0.985	12.63	12.630	0.03
2	0	0	1	10.87	10.870	0.03
3	0	0.015	0.985	13.40	13.400	0.04
4	0.01	0.01	0.98	13.71	13.560	1.07
5	0	0.03	0.97	15.88	15.880	0.03
6	0.015	0.015	0.97	12.75	12.750	0.00
7	0.03	0	0.97	13.14	13.140	0.04
8	0.005	0.005	0.99	13.75	12.611	8.28
9	0.005	0.02	0.975	13.25	14.328	8.14
10	0.02	0.005	0.975	12.59	13.155	4.49

Los puntos del test 8,9,10 fueron usados para validar el modelo obtenido, a partir de estos resultados se calcularon las métricas de desempeño. Obteniendo así valores de:

- RMSE = 0.96
- MAE = 0.93

5.2 Análisis descriptivo del conjunto de datos

Con las evaluaciones estáticas fue posible construir un conjunto de datos representativo. El dataset consiste en 252 puntos experimentales y 7 variables fisicoquímicas del sistema como tipo y concentración de sal, fuerza iónica, concentración de inhibidor, pH, temperatura y tiempo de inmersión, además de la variable objetivo correspondiente a la tasa de corrosion. En la **Tabla 9** se presenta el resumen estadístico descriptivo de las variables del dataset.

Tabla 9. Análisis descriptivo de las variables experimentales

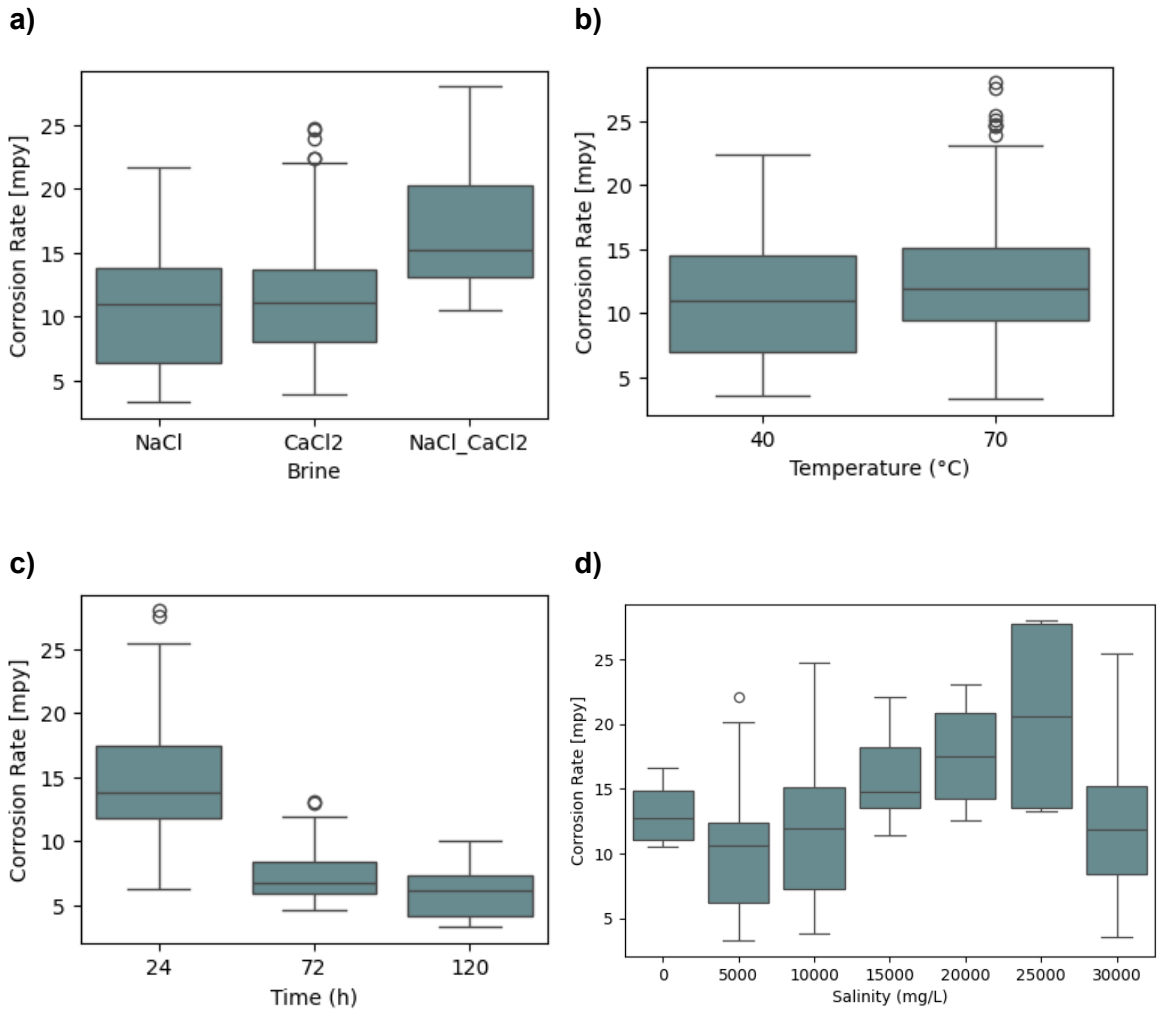
	Inhibitor	Time	pH	Temperature	Salinity	Ionic_strenght	CR
Mean	140.47	46.85	4.08	57.14	15634.92	0.35	11.99
Std	157.00	31.89	0.22	14.87	10804.01	0.25	5.21
Min	0.00	24.00	3.31	40.00	0.00	0.00	3.31
25%	0.00	24.00	3.95	40.00	5000.00	0.17	7.58
50%	100.00	24.00	4.07	70.00	10000.00	0.27	11.49
75%	300.00	72.00	4.23	70.00	30000.00	0.51	14.84
Max	500.00	120.00	4.61	70.00	30000.00	0.81	28.01

Se puede observar que hay una considerable dispersión en los datos, particularmente en la salinidad y concentración del inhibidor. Para la concentración de CQDs se indica una distribución muy asimétrica, lo que refleja una clara distinción entre muestras sin presencia de inhibidor o con bajas dosis y algunos con concentraciones altas. El tiempo de inmersión tiene una mediana de 24h y una media de 48.6h, evidencia que gran parte de las mediciones corresponde a experimentos para las primeras etapas corrosivas, aunque se incluyen tiempos hasta 120h.

La salinidad con rangos de 0 a 30000 mg L^{-1} presenta una media cerca de 15635 mg L^{-1}, con una alta desviación estándar, lo que refleja la inclusión de condiciones de baja y alta salinidad en el diseño experimental. En cuanto a la fuerza iónica asociada, varía entre 0 y 0.811 M, concordando con las diferencias entre las soluciones por el efecto del ion monovalente Na^+ y divalente Ca^{2+} estudiados. El pH se mantuvo relativamente estable, con una desviación estándar baja de 0.22, lo que facilita comparar efectos iónicos sin la presencia de grandes variaciones en la acidez del medio.

Finalmente, la tasa de corrosión presenta un amplio rango desde 3.31 mpy a 28.01 mpy, con media cerca de 12 mpy y una desviación estándar de 5.21 mpy, lo que indica que las mayores tasas de corrosión suceden bajo condiciones específicas. La distribución de la tasa de corrosion sugiere una variabilidad moderada ante las condiciones experimentales evaluadas.

En general, este análisis estadístico sugiere distribuciones asimétricas para las variables de inhibidor, fuerza iónica y salinidad, y una CR con dispersión suficiente para construir y entrenar los modelos predictivos. En la **Figura 11** se presentan los diagramas de caja que muestran la variación de la tasa de corrosión frente a los principales factores experimentales.



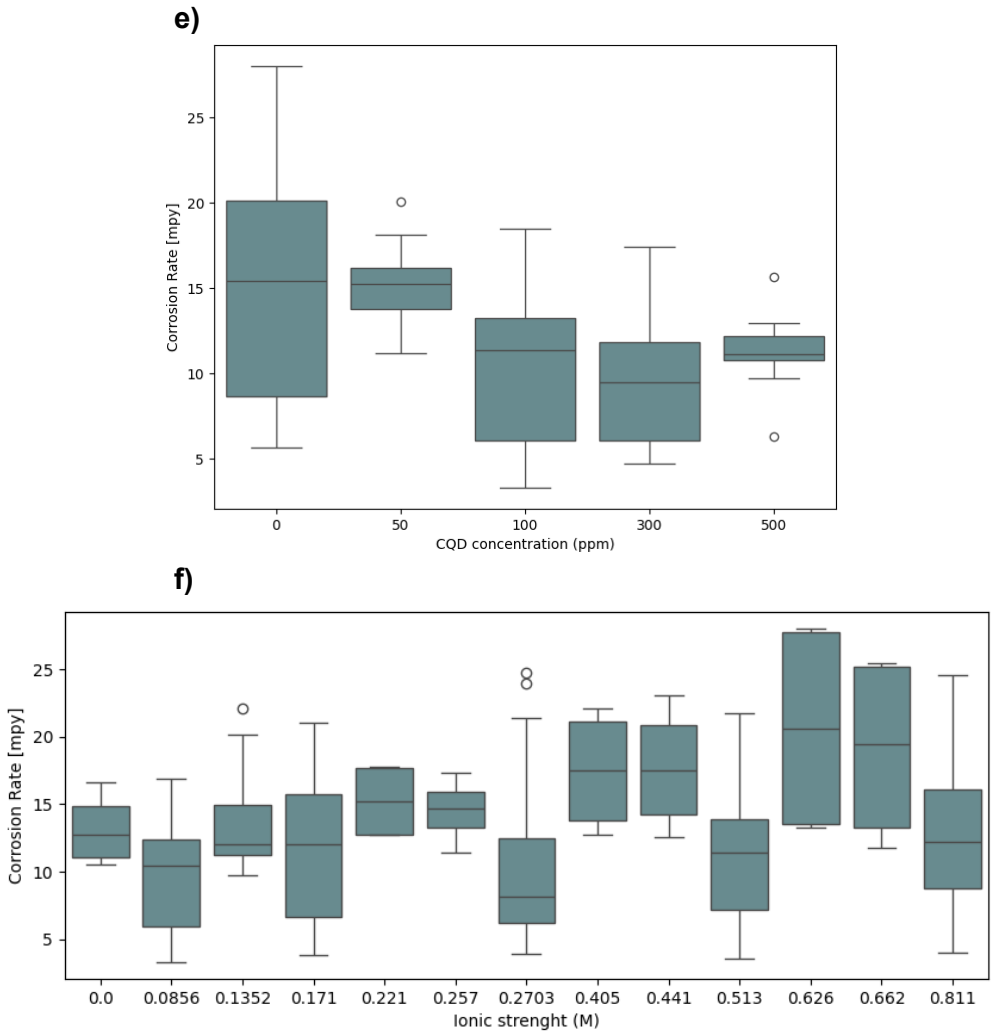


Figura 11. Distribución de la tasa de corrosión frente a variables experimentales **a)** Tipo de ion **b)** Temperatura **c)** Tiempo de inmserión **d)** Salinidad **e)** Concentración de CQD y **f)** Fuerza iónica

5.2.1 Análisis de Correlación

La matriz de correlación de Pearson proporciona información sobre las relaciones entre los distintos factores que influyen en la corrosión. Como se muestra en la **Figura 12**, los coeficientes de correlación entre las variables estudiadas revelan interacciones clave que afectan la tasa de corrosión.

El análisis de correlación entre las variables experimental muestra una fuerte correlación positiva entre salinidad y fuerza iónica (r = 0.92), es decir, que ambas variables están aportando información muy similar.

Para evitar colinealidad, se decidió eliminar la variable salinidad, a pesar que la fuerza iónica es un parámetro derivado de la salinidad, la inclusión de esta variable resulta más adecuada, ya que representa de manera más directa el efecto de los iones presentes en la solución sobre las tasas de corrosión.

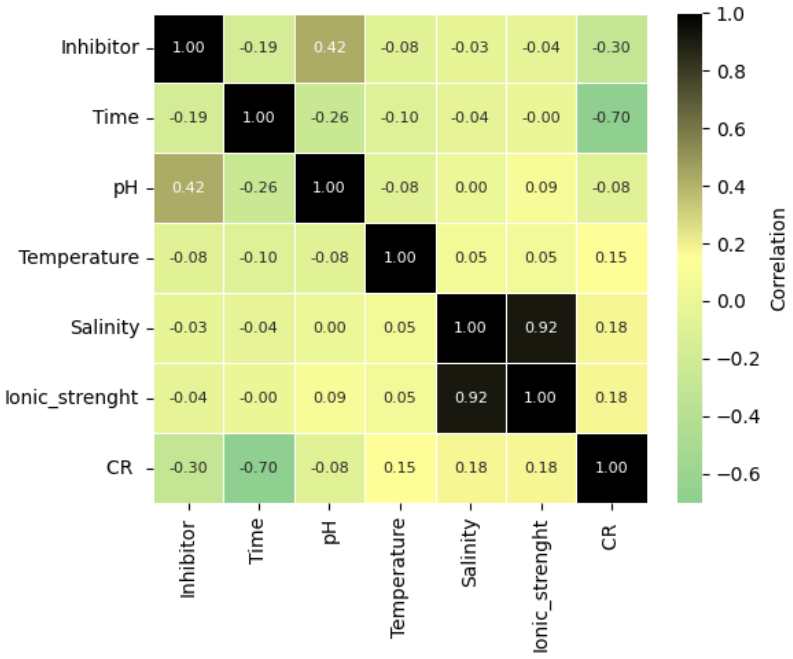


Figura 12. Matriz de correlación

La matriz de correlación ajustada muestra que CR presenta una correlación negativa con la concentración de CQDs, lo cual confirma su acción inhibitoria sobre la disolución del acero. De igual manera, se observa una correlación negativa (r = 0.70) más marcada con el tiempo de inmersión, lo cual puede atribuirse a la formación de las capas pasivas de productos de corrosión que reducen el ataque de las especies corrosivas del medio. En contraste, la fuerza iónica mantiene una correlación positiva con CR, reflejando que un mayor contenido de iones en solución favorece los procesos corrosivos.

El pH muestra correlaciones débiles con CR, pero una correlación positiva (0.42) con la adición de CQDs, lo que sugiere posibles interacciones químicas dependientes de la concentración de CQDs. Finalmente, la temperatura muestra una leve correlación positiva

(0.15) con CR, consistente con el efecto esperado de la aceleración de la cinética de las reacciones electroquímicas, sin embargo, no llega a ser determinante frente al efecto de la fuerza iónica, tiempo y presencia de CQDs.

5.3 Desempeño de modelos de aprendizaje de máquinas implementados para la predicción de tasas de corrosión

Con el fin de evaluar la capacidad predictiva de los modelos de aprendizaje supervisado sobre la tasa de corrosión, se diseñaron dos escenarios de análisis. En el primero, se consideraron únicamente los experimentos a 24 h de inmersión, lo que permite discriminar los efectos iniciales de las variables fisicoquímicas sobre el comportamiento corrosivo sin la influencia de la evolución temporal o la formación de capas pasivas protectoras. En el segundo escenario, se incorporaron todos los tiempos de exposición, con el propósito de capturar la dinámica completa del sistema y evaluar la robustez de los modelos frente a condiciones experimentales más diversas. Esta aproximación comparativa permite examinar la influencia relativa de las variables en diferentes etapas del proceso de corrosión, así como valorar la consistencia de las predicciones obtenidas.

5.3.1 Predicción de la tasa de corrosión en condiciones iniciales

En este primer escenario se analizaron las tasas de corrosión luego de 24 horas de inmersión, con el propósito de evaluar la capacidad predictiva de los modelos durante la etapa inicial de la corrosión inducida por CO_2. Este enfoque nos permite aislar el efecto del tipo de ion, fuerza iónica y concentración de CQDs. En la **Figura 13** se presentan las curvas de ajuste de las predicciones de la tasa de corrosión para los modelos de aprendizaje de máquinas evaluados.

El eje X corresponde a los valores reales de la tasa de corrosión, mientras que el eje Y muestra los valores predichos por cada modelo. La línea X = Y se incluye como referencia, representando un escenario ideal en el cual las predicciones coinciden perfectamente con los valores experimentales. En el caso óptimo, todos los puntos de datos se encontrarían próximos a esta línea de referencia, lo que evidenciaría una alta precisión predictiva.

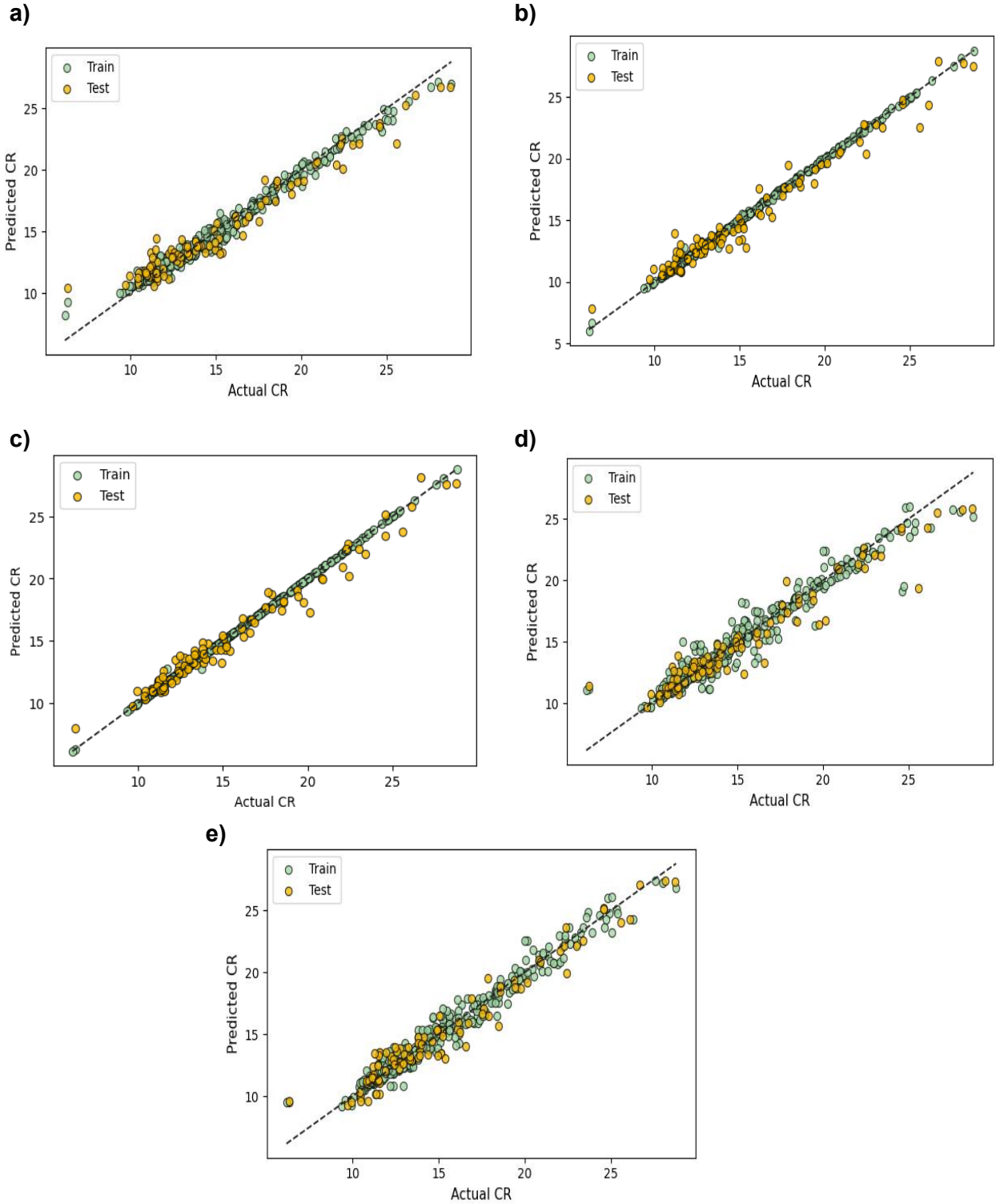


Figura 13. Comparación de los valores reales y predichos con diferentes modelos de aprendizaje supervisado evaluados a 24h **a)** RF **b)** XGBoost **c)** KNN **d)** SVR y **e)** MLP

RF muestra dispersión mínima para los datos de entrenamiento como en prueba, sin embargo, presenta una ligera dispersión para tasas de corrosión altas (>20 mpy), lo que podía atribuirse a la variabilidad entre las interacciones iónicas causadas por la presencia de Ca^{2+} y los CQDs, a pesar de esto RF tiene un buen ajuste, es robusto frente a valores atípicos y tiene una buena capacidad de generalización, lo que resalta la fortaleza del ensamble de árboles de decisiones para manejar la no linealidad o multicolinealidad de estas variables.

La **Figura 14** muestra las variables más dominantes en la predicción de CR. Se observa que tanto la concentración de los CQDs como la fuerza iónica, son dos de los factores más influyentes, respaldando así que el efecto del tipo de ion, como la presencia de CQDs dominan los procesos corrosivos, por la posibles interacciones entre el tipo de ion y los CQDs que dan lugar a la formación de capas protectoras mitigando las tasas de corrosión.

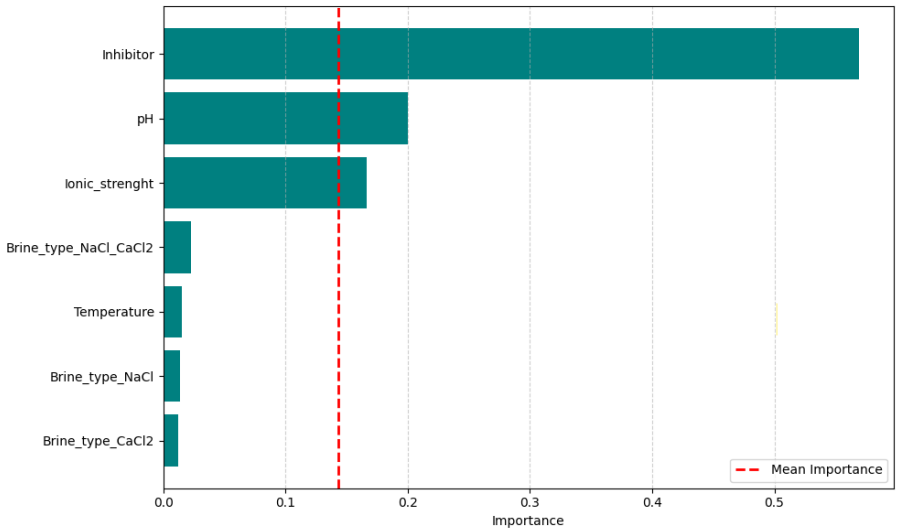


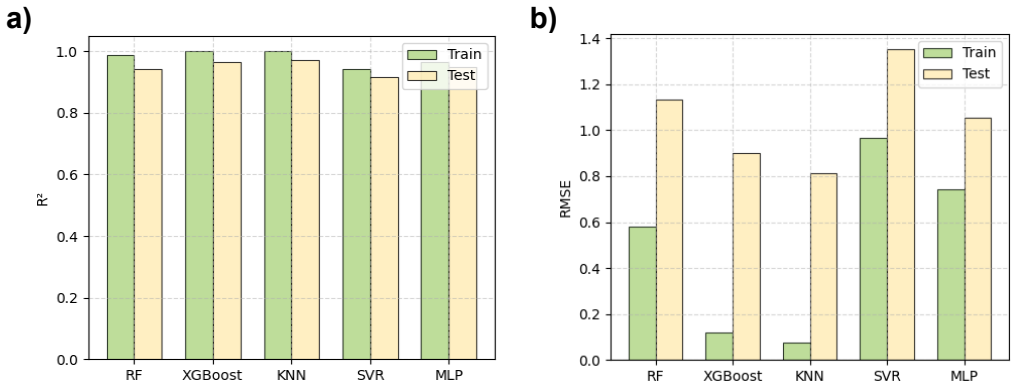
Figura 14. Importancia de caracteristicas para RF implementado en la predicción de tasas de corrosion a 24h

Similarmente, XGBoost presenta un muy buen ajuste entre los datos de entrenamiento y prueba, como se ve en la distribución uniforme de los valores predichos y reales, especialmente en las tasas de corrosion entre 15 y 25 mpy, donde los puntos del conjunto de testeo se superponen casi perfectamente sobre los puntos de entrenamiento y validación, indicando que el boosting es efectivo para ponderar las contribuciones de las variables claves. Por otro lado, KNN muestra un buen rendimiento general, pero con

mayores puntos de dispersión en valores extremos, lo que se traduce en subestimación de tasas de corrosión altas o sobreestimación para tasas de corrosión bajas, este resultado indica que la predicción basada en vecinos captura bien las tasas de corrosión cuando hay similitudes entre los valores causados por cercanías entre parámetros como pH o temperatura, pero tiene limitaciones frente a valores extremos, cuando el fenómeno corrosivo es dominado por otras variables.

El modelo SVR muestra de igual forma un buen ajuste, con dispersión mínima en los conjuntos de entrenamiento, pero con desviaciones ligeras en el conjunto de prueba, lo que indicar un margen de error subóptimo en regiones donde la tasa de corrosión es baja. Finalmente, el modelo MLP muestra una dispersión uniforme en los conjuntos de entrenamiento y prueba, lo que destaca la flexibilidad de las redes neuronales para aprender patrones complejos de interacción entre las variables experimentales.

Estas observaciones son evaluadas junto con las métricas de desempeño. En **Figura 15** se muestran las métricas de error evaluadas para cada modelo, tanto en los conjuntos de entrenamiento y validación como en los de testeo.



c)

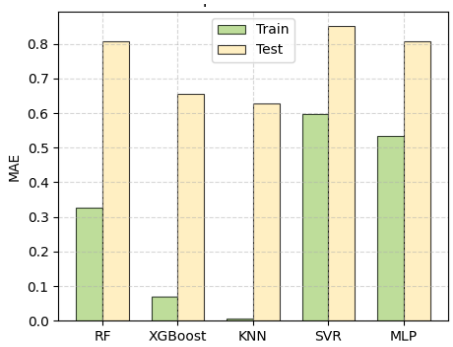


Figura 15. Métricas de desempeño para los modelos evaluados en la predicción de tasas de corrosión a 24h **a)** R^2 **b)** RMSE y **c)** MAE

Los resultados obtenidos evidencian un alto R^2 (>0.9) para todos los modelos, lo que indican las altas capacidades de generalización, sin embargo, cuando se evalúa RMSE y MAE, se encontró que, para RF estos valores aumentaron de 0.5797 y 0.4547, respectivamente, desde el conjunto de entrenamiento a 1.1345 y 0.8624 en el conjunto de prueba, lo que sugiere que el modelo se ajusta mejor a los datos de entrenamiento que a datos nuevos, lo que sugiere un ligero sobreajuste del modelo.

Para XGBoost y KNN se presentaron sobreajustes significativos. Aunque XGBoost obtuvo un R^2 de 0.96, y alcanza el mejor equilibrio entre el conjunto de entrenamiento y prueba, a pesar de esto, la gran diferencia entre las métricas de entrenamiento (RMSE: 0.1183, MAE: 0.0711) y prueba (RMSE: 0.8991, MAE: 0.6567), indica que el modelo memorizó los datos de entrenamiento, este resultado, concuerda con la dispersión de los datos presentada en el ajuste de los datos predichos y reales. De manera similar, KNN presenta el menor RMSE y MAE en prueba (0.8142 y 0.6273), lo que lo hace un modelo eficiente para predicciones locales, los valores extremadamente bajos en entrenamiento (MAE: 0.0061), sugiere un alto potencial de sensibilidad al ruido de los datos vecinos.

Por otro lado, SVR muestra ajustes conservadores (R^2: 0.9170), y RMSE y MAE son los más altos en el conjunto de prueba, lo que indica que es el menos preciso de los modelos evaluados. Y MLP muestra un buen rendimiento, con menos sobreajuste que los demás modelos. La diferencia entre las métricas de entrenamiento (RMSE: 0.7422, MAE: 0.5340) y prueba (RMSE: 1.0548, MAE: 0.8064), son menores que para XGBoost y KNN, lo que indica una mejor capacidad de generalización.

Por último, en la tabla **Tabla 10** se presentan los mejores hiperparámetros encontrados en la optimización de los modelos. Los hiperparámetros se optimizaron mediante búsqueda en grilla y validación cruzada de 5-fold, asegurando una generalización robusta en este dataset de tamaño moderado.

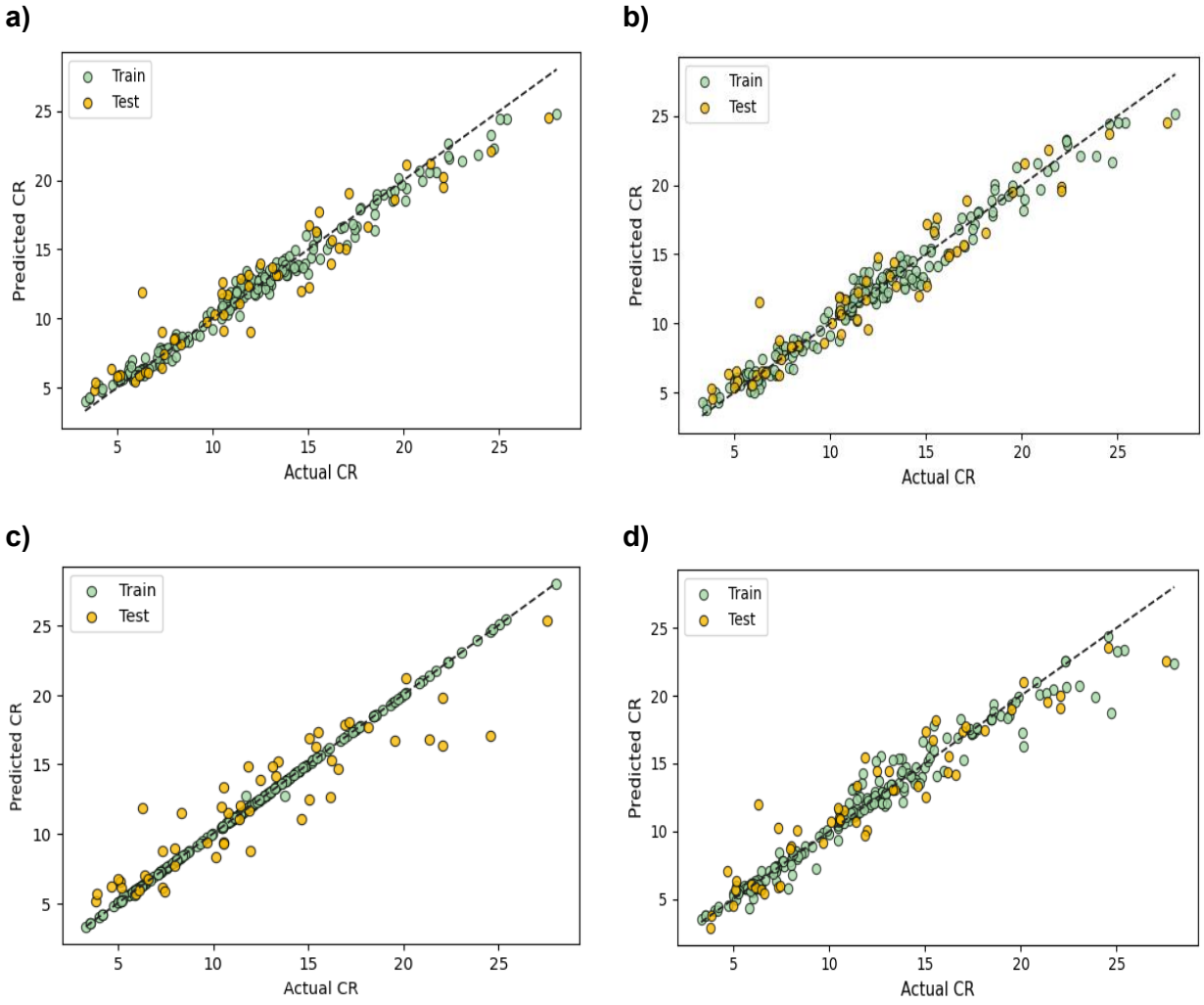
Las métricas obtenidas validan la selección de hiperparámetros, que tienen como enfoque minimizar la diferencia entre entrenamiento y prueba para evitar sobreajuste en la predicción de las tasas de corrosión.

Tabla 10. Hiperparámetros resultantes de la optimización de los modelos implementados para la predicción de tasas de corrosion a 24h

Modelo	Hiperparámetros	Valores
RF	criterion	squared_error
	max_depth	10
	max_features	sqrt
	max_samples	0.9
	min_samples_leaf	1
	min_samples_split	2
	n_estimators	300
XGBoost	colsample_bytree	0.9
	learning_rate	0.1
	max_depth	9
	min_child_weight	5
	n_estimators	300
	subsample	0.7
KNN	algorithm	auto
	leaf_size	10
	metric	manhattan
	n_neighbors	5
	weights	distance
SVR	C	100
	epsilon	0.2
	gamma	scale
	kernel	rbf
MLP	activation	relu
	alpha	0.001
	hidden_layer_sizes	(100, 100)
	learning_rate	constant
	max_iter	1000
	solver	adam

5.3.2 Predicción de la tasa de corrosión en el tiempo

En el segundo escenario, se incluyeron todas las condiciones de tiempo de inmersión (24, 72 y 120 h), con el objetivo de capturar la evolución temporal del sistema y analizar la capacidad de los modelos para generalizar en un contexto más complejo. Con esta estrategia se busca identificar la sensibilidad de los modelos desde las condiciones iniciales y su robustez frente a la dinámica completa de la corrosión en medios carbonatados. En la **Figura 16** se presentan las curvas de ajuste de las predicciones de la tasa de corrosión para los modelos de aprendizaje de máquinas evaluados en el tiempo.



e)

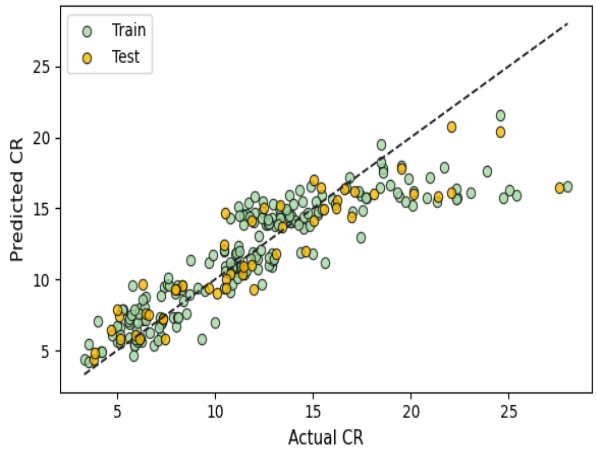


Figura 16. Comparación de los valores reales y predichos con diferentes modelos de aprendizaje supervisado evaluados en el tiempo **a)** RF **b)** XGBoost **c)** KNN **d)** SVR y **e)** MLP

La **Figura 17** muestra las variables más dominantes en la predicción de CR. Se observa que, para este nuevo escenario estudiado, se presenta un cambio respecto a las variables más relevantes. Se encontró ahora que el tiempo de inmersión y la concentración de los CQDs se vuelven los factores más críticos en la predicción de las tasas de corrosión.

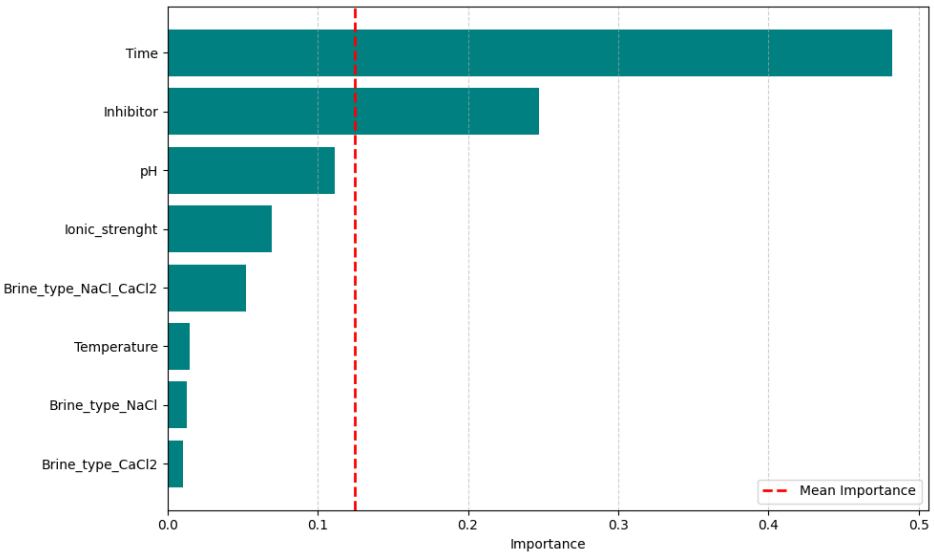


Figura 17. Importancia de caracteristicas para RF implementado en la predicción de tasas de corrosion en el tiempo

Este resultado tiene sentido cuando se analiza la corrosión en el tiempo. Según los resultados obtenidos en **Figura 8**, se confirma que las tasas de corrosión disminuyen con

el aumento del tiempo lo cual en ausencia de inhibidores puede deberse a la formación de las películas protectoras, y en presencia del inhibidor, esta disminución es más pronunciada, porque su función es adsorberse sobre la superficie del metal e intervenir en la difusión de iones corrosivos. En el escenario anterior, las tasas de corrosión eran influenciadas con gran importancia por la fuerza iónica, sin embargo, en este escenario el mecanismo de corrosión cambia y el tiempo se vuelve predominante, por lo que opaca la importancia de esta variable en la predicción de las tasas de corrosión.

El comportamiento y el rendimiento de los modelos cambio significativamente con el nuevo conjunto de datos. En **Figura 18** se muestran las nuevas métricas de error evaluadas para cada modelo, tanto en los conjuntos de entrenamiento y validación como en los de testeo.

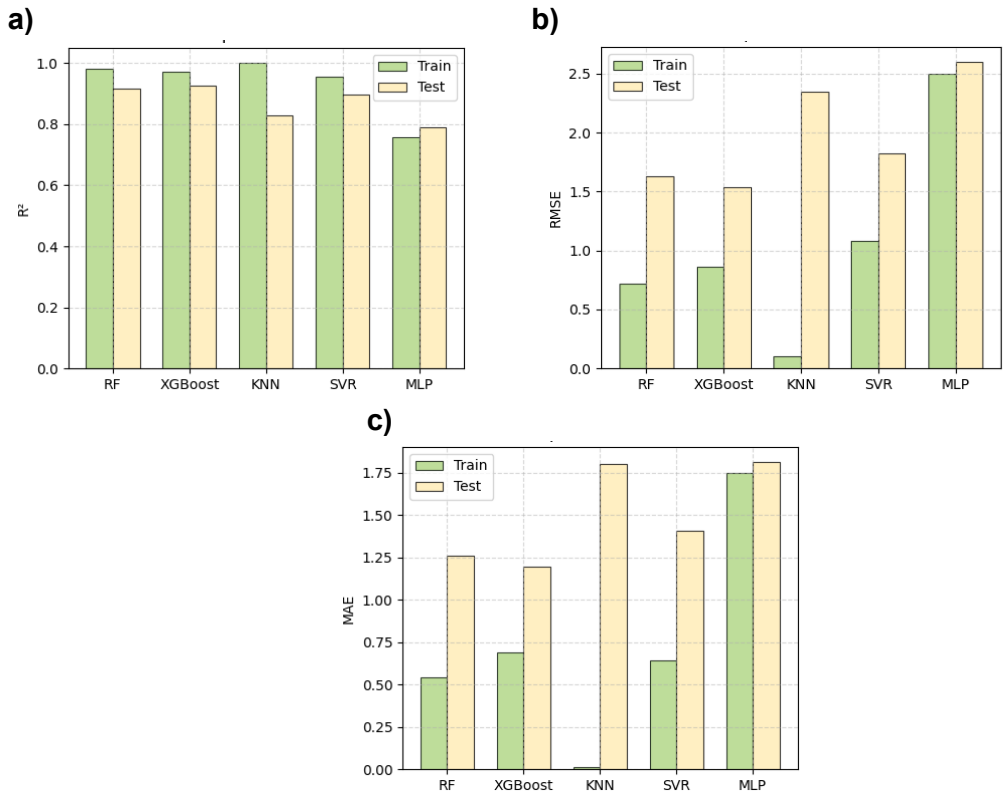


Figura 18. Métricas de desempeño para los modelos evaluados en la predicción de tasas de corrosión en el tiempo **a)** R^2 **b)** RMSE y **c)** MAE

Los resultados demuestran que la mayoria de los modelos mantienen un desempeño sólido, validando su capacidad de generalización a nuevos datos. RF y XGBoost son los

modelos mas efectivos, con R^2 de 0.9168 y 0.9264, respectivamente, lo que confirma su capacidad de capturar la variablididad de los datos.

KNN presenta un mayor sobreajuste que en el escenario anterior, ya que en el conjunto de entrenamiento obtuvo R^2 de 0.9996, RMSE de 0.1071 y MAE de 0.0152, pero la diferencia entre su desempeño en prueba es considerablemente peor para RMSE (2.53515) y MAE (1.8048), esto sugiere que el modelo memorizó los datos de entrenamiento en lugar de aprender los nuevos patrones de las tasas de corrosión evaluadas.

En contraste, MLP obtuvo métricas de prueba mejores que las de entrenamiento (R^2_{test}: 0.7892 vs R^2_{train}:0.7579). Si bien su desempeño es más bajo entre los modelos evaluados, tiene una buena capacidad para evitar sobreajuste.
Finalmente, en la tabla **Tabla 11** se presentan los mejores hiperparámetros encontrados en la optimización de los modelos.

Tabla 11. Hiperparámetros resultantes de la optimización de los modelos implementados para la predicción de tasas de corrosion en el tiempo

Modelo	Hiperparámetros	Valores
RF	criterion	squared_error
	max_depth	10
	max_features	log2
	max_samples	0.9
	min_samples_leaf	1
	min_samples_split	2
	n_estimators	100
XGBoost	colsample_bytree	0.9
	learning_rate	0.1
	max_depth	3
	min_child_weight	1
	n_estimators	100
	subsample	0.7
KNN	algorithm	ball_tree
	leaf_size	30
	metric	manhattan
	n_neighbors	5
	weights	distance
SVR	C	100
	epsilon	0.2
	gamma	auto
	kernel	rbf
	activation	relu

	alpha	0.0001
	hidden_layer_sizes	(50, 50)
MLP	learning_rate	constant
	max_iter	1000
	solver	adam

Se encontró que, para RF un subconjunto más pequeño de características son las predictivas, al presentar una reducción de max_features de sqrt a log2. Además, los resultados muestran que un menor número de iteraciones es suficiente para alcanzar un buen desempeño. De manera similar a RF, para XGBoost se presentó una reducción de n_estimators de 300 a 100. Para MLP, la arquitectura de la red se hizo más pequeña y simple, el hiperparámetro hidden_layer_sizes se redujo de (100,100) a (50,50), lo que indica que para el segundo escenario no se requiere una red tan grande para aprender nuevos patrones y prevenir el sobreajuste.

Conclusiones y recomendaciones

La predicción de la tasa de corrosión en sistemas salinos carbonatados constituye un reto complejo debido a la interacción de múltiples factores. En este trabajo permitió evaluar la influencia de Ca^{2+} en los procesos corrosivos y como los CQDs se desempeñan como inhibidores cuando son evaluados bajo estas condiciones experimentales que no se habían considerado anteriormente. Estas evaluaciones nos permitieron conformar un conjunto de datos representativo de procesos corrosivos con el fin de implementar modelos de aprendizaje de maquinas teniendo como variables fuerza iónica, tiempo de inmersión, temperatura y CQDs.

Además, en este trabajo la implementación de un diseño de mezclas simplex-centroide permitió ajustar modelos cúbicos especiales que reprodujeron adecuadamente los valores experimentales dentro de cada escenario, sin embargo, en pruebas de validación externa los resultados su desempeño fue significativamente inferior sin inhibidor se obtuvo un (R^2 = 0.43, RMSE = 4.34 mpy y MAE = 3.66 mpy). Esto confirma que el SCDM describe bien el dominio experimental, pero pierde capacidad de generalización fuera de él.

En contraste, los modelos de aprendizaje de máquinas ofrecieron un marco más flexible y robusto, al integrar simultáneamente todas las variables experimentales (salinidad, composición iónica, inhibidor, pH, tiempo). Entre los algoritmos evaluados, Random Forest y XGBoost destacaron por su desempeño superior: en los conjuntos de prueba alcanzaron $R^2 > 0.91$, con RMSE entre 0.89 -1.1 mpy y MAE inferiores a 1 mpy, evidenciando una alta capacidad predictiva. Otros algoritmos como MLP lograron resultados aceptables (R^2: 0.85, RMSE: 1.2 mpy, MAE: 0.8 mpy), mostrando su buena capacidad de generalización.

En un segundo escenario, se incorporó además el tiempo de inmersión (24, 72 y 120 h), lo que permitió capturar la evolución temporal del proceso corrosivo. En este caso, los modelos mantuvieron un desempeño elevado: Random Forest y XGBoost alcanzaron R^2 superiores a 0.93, con RMSE cercanos a 1.7 mpy y MAE cercanos a 1.25 mpy, lo que

refleja una disminución ante el escenario de solo 24h, estos modelos logran capturar con precisión el efecto acumulativo del tiempo. Otros algoritmos, como SVR y MLP, también mostraron un ajuste aceptable (R^2 entre 0.80 y 0.87, RMSE ≈ 3.0 mpy, MAE ≈ 2.3 mpy). Estos resultados confirman que los algoritmos de ML no solo representan adecuadamente el efecto de los iones, sino también la dinámica del sistema a lo largo del tiempo, manteniendo su capacidad predictiva en condiciones más complejas.

Un aspecto relevante es que los modelos de ML lograron capturar de forma integral el efecto inhibidor de los CQDs, reflejando no solo la reducción en la tasa de corrosión sino también la menor sensibilidad del sistema frente al incremento de salinidad y fuerza iónica. A diferencia del diseño de mezclas, que requiere un modelo separado para cada condición, los algoritmos de aprendizaje de máquinas integraron en una sola estructura los escenarios con y sin inhibidor, simplificando el proceso y mejorando la capacidad de predicción.

En conjunto, los hallazgos de este trabajo permiten concluir que los modelos de aprendizaje de máquinas representan una alternativa más potente y generalizable. Su capacidad para manejar la complejidad del sistema corrosivo, integrar múltiples variables y mantener métricas de desempeño superiores en validación externa los posiciona como una metodología de predicción más precisa y aplicable en escenarios industriales, contribuyendo al diseño de estrategias de mitigación de la corrosión más efectivas.

Bibliografía

[1] F. Nath, M. N. Mahmood, and N. Yousuf, “Recent advances in CCUS: A critical review on technologies, regulatory aspects and economics,” *Geoenergy Science and Engineering*, vol. 238, p. 212726, 2024.

[2] National Oceanic and Atmospheric Administration (NOAA), “Despite pandemic shutdowns, carbon dioxide and methane surged in 2020.” Accessed: Sep. 05, 2025. [Online]. Available: https://research.noaa.gov/2021/04/07/despite-pandemic-shutdowns-carbon-dioxide-and-methane-surged-in-2020/

[3] GML, “Carbon Cycle Greenhouse Gases.” Accessed: Sep. 05, 2025. [Online]. Available: https://gml.noaa.gov/ccgg/

[4] G. P. D. De Silva, P. G. Ranjith, and M. S. A. Perera, “Geochemical aspects of CO2 sequestration in deep saline aquifers: A review,” *Fuel*, vol. 155, pp. 128–143, 2015.

[5] A. Morgan, R. Grigg, and W. Ampomah, “A gate-to-gate life cycle assessment for the CO2-EOR operations at farnsworth unit (FWU),” *Energies (Basel)*, vol. 14, no. 9, p. 2499, 2021.

[6] K. Brown, S. Whittaker, M. Wilson, W. Srisang, H. Smithson, and P. Tontiwachwuthikul, “The history and development of the IEA GHG Weyburn-Midale CO2 Monitoring and Storage Project in Saskatchewan, Canada (the world largest CO2 for EOR and CCS program),” *Petroleum*, vol. 3, no. 1, pp. 3–9, 2017.

[7] P. Zhai *et al.*, “Global Warming of 1.5 C. An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty,” *Sustainable Development, and Efforts to Eradicate Poverty*, p. 32, 2018.

[8] Working Group III of the Intergovernmental Panel on Climate Change, “IPCC Special Report on Carbon Dioxide Capture and Storage,” 2005. Accessed: Sep. 05, 2025. [Online]. Available: https://www.ipcc.ch/report/carbon-dioxide-capture-and-storage/

[9] National Oceanic and Atmospheric Administration (NOAA) and Global Monitoring Laboratory, “Trends in atmospheric carbon dioxide (CO2).” Accessed: Sep. 05, 2025. [Online]. Available: https://gml.noaa.gov/ccgg/trends/

[10] A. Dubey and A. Arora, “Advancements in carbon capture technologies: A review,” *J Clean Prod*, vol. 373, p. 133932, 2022.

[11] Y. A. Alli, N. E. Magida, F. Matebese, N. Romero, A. S. Ogunlaja, and K. Philippot, “Biomimetic photocatalysts for the transformation of CO2: design, properties, and mechanistic insights,” *Mater Today Energy*, vol. 34, p. 101310, 2023.

[12] International Energy Agency, “Net zero by 2050, IEA, Paris.” Accessed: Sep. 08, 2025. [Online]. Available: https://www.iea.org/reports/net-zero-by-2050

[13] P. Zhai *et al.*, “Global Warming of 1.5 C. An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty,” *Sustainable Development, and Efforts to Eradicate Poverty*, p. 32, 2018.

[14] E. Hanson, C. Nwakile, and V. O. Hammed, “Carbon capture, utilization, and storage (CCUS) technologies: Evaluating the effectiveness of advanced CCUS solutions for reducing CO2 emissions,” *Results in Surfaces and Interfaces*, vol. 18, p. 100381, 2025.

[15] P. R. Yaashikaa, P. S. Kumar, A. Saravanan, S. Karishma, and G. Rangasamy, “A biotechnological roadmap for decarbonization systems combined into bioenergy production: Prelude of environmental life-cycle assessment,” *Chemosphere*, vol. 329, p. 138670, 2023.

[16] W. Ampomah, A. Morgan, D. O. Koranteng, and W. I. Nyamekye, “CCUS perspectives: Assessing historical contexts, current realities, and future prospects,” *Energies (Basel)*, vol. 17, no. 17, p. 4248, 2024.

[17] Q. Deng, X. Ling, K. Zhang, L. Tan, G. Qi, and J. Zhang, “CCS and CCUS technologies: Giving the oil and gas industry a green future,” *Front Energy Res*, vol. 10, p. 919330, 2022.

[18] L. Yang, W. Rui, Z. Qingmin, Z. Yuanlong, F. Xin, and X. Zhaojie, “CO2-enhanced oil recovery with CO2 utilization and storage: Progress and practical applications in China,” *Unconventional Resources*, vol. 4, p. 100096, 2024.

[19] J. N. Turkson *et al.*, “Carbonated water injection for enhanced oil recovery and CO2 geosequestration in different CO2 repositories: a review of physicochemical processes and recent advances,” *Energy & Fuels*, vol. 38, no. 8, pp. 6579–6612, 2024.

[20] H. Yu *et al.*, “Enhanced oil recovery and CO2 storage by enhanced carbonated water injection: a mini-review,” *Energy & Fuels*, vol. 38, no. 14, pp. 12409–12432, 2024.

[21] A. Talebi, A. Hasan-Zadeh, Y. Kazemzadeh, and M. Riazi, “A review on the application of carbonated water injection for EOR purposes: Opportunities and challenges,” *J Pet Sci Eng*, vol. 214, p. 110481, 2022.

[22] G. Schmitt, R. H. Hausler, and H. P. Giddard, “Advances in CO2 corrosion,” *NACE, Houston*, vol. 1, 1984.

[23] A. Kahyarian, M. Singer, and S. Nesic, “Modeling of uniform CO2 corrosion of mild steel in gas transportation systems: a review,” *J Nat Gas Sci Eng*, vol. 29, pp. 530–549, 2016.

[24] H. Mansoori, R. Mirzaee, F. Esmaeilzadeh, A. Vojood, and A. S. Dowrani, “Pitting corrosion failure analysis of a wet gas pipeline,” *Eng Fail Anal*, vol. 82, pp. 16–25, 2017.

[25] Y. Hua, R. Barker, and A. Neville, “Comparison of corrosion behaviour for X-65 carbon steel in supercritical CO2-saturated water and water-saturated/unsaturated supercritical CO2,” *J Supercrit Fluids*, vol. 97, pp. 224–237, 2015.

[26] M. B. Kermani and A. Morshed, “Carbon dioxide corrosion in oil and gas productiona compendium,” *Corrosion*, vol. 59, no. 08, 2003.

[27] R. Barker, D. Burkle, T. Charpentier, H. Thompson, and A. Neville, “A review of iron carbonate (FeCO3) formation in the oil and gas industry,” *Corros Sci*, vol. 142, pp. 312–341, 2018.

[28] A. B. Forero, M. Núñez, and I. S. Bott, “Analysis of the corrosion scales formed on API 5L X70 and X80 steel pipe in the presence of CO2,” *Materials Research*, vol. 17, pp. 461–471, 2014.

[29] H. Fang, B. Brown, and S. Nešić, “Sodium chloride concentration effects on general CO2 corrosion mechanisms,” *Corrosion*, vol. 69, no. 3, pp. 297–302, 2013.

[30] R. Elgaddafi, R. Ahmed, and S. Shah, “Corrosion of carbon steel in CO2 saturated brine at elevated temperatures,” *J Pet Sci Eng*, vol. 196, p. 107638, 2021.

[31] H. Mansoori, F. ESMAEILZADEH, D. MOWLA, and A. H. Mohammadi, “Case study: production benefits from increasing C-values,” *Oil & gas journal*, vol. 111, no. 6, pp. 64–73, 2013.

[32] R. Rizzo, S. Gupta, M. Rogowska, and R. Ambat, “Corrosion of carbon steel under CO2 conditions: Effect of CaCO3 precipitation on the stability of the FeCO3 protective layer,” *Corros Sci*, vol. 162, p. 108214, 2020.

[33] S. N. Esmaeely, D. Young, B. Brown, and S. Nešić, “Effect of incorporation of calcium into iron carbonate protective layers in CO2 corrosion of mild steel,” *Corrosion*, vol. 73, no. 3, pp. 238–246, 2017.

[34] L. M. Tavares, E. M. da Costa, J. J. de Oliveira Andrade, R. Hubler, and B. Huet, “Effect of calcium carbonate on low carbon steel corrosion behavior in saline CO2 high pressure environments,” *Appl Surf Sci*, vol. 359, pp. 143–152, 2015.

[35] C. Ding, K. Gao, and C. Chen, “Effect of Ca2+ on CO2 corrosion properties of X65 pipeline steel,” *International Journal of Minerals, Metallurgy and Materials*, vol. 16, no. 6, pp. 661–666, 2009.

[36] X. Liu, J. Zheng, J. Fu, J. Ji, and G. Chen, “Multi-level optimization of maintenance plan for natural gas pipeline systems subject to external corrosion,” *J Nat Gas Sci Eng*, vol. 50, pp. 64–73, 2018.

[37] M. E. A. Ben Seghier, D. Höche, and M. Zheludkevich, “Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques,” *J Nat Gas Sci Eng*, vol. 99, p. 104425, 2022.

[38] M. Askari, M. Aliofkhazraei, and S. Afroukhteh, “A comprehensive review on internal corrosion and cracking of oil and gas pipelines,” *J Nat Gas Sci Eng*, vol. 71, p. 102971, 2019.

[39] A. Kahyarian, M. Achour, and S. Nesic, “Mathematical modeling of uniform CO2 corrosion,” *Trends in Oil and Gas Corrosion Research and Technologies*, pp. 805–849, 2017.

[40] S. Nesic, M. Nordsveen, N. Maxwell, and M. Vrhovac, “Probabilistic modelling of CO2 corrosion laboratory data using neural networks,” *Corros Sci*, vol. 43, no. 7, pp. 1373–1392, 2001.

[41] M. Jones *et al.*, “Numerical exploration of a fully mechanistic mathematical model of aqueous CO2 corrosion in steel pipelines,” *Corros Sci*, vol. 236, p. 112235, 2024.

[42] A. Charles, K. Sivaraj, and S. Thanikaikarasan, “Synthesis, characterization and corrosion studies of Schiff bases derived from pyrrole-2-carbaldehyde,” *Mater Today Proc*, vol. 33, pp. 3135–3138, 2020.

[43] Y. Tang, F. Zhang, S. Hu, Z. Cao, Z. Wu, and W. Jing, “Novel benzimidazole derivatives as corrosion inhibitors of mild steel in the acidic media. Part I: Gravimetric, electrochemical, SEM and XPS studies,” *Corros Sci*, vol. 74, pp. 271–282, 2013.

[44] X. Luo, X. Pan, S. Yuan, S. Du, C. Zhang, and Y. Liu, “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*, vol. 125, pp. 139–151, 2017.

[45] 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.

[46] X. Wu, J. Li, C. Deng, L. Yang, J. Lv, and L. Fu, “Novel carbon dots as effective corrosion inhibitor for N80 steel in 1 M HCl and CO2-saturated 3.5 wt% NaCl solutions,” *J Mol Struct*, vol. 1250, p. 131897, 2022.

[47] W. T. Nash, C. J. Powell, T. Drummond, and N. Birbilis, “Automated corrosion detection using crowdsourced training for deep learning,” *Corrosion*, vol. 76, no. 2, pp. 135–141, 2020.

[48] C. I. Ossai, “A data-driven machine learning approach for corrosion risk assessment—a comparative study,” *Big Data and Cognitive Computing*, vol. 3, no. 2, p. 28, 2019.

[49] G. Sanchez, W. Aperador, and A. Cerón, “Corrosion grade classification: a machine learning approach,” *Indian Chemical Engineer*, vol. 62, no. 3, pp. 277–286, 2020.

[50] M. Rogowska, J. Gudme, A. Rubin, K. Pantleon, and R. Ambat, “Effect of Fe ion concentration on corrosion of carbon steel in CO2 environment,” *Corrosion Engineering, Science and Technology*, vol. 51, no. 1, pp. 25–36, 2016.

[51] L. Yan, Y. Diao, Z. Lang, and K. Gao, “Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach,” *Sci Technol Adv Mater*, vol. 21, no. 1, pp. 359–370, 2020.

[52] M. Aghaaminiha *et al.*, “Machine learning modeling of time-dependent corrosion rates of carbon steel in presence of corrosion inhibitors,” *Corros Sci*, vol. 193, p. 109904, 2021.

[53] S. Nešić, “Key issues related to modelling of internal corrosion of oil and gas pipelines–A review,” *Corros Sci*, vol. 49, no. 12, pp. 4308–4338, 2007.

[54] T. M. Mitchell, “Does machine learning really work?,” *AI Mag*, vol. 18, no. 3, p. 11, 1997.

[55] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. “ O’Reilly Media, Inc.,” 2022.

[56] H. A. Al-Jamimi, S. Al-Azani, and T. A. Saleh, “Supervised machine learning techniques in the desulfurization of oil products for environmental protection: A review,” *Process Safety and Environmental Protection*, vol. 120, pp. 57–71, 2018.

[57] L. Breiman, “Random forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001.

[58] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, “Improved random forest for classification,” *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012–4024, 2018.

[59] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[60] S. R. Gunn, “Support vector machines for classification and regression,” *Technical Report, Image Speech and Intelligent Systems Research Group, 1997*, vol. 1, no. 1, pp. 1–52, 1997.

[61] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans Inf Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[62] S. Xia, Z. Xiong, Y. Luo, L. Dong, and G. Zhang, “Location difference of multiple distances based k-nearest neighbors algorithm,” *Knowl Based Syst*, vol. 90, pp. 99–110, 2015.

[63] Z. Dong *et al.*, “Machine learning-based corrosion rate prediction of steel embedded in soil,” *Sci Rep*, vol. 14, no. 1, p. 18194, 2024.

[64] C. Ding, K. Gao, and C. Chen, “Effect of Ca2+ on CO2 corrosion properties of X65 pipeline steel,” *International Journal of Minerals, Metallurgy and Materials*, vol. 16, no. 6, pp. 661–666, 2009.

[65] S. Navabzadeh Esmaeely, Y.-S. Choi, D. Young, and S. Nešić, “Effect of calcium on the formation and protectiveness of iron carbonate layer in CO2 corrosion,” *Corrosion*, vol. 69, no. 9, pp. 912–920, 2013.

[66] A. Shamsa, R. Barker, Y. Hua, E. Barmatov, T. L. Hughes, and A. Neville, “The role of Ca2+ ions on Ca/Fe carbonate products on X65 carbon steel in CO2 corrosion environments at 80 and 150° C,” *Corros Sci*, vol. 156, pp. 58–70, 2019.

[67] X. Ren *et al.*, “The influence of Ca2+ on the growth mechanism of corrosion product film on N80 steel in CO2 corrosion environments,” *Corros Sci*, vol. 218, p. 111168, 2023.

[68] J. Li, J. Lv, L. Fu, M. Tang, and X. Wu, “New ecofriendly nitrogen-doped carbon quantum dots as effective corrosion inhibitor for saturated CO2 3% NaCl solution,” *Russian Journal of Applied Chemistry*, vol. 93, no. 3, pp. 380–392, 2020.

[69] H. Cen, Z. Chen, and X. Guo, “N, S co-doped carbon dots as effective corrosion inhibitor for carbon steel in CO2-saturated 3.5% NaCl solution,” *J Taiwan Inst Chem Eng*, vol. 99, pp. 224–238, 2019.

[70] M. Aghaaminiha *et al.*, “Machine learning modeling of time-dependent corrosion rates of carbon steel in presence of corrosion inhibitors,” *Corros Sci*, vol. 193, p. 109904, 2021.

[71] T. Wang, J. Li, and Y. Liu, “Based on Pearson correlation coefficient and Monte Carlo simulation method, the calculation formula of frictional resistance is optimized for soft shaft bending and pulling of steel wire,” in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 012088.

[72] K. Gilbert, P. C. Bennett, W. Wolfe, T. Zhang, and K. D. Romanak, “CO2 solubility in aqueous solutions containing Na+, Ca2+, Cl−, SO42− and HCO3-: The effects of electrostricted water and ion hydration thermodynamics,” *Applied Geochemistry*, vol. 67, pp. 59–67, 2016.

[73] S. N. Esmaeely, D. Young, B. Brown, and S. Nešić, “Effect of incorporation of calcium into iron carbonate protective layers in CO2 corrosion of mild steel,” *Corrosion*, vol. 73, no. 3, pp. 238–246, 2017.

[74] R. Rizzo, S. Gupta, M. Rogowska, and R. Ambat, “Corrosion of carbon steel under CO2 conditions: Effect of CaCO3 precipitation on the stability of the FeCO3 protective layer,” *Corros Sci*, vol. 162, p. 108214, 2020.

[75] H. A. Alsaiari, N. Zhang, S. Work, A. Kan, and M. B. Tomson, “A new correlation to predict the stoichiometry of mixed scale: Iron-calcium carbonate,” in *SPE International Oilfield Scale Conference and Exhibition*, SPE, 2012, p. SPE-155127.

[76] M. Cui, S. Ren, H. Zhao, L. Wang, and Q. Xue, “Novel nitrogen doped carbon dots for corrosion inhibition of carbon steel in 1 M HCl solution,” *Appl Surf Sci*, vol. 443, pp. 145–156, 2018.

[77] A. Yakusheva, A. Sayapina, L. Luchnikov, D. Arkhipov, G. Karunakaran, and D. Kuznetsov, “Carbon Quantum Dots’ Synthesis with a Strong Chemical Claw for Five Transition Metal Sensing in the Irving–Williams Series,” *Nanomaterials*, vol. 12, no. 5, p. 806, 2022.

[78] K. Luo and X. Jiang, “Fluorescent carbon quantum dots with Fe (III/II) irons as bridge for the detection of ascorbic acid and H2O2,” *J Fluoresc*, vol. 29, no. 3, pp. 769–777, 2019.

[79] X. Wu, J. Li, C. Deng, L. Yang, J. Lv, and L. Fu, “Novel carbon dots as effective corrosion inhibitor for N80 steel in 1 M HCl and CO2-saturated 3.5 wt% NaCl solutions,” *J Mol Struct*, vol. 1250, p. 131897, 2022.

[80] L. Dong *et al.*, “High-efficiency corrosion inhibitor of biomass-derived high-yield carbon quantum dots for Q235 carbon steel in 1 M HCl solution,” *ACS Omega*, vol. 8, no. 49, pp. 46934–46945, 2023.