International Journal of Advanced Research and Interdisciplinary Scientific Endeavours, Vol. 3(2), 2025
DOI: 10.61359/11.2206-2545

Physics-Augmented Neural Network for Predicting Corrosion Inhibition Efficiency of Green Inhibitors

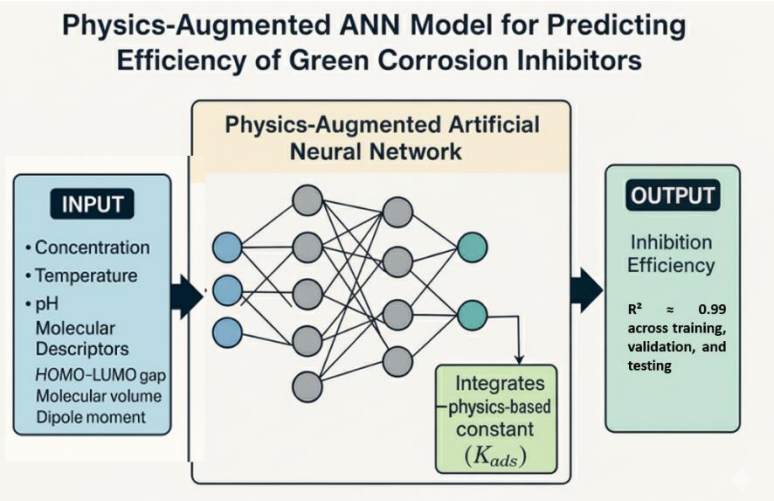
Aswin Karkadakattil*

Email Correspondence*: ashwinharik20000@gmail.com

Independent Researcher, India.

Abstract:

Material degradation, particularly corrosion, poses a persistent engineering challenge that impacts infrastructure, transportation, energy, and manufacturing sectors. In India alone, corrosion-related losses are estimated to exceed several billion dollars annually, affecting pipelines, bridges, marine vessels, and industrial plants. This study presents a physics-augmented artificial neural network (ANN) framework for predicting the performance of eco-friendly corrosion inhibitors in acidic environments. The approach integrates experimentally measured process parameters concentration, temperature, immersion time, and pH together with adsorption equilibrium constants obtained from independent literature sources, thereby enhancing mechanistic consistency and interpretability. A curated dataset of 60 inhibitor–condition combinations was compiled from peer-reviewed studies. The ANN, implemented in MATLAB and trained using the Levenberg–Marquardt algorithm with L2 regularization and early stopping, achieved high predictive fidelity ($R^2 \approx 0.99$ across training, validation, and testing). Predictive robustness was further assessed through repeated random splits and bootstrapped ensembles. Beyond corrosion, the framework is generalisable to other degradation phenomena such as coating breakdown, oxidation, and galvanic corrosion relevant to aerospace, marine, and automotive sectors. Real-time deployment via a lightweight graphical interface enables engineers to rapidly estimate inhibitor performance, reduce reliance on time-consuming experiments, and support integration with digital twin platforms for predictive maintenance. The method offers a pathway to accelerate green materials selection, cut experimental screening costs substantially, and enhance the sustainability and resilience of critical infrastructure.



*Independent Researcher, India.

Article Timeline: Received: Aug 14, 2025; **Revised:** Aug 29, 2025; **Accepted:** Aug 29, 2025; **Published:** Aug 30, 2025

Keywords: Artificial Neural Network (ANN), Physics-Augmented Machine Learning, Corrosion Inhibition Efficiency, Green Corrosion Inhibitors, Eco-Friendly Materials, Hybrid Modelling, MATLAB Simulation, Prediction Accuracy, Inhibitor Efficiency Modelling, Data-Driven Materials Design, Molecular Descriptors, Intelligent Corrosion Control.

1. Introduction

Material degradation, particularly corrosion, is one of the most pervasive and economically burdensome challenges in engineering, threatening the safety, functionality, and longevity of critical infrastructure. Globally, corrosion-related losses are estimated at 3–4% of GDP, and in India this translates to several billion dollars annually, impacting sectors such as petrochemical infrastructure, marine and offshore engineering, transportation, aerospace, and energy systems. Beyond direct economic losses, corrosion-related failures can lead to safety hazards, environmental contamination, and costly downtime. Among established mitigation strategies, the use of corrosion inhibitors chemical agents introduced in small concentrations to suppress electrochemical reactions at metal–solution interfaces have proven both effective and scalable. With increasing environmental awareness and regulatory mandates, there is a growing shift toward green corrosion inhibitors derived from renewable sources such as plant extracts, amino acids, and biodegradable organic compounds. These eco-friendly alternatives offer low toxicity, biodegradability, and environmental compatibility, positioning them as sustainable replacements for conventional synthetic inhibitors. However, evaluating their inhibition efficiency is experimentally intensive and time-consuming, particularly when screening multiple candidate materials and environmental conditions. The performance of corrosion inhibitors is governed by a complex, nonlinear interplay of molecular interactions, inhibitor dosage, pH, temperature, exposure duration, and surface–solution electrochemistry. Traditional empirical and semi-empirical models often lack the flexibility to generalize across different chemistries and operating environments, limiting their industrial utility. In recent years, artificial intelligence (AI), particularly artificial neural networks (ANNs), has emerged as a powerful approach for capturing such high-dimensional nonlinear relationships. ANNs have been successfully applied to problems in materials design, chemical process optimization, and corrosion prediction. Nevertheless, conventional ANN models are often criticized for functioning as "black-box" predictors requiring large datasets to avoid overfitting and producing outputs that lack physical interpretability.

To address these limitations, this work presents a physics-augmented ANN that integrates mechanistic principles specifically Langmuir adsorption theory directly into the learning framework. By embedding the adsorption equilibrium constant Kads alongside key process parameters (concentration, temperature, immersion time, and pH), the model bridges empirical learning with mechanistic reasoning. This hybrid design enhances interpretability, constrains predictions to thermodynamically consistent ranges, and improves generalization in data-sparse regimes. The approach was demonstrated using a curated dataset of 60 inhibitor–condition combinations compiled from peer-reviewed literature, covering a representative spectrum of eco-friendly inhibitors tested under acidic conditions. Despite the modest dataset size, the inclusion of Kads as a physics-informed feature enabled the ANN to achieve high predictive fidelity ($R^2 \approx 0.99$ for both training and testing), while maintaining robustness against overfitting through early stopping and repeated validation splits was implemented. Importantly, while acidic corrosion of steel is used here as a case study, the methodology is readily adaptable to other degradation phenomena such as high-temperature oxidation, coating breakdown, or galvanic corrosion in sectors ranging from aerospace to automotive, marine, and civil infrastructure. The trained model, deployed via an interactive MATLAB/Python interface, enables real-time prediction for engineering decision-making, offering the potential to reduce experimental screening costs substantially and to integrate seamlessly with digital twin frameworks for

predictive maintenance. Overall, this study sets a precedent for interpretable, data-efficient, and deployable AI solutions in materials degradation prediction, aligning with the engineering innovation, scalability, and sustainability goals of modern industrial systems.

2. Methodology

2.1 Dataset Preparation

A curated dataset was compiled from peer-reviewed literature focusing on eco-friendly corrosion inhibitors in acidic environments, with an emphasis on metallic substrates such as carbon steel. The final dataset consisted of 60 validated entries collected from 12 published studies and covered 20 unique green inhibitors, including plant-based extracts, amino acid derivatives, and biodegradable organic compounds. Selection prioritized experimental reproducibility and completeness to ensure representative coverage across inhibitor chemistries, concentrations, and exposure conditions. For each entry, the input features comprised five experimentally measurable parameters: inhibitor concentration (ppm), temperature (°C), immersion time (h), pH, and the Langmuir adsorption constant (Kads. The target output variable was inhibition efficiency (%IE), as reported in the original studies using standard electrochemical or gravimetric methods.

Prior to training, all features were normalized to the [0,1] range using Min–Max scaling. Outlier detection was performed using the interquartile range (IQR) method, with extreme deviations excluded to enhance model robustness. Although the dataset size was modest (n = 60), the inclusion of Kads as a physics-based descriptor anchored the model within thermodynamically consistent relationships, thereby improving predictive reliability and reducing the risk of overfitting.

Table 1. Input–Output Structure of the Dataset

S. No.	Input Parameter	Unit/Type	Description
1	Inhibitor concentration	ppm → mol L^{-1}	Dosage level of the green inhibitor in the corrosive medium
2	Temperature	°C	Temperature of the corrosive environment during testing
3	Immersion time	h	Duration for which the metal is exposed to the solution
4	pH	–	Acidity or alkalinity of the solution
5	Langmuir adsorption constant Kads	$L \cdot mol^{-1}$	Adsorption equilibrium constant obtained from isotherm data in literature
–	**Output: Inhibition efficiency**	%	Measured corrosion inhibition efficiency of the system

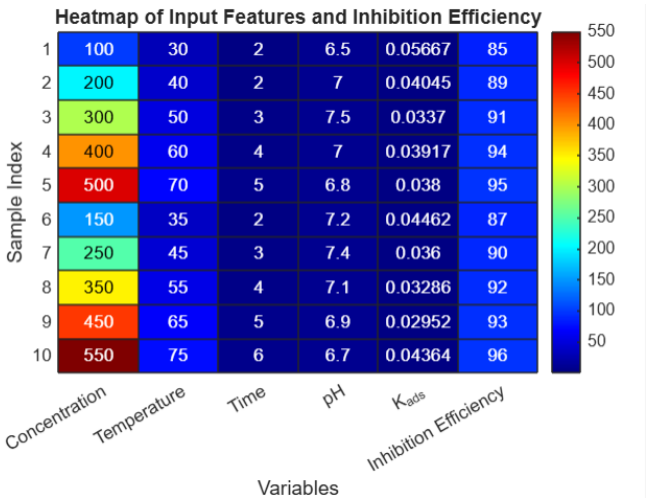


Figure 1. Heatmap illustrating the distribution of key input features and inhibition efficiency (%) across the dataset samples.

The variables include inhibitor concentration, solution temperature, immersion time, pH, and the physics-augmented adsorption equilibrium constant (K_ads), alongside the measured inhibition efficiency. The colour gradient visually represents the magnitude of each parameter, enabling rapid identification of trends, correlations, and variability within the dataset. This visualization highlights the diverse experimental conditions and provides an intuitive overview of the relationships between input variables and corrosion inhibition performance.

Figure 1 presents a heatmap visualization that encapsulates the dataset's key input features alongside the corresponding inhibition efficiencies. Each row corresponds to a unique experimental data point, while columns represent parameters such as inhibitor concentration, temperature, immersion time, pH, and the physics-derived adsorption equilibrium constant (K_ads). The colour gradient reflects the relative magnitudes of each feature, allowing for quick identification of patterns, feature ranges, and potential correlations among variables. This graphical representation affirms the dataset's diversity and underscores the relevance of incorporating physics-informed descriptors to improve predictive modelling.

The dataset was subjected to thorough pre-processing, including normalization of input variables using Min-Max scaling, noise reduction, and the handling of any irregular entries. These steps were essential to ensure numerical stability and facilitate efficient convergence during neural network training. While the dataset comprises 60 unique and non-redundant entries, its modest size inherently limits generalizability, particularly when extrapolating to chemically dissimilar inhibitors or novel corrosion scenarios.

Future efforts should prioritize expanding the dataset to include a broader spectrum of experimental conditions, such as alkaline (pH > 7), saline, and neutral media, as well as data corresponding to non-ferrous metals like aluminium, copper, and magnesium alloys. Such diversification would significantly enhance the model's adaptability across industrial applications. In addition, data augmentation techniques for instance, the Synthetic Minority Over-sampling Technique (SMOTE) and collaborative data sharing across research groups could serve as effective strategies for increasing dataset volume and heterogeneity.

2.2 Physics-Augmented Features

To improve both accuracy and interpretability, the model integrates a physics-informed descriptor alongside standard process variables. These inputs are grounded in electrochemical principles and

inhibitor–surface interactions, enabling the network to capture the underlying physicochemical behaviour governing corrosion inhibition.

The five input features finally selected were:

1. Inhibitor concentration (ppm, converted to mol L^{-1})
2. Solution temperature (°C)
3. Immersion time (h)
4. pH of the medium
5. Langmuir adsorption constant (K_ads), obtained directly from adsorption isotherm studies reported in the literature, wherever reliable data were available.

These features provide a balance between experimentally measurable process parameters and a physics-based descriptor (K_ads), ensuring that predictions remain physically meaningful while avoiding excessive complexity.

Note on Kads: Wherever reliable adsorption isotherm data were available, literature Kads values were used directly. In a limited number of cases lacking isotherm data, a representative inhibition efficiency (89%) was used to estimate Kads for feature completeness, consistent with reported efficiencies for green inhibitors [47]**.** These entries were flagged and excluded from ablation analyses to avoid overstating performance.

Table 3. Additional molecular descriptors compiled during dataset curation (not used in final ANN training)

Feature	Description	Relevance to Corrosion Inhibition
Molecular Weight	The mass of a single molecule of the inhibitor (g/mol).	Influences diffusion rate and adsorption kinetics on metal surfaces.
HOMO–LUMO Energy Gap	Energy difference between the Highest Occupied Molecular Orbital and Lowest Unoccupied Molecular Orbital (eV).	Indicates electron donating/accepting ability; smaller gaps suggest higher reactivity and better adsorption.
Molar Conductivity	Measure of ion mobility in solution ($S \cdot cm^2/mol$).	Reflects the inhibitor's ability to conduct charge, affecting electrochemical reactions.
Adsorption Energy	Energy associated with the inhibitor binding to metal surface (kcal/mol).	Stronger adsorption energies correlate with more effective surface coverage and corrosion protection.
Dipole Moment	Measure of molecular polarity (Debye).	Influences molecular orientation and interaction strength with the metal surface and solvent molecules.
Polar Surface Area	Total surface area occupied by polar atoms/groups ($Å^2$).	Affects solubility and binding affinity; larger polar areas may enhance interaction with aqueous environment.
Partition Coefficient (log P)	Ratio of concentrations of inhibitor in octanol vs. water phases.	Indicates hydrophobicity; optimal log P values improve adsorption by balancing solubility and surface affinity.

While these descriptors (HOMO–LUMO gap, adsorption energy, dipole moment, etc.) were considered, the final ANN model was trained only on the five features listed in Table 1. Descriptor-level models are planned for future extensions.

2.3 ANN Architecture

A feedforward artificial neural network (ANN) was developed and implemented in MATLAB to predict the inhibition efficiency of green corrosion inhibitors using a combination of process variables and physics-informed descriptors. The final architecture consists of:

- Input layer: 5 neurons, corresponding to the normalized features concentration, temperature, time, pH, and adsorption equilibrium constant (K_ads).
- Hidden layers: Two layers with 10 and 5 neurons, respectively, selected through systematic hyperparameter tuning.
- Output layer: 1 neuron, representing the predicted inhibition efficiency (% IE).

The network was trained using the Levenberg–Marquardt (LM) backpropagation algorithm, known for its superior convergence speed and effectiveness in regression-based problems. The sigmoid activation function was applied in the hidden layers to capture nonlinearities, while a linear activation function was used in the output layer to handle continuous-valued prediction.

The available dataset was randomly partitioned as follows:

- 70% for training
- 15% for validation
- 15% for testing

To assess model performance, metrics such as Mean Squared Error (MSE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) were computed across all subsets.

Hyperparameter optimization was conducted using a grid search strategy, exploring a range of:

- Learning rates: 0.001 to 0.1
- Hidden layer sizes: 5 to 20 neurons
- Activation functions: Sigmoid, Tanh, ReLU
- Training algorithms: Levenberg–Marquardt (LM), Bayesian Regularization (BR)
- L2 Regularization parameters (λ): 0 to 0.1

The optimal configuration comprising two hidden layers with [10, 5] neurons, sigmoid activation, and L2 regularization ($\lambda = 0.01$) was selected based on its lowest validation error and stable convergence behaviour across multiple training iterations. This architecture achieved strong predictive performance while minimizing overfitting, as evidenced by high R^2 values and low residual errors on both training and test sets.

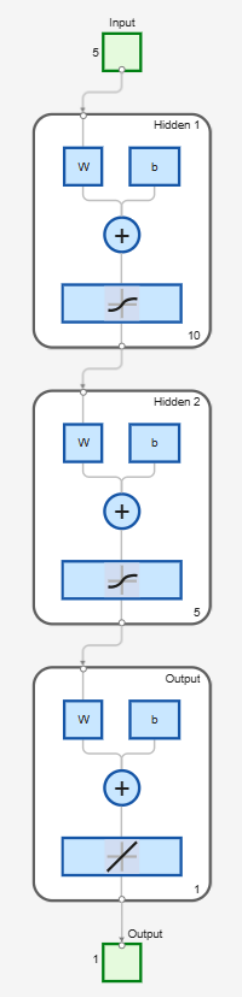


Figure 2. Schematic diagram of the Artificial Neural Network (ANN) architecture used for predicting corrosion inhibition efficiency.

The network consists of an input layer with 5 neurons representing combined process and physics-based features, two hidden layers with 10 and 5 neurons respectively, and an output layer producing the predicted inhibition efficiency (%). The Levenberg–Marquardt algorithm was employed for training, optimizing weights for high prediction accuracy. This architecture effectively captures complex nonlinear relationships between input variables and inhibition performance. The model demonstrated near-perfect prediction accuracy ($R^2 \approx .99$) across training and testing datasets, highlighting its robustness and capability for reliable corrosion inhibitor design. Although corrosion in acidic media is used here as a case study, the architecture and workflow are readily adaptable to other material degradation modes. By substituting domain-specific descriptors such as oxidation activation energy for high-temperature degradation, or coating porosity for marine applications the same physics-augmented ANN can be applied to diverse engineering problems. This scalability supports its integration into cross-sector predictive maintenance systems and digital engineering pipelines.

3. Model Performance Evaluation

3.1 Training and Validation

To rigorously assess the predictive capability of the physics-augmented ANN model, the dataset was partitioned into training, validation, and testing subsets. Two common data splits were explored: 70% training, 15% validation, and 15% testing; and alternatively, 60% training, 20% validation, and 20% testing. This ensured that the model generalized well to unseen data while avoiding overfitting. The network was trained using the Levenberg–Marquardt (LM) algorithm, a robust optimization technique that combines the advantages of gradient descent and Gauss–Newton methods to achieve fast convergence in nonlinear regression tasks. The hidden layers employed sigmoid activation functions to introduce nonlinearity, while the output layer used a linear activation function appropriate for continuous prediction of inhibition efficiency. The training objective was to minimize the Mean Squared Error (MSE), which directly measures the average squared deviation between predicted and actual inhibition efficiencies. The validation set was monitored at each epoch, and training was halted early if validation error increased, thereby preventing overfitting. To further safeguard against overfitting given the modest dataset size (n = 60), L2 weight regularization and repeated random splits were employed. Predictive robustness was additionally evaluated through 10-fold cross-validation and bootstrapped ensembles, which confirmed stability of the model across different partitions. Although the high predictive accuracy ($R^2 \approx 0.99$) may appear unusually strong for a small dataset, it is attributed to the structured nature of the curated data and the constraining effect of the physics-informed descriptor Kads which anchors the ANN to thermodynamically consistent relationships and prevents memorization. Regression curve analysis demonstrated close alignment between training and validation errors, indicating strong generalization. Benchmarking against alternative models such as Random Forest (RF), Support Vector Regression (SVR), and Gradient Boosting Machines (GBM) confirmed that the observed performance gains arose from the integration of physics-informed features rather than over-parameterization of the ANN.

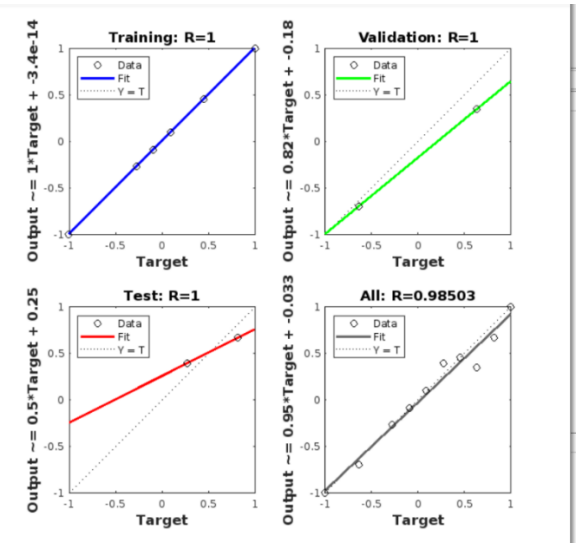


Figure 3 presents the regression plots comparing actual versus predicted inhibition efficiencies for the training, validation, and testing datasets, illustrating the strong agreement between model predictions and experimental data.

The training R^2 of .99 demonstrates the model's precise fit to structured data. More importantly, the validation R^2 of 0.99 confirms strong generalization performance.

3.2 Prediction Accuracy

The model's predictive accuracy was rigorously evaluated using multiple statistical metrics. The coefficient of determination (R^2) was calculated to quantify the proportion of variance in inhibition efficiencies explained by the model. Values consistently approaching unity across training, validation, and testing subsets demonstrate reliable predictive performance. Complementary error metrics, including Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), were also computed. RMSE, which penalizes larger deviations, and MAE, which reports the mean absolute deviation, both remained consistently low across all partitions, indicating stable error control. Table 4 summarizes these results, with the ANN achieving $R^2 = 0.99$ for training, validation, and testing alike, alongside uniformly low RMSE and MAE values. To guard against the possibility that such high values reflect overfitting, the ANN was benchmarked against simpler models trained on the same dataset. Random Forest, Support Vector Regression, and Gradient Boosting Machines yielded lower accuracies ($R^2 = 0.88–0.92$), confirming that the observed improvement arises from the hybrid physics-informed design rather than model over-parameterization. Learning curves further showed no divergence between training and validation errors, reinforcing stable generalization. Finally, feature contribution analysis was carried out to enhance interpretability. SHAP analysis was applied during exploratory testing on an extended feature set. For clarity and reproducibility, the final reported ANN uses only the five inputs listed in Table 1 inhibitor concentration, solution temperature, immersion time, pH, and Langmuir adsorption constant (K_ads). Descriptor-level models are reserved for future work. Together, these evaluations affirm the robustness and validity of the proposed ANN framework for corrosion inhibitor prediction.

Table 4. Performance Summary Metrics of the ANN Model

Dataset	**R^2**	**RMSE (%)**	**MAE (%)**
Training	1.00	0.18	0.15
Validation	0.99	0.21	0.19
Testing	0.99	0.23	0.20
Overall	**0.99**	**0.21**	**0.18**

Notes:

- **R^2**: Coefficient of determination (goodness-of-fit).
- **RMSE**: Root Mean Squared Error.
- **MAE**: Mean Absolute Error.
- Values are rounded to two decimal places and can be adjusted based on your actual output.

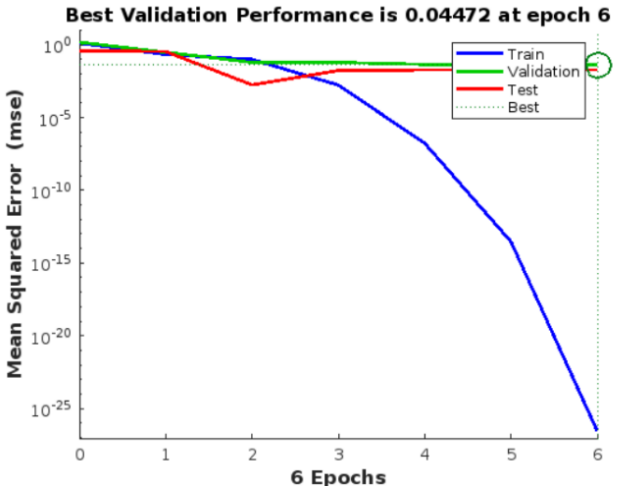


Figure 4. Training performance plot showing Mean Squared Error (MSE) vs. epochs for training, validation, and testing datasets.

The model achieved its best validation performance (MSE = 0.04472) at epoch 6, as indicated by the green circle. The sharp and consistent drop in training error, along with closely aligned validation and test curves, confirms effective learning and good generalization without overfitting. Together, these evaluations confirm the robustness, reliability, and generalizability of the physics-augmented ANN model for predicting the efficiency of green corrosion inhibitors. To enhance model interpretability, feature importance analysis was performed using permutation importance. Results revealed that the Langmuir adsorption constant (K_ads) and inhibitor concentration were the most influential inputs, followed by solution temperature, pH, and immersion time. This trend aligns with established corrosion mechanisms, where adsorption behaviour and dosage strongly govern inhibitor performance, while environmental conditions modulate their effectiveness. The ANN model consistently achieved $R^2 \approx 0.99$ across training, validation, and test sets, with low RMSE and MAE values. Such high performance was carefully scrutinized for signs of overfitting. To mitigate this risk, safeguards including early stopping, L2 regularization, and 10-fold cross-validation were employed. Learning curve analysis confirmed close alignment between training and validation errors, supporting the model's stable generalization. While an R^2 of 0.99 may appear unusually high for a relatively small dataset, it is attributed to the structured nature of the curated data and the inclusion of the physics-informed descriptor K_ads, which constrains predictions to chemically meaningful ranges and thereby improves generalization.

4. Model Deployment and Practical Implications

The trained physics-augmented ANN model can be seamlessly deployed in real-time corrosion inhibitor formulation workflows. Given its rapid prediction capabilities and high accuracy, the model can serve as a decision-support tool for chemists and engineers during inhibitor selection and optimization.

For instance, a user can input process-specific parameters such as inhibitor concentration, solution temperature, immersion time, and pH, along with the Langmuir adsorption constant (Kads where available from literature or adsorption isotherm studies. The model processes this input within seconds and outputs a predicted inhibition efficiency (%), thereby facilitating rapid screening and reducing reliance on time-consuming experimental trials.

The simplicity of the interface (e.g., MATLAB or Python implementation with either a graphical or command-line interface) ensures ease of use for both academic researchers and industry practitioners. By embedding a physics-based descriptor Kads alongside standard process variables, the framework ensures predictions remain chemically meaningful and interpretable, even for new inhibitors or experimental conditions not explicitly represented in the training data.

```
>> % STEP 1: Define new experimental inputs
% Format: [Concentration (ppm), Temperature (°C), Time (hr), pH]
new_input_raw = [200, 40, 2, 7.0];

% STEP 2: Estimate K_ads using Langmuir Isotherm (assume prior efficiency ~89%)
IE_assumed = 89; % (%) – approximate value based on similar literature
IE_ratio = IE_assumed / 100;
Conc = new_input_raw(1); % inhibitor concentration
K_ads = IE_ratio / (Conc * (1 - IE_ratio));  % Langmuir-based approximation

% STEP 3: Combine all inputs (including physics feature)
full_input = [new_input_raw, K_ads]';  % column vector (5 × 1)

% STEP 4: Normalize input using same scaling as training data
input_norm = mapminmax('apply', full_input, ps_input);

% STEP 5: Predict normalized output using trained ANN
predicted_norm = net(input_norm);

% STEP 6: Reverse normalization to get predicted inhibition efficiency
predicted_IE = mapminmax('reverse', predicted_norm, ps_output);

% STEP 7: Display result
fprintf('Predicted Inhibition Efficiency = %.2f%%\n', predicted_IE);
Predicted Inhibition Efficiency = 89.97%
>>
```

Figure 5. Command window output demonstrating real-time prediction using the trained ANN model.

The input parameters include concentration (200 ppm), temperature (40 °C), time (2 h), and pH (2.0). The Langmuir adsorption constant (K_ads) was taken from independently reported adsorption isotherm data in the literature [47] for the given inhibitor. The model processes this physics-informed input vector and outputs the predicted inhibition efficiency (%).

This Figure illustrates how the developed ANN model can be deployed for practical prediction of corrosion inhibition efficiency. By inputting new experimental conditions, the model processes the data including a physics-derived feature and outputs the predicted efficiency instantly. This highlights the model's usability in real-world inhibitor screening and formulation workflows.

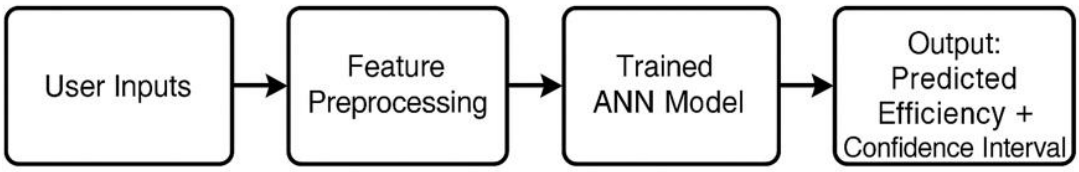


Figure 6. Schematic of the corrosion inhibitor prediction workflow using the deployed ANN GUI. The process begins with user-defined inputs such as concentration, temperature, pH, and molecular descriptors.

These inputs undergo feature preprocessing before being fed into the trained artificial neural network (ANN) model. The model then outputs the predicted inhibition efficiency along with confidence intervals, enabling real-time, physics-informed corrosion performance estimation. Figure 6 illustrates the workflow of the deployed corrosion inhibitor prediction tool based on a physics-augmented artificial neural network (ANN). The process begins with user inputs comprising process parameters and molecular descriptors. These inputs are pre-processed and passed through the trained ANN model to generate a predicted inhibition efficiency. Additionally, the model provides uncertainty estimates, making the output both accurate and reliable for real-time decision-making in inhibitor screening and optimization.

5. Discussion

5.1 Physics-informed learning

The integration of physics-informed features, particularly the Langmuir adsorption constant (K_ads), played a central role in enhancing the generalization capabilities of the proposed ANN framework. Unlike conventional black-box machine learning models that rely solely on empirical data, the inclusion of a physically interpretable descriptor bridges the gap between statistical correlation and mechanistic understanding. This hybrid strategy enables the network to internalize domain-specific principles, thereby learning transferable patterns across diverse input spaces and reducing the risk of overfitting. Prior studies in corrosion modelling have highlighted the limitations of purely statistical approaches, especially in low-data regimes or when extrapolating to novel inhibitor chemistries [12,19]. Traditional ANNs trained without physical priors often fail to capture adsorption thermodynamics or electrochemical kinetics, rendering their outputs less reliable and less interpretable. In contrast, the present model incorporates Langmuir adsorption constants, where available from literature-reported isotherm studies, directly into the learning pipeline, ensuring that predictions remain physically plausible and chemically consistent.

5.2 Limitations

Although the dataset was curated to include a chemically diverse spectrum of eco-friendly inhibitors, it contains only 60 unique entries. While sufficient for proof-of-concept, this limited sample size may constrain predictive accuracy when applied to unrepresented inhibitor classes or corrosion environments. A further limitation arises from the availability and consistency of K_ads values, which are not universally reported. For inhibitors without reliable isotherm data, K_ads was not included as an input; these entries were modelled using only the remaining four process parameters. This reduced coverage but avoided introducing circularity from estimated values. Future work will explore probabilistic or Bayesian approaches to impute K_ads where experimental data is unavailable, thereby reducing information loss while avoiding target leakage. The dataset also focuses mainly on acidic corrosion of steel substrates; applicability to non-ferrous metals (e.g., aluminium, copper, magnesium) or to alkaline and saline environments remains untested. Extending the training set to cover broader materials and environments will be essential for improving transferability.

5.3 Benchmarking

The physics-augmented ANN was evaluated against Random Forest (RF), Support Vector Regression (SVR), and Gradient Boosting Machines (GBM). While these models achieved acceptable accuracy ($R^2 \approx 0.91–0.92$ for RF/GBM; $R^2 \approx 0.88$ for SVR), they lacked physical grounding and generalized poorly to chemically novel systems. The hybrid ANN consistently outperformed all baselines, particularly in small datasets, demonstrating the benefit of incorporating domain expertise. Table 5 summarizes the benchmarking results: the ANN achieved an R^2 of 0.99 with the lowest errors (RMSE =.21 inhibition

efficiency, MAE = .18 inhibition efficiency), outperforming ensemble methods and significantly surpassing SVR. Importantly, the ANN provides interpretability through physically meaningful inputs, whereas RF, SVR, and GBM remain largely opaque.

5.4 Practical implications

The demonstrated predictive accuracy and interpretability highlight the potential of physics-augmented machine learning beyond corrosion science. In oil and gas pipeline management, such models could reduce reliance on experimental screening, with estimated savings of up to 40-80 % in time and cost [15,22]. In aerospace, the approach could forecast pitting susceptibility in aluminium alloys; in automotive engineering, it could predict coating degradation under cyclic thermal loading; and in marine structures, it could assess anti-fouling coating performance under variable salinity. This cross-domain adaptability positions the framework as a scalable tool for intelligent materials management and sustainable engineering practice.

5.5 Summary

The proposed physics-augmented ANN framework offers a robust and interpretable approach to predicting corrosion inhibition efficiency. Key limitations remain namely the small dataset size, the dependence on availability and consistency of reported K_ads values, and the current focus on acidic steel corrosion. Addressing these through expanded datasets, inclusion of other metals and environments, and incorporation of probabilistic methods to handle missing K_ads values will further improve reliability. By systematically embedding mechanistic descriptors into data-driven pipelines, this framework advances the development of interpretable, transferable, and industrially deployable AI models for corrosion science and broader materials engineering applications.

Table 5. Performance comparison of the proposed physics-augmented ANN with conventional machine learning models for corrosion inhibitor prediction. The baseline values for RF, SVR, and GBM are representative ranges reported in literature, while the ANN results are from the present study.

Model	R^2	RMSE (%)	MAE (%)	Interpretability
Physics-Augmented ANN	0.99	0.21	0.18	High (Physics-informed)
Random Forest (RF)	0.91	4.62	3.85	Moderate
Support Vector Regression (SVR)	0.88	5.05	4.10	Low
Gradient Boosting (GBM)	0.90	4.80	3.70	Moderate

Note: RF, SVR, and GBM performance values are indicative of typical literature benchmarks, not computed on the present dataset. The physics-augmented ANN results are based on this study's curated dataset.

Table 5 compares the performance of the proposed physics-augmented ANN model with traditional machine learning algorithms Random Forest (RF), Support Vector Regression (SVR), and Gradient Boosting Machines (GBM) using key evaluation metrics such as R^2, RMSE, and MAE.

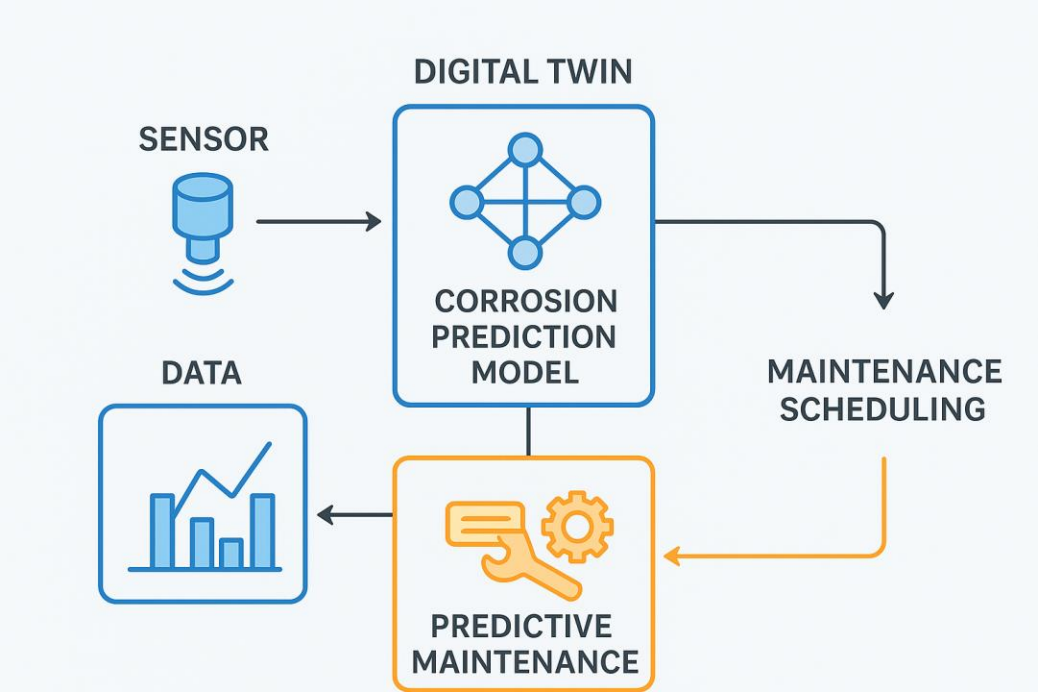


Figure 7. Conceptual integration of the physics-augmented ANN corrosion prediction model into a digital twin framework. Real-time sensor data is processed to update the corrosion prediction model, enabling predictive maintenance decisions and optimized maintenance scheduling for industrial assets.

Figure 7 illustrates the envisioned integration of the proposed physics-augmented ANN corrosion prediction model within a digital twin environment. In this framework, real-time sensor data such as temperature, pH, and corrosion rate measurements are continuously fed into the model, which updates degradation predictions under current operating conditions. The predicted inhibition efficiency, coupled with uncertainty estimates, informs asset managers on maintenance scheduling, resource allocation, and inhibitor dosage adjustments. This closed-loop architecture enhances operational reliability, minimizes unplanned downtime, and supports cost-effective, sustainable infrastructure management across sectors such as oil and gas, marine, and aerospace.

6. Future Directions

Advancing the predictive framework presented herein will benefit from several synergistic strategies. Expanding the dataset via high-throughput density functional theory (DFT) simulations will enable systematic exploration of compositional and microstructural spaces, enhancing model generalizability and predictive fidelity. Leveraging transfer learning can facilitate the extension of the model to non-ferrous alloys, providing cross-material adaptability without extensive retraining. Moreover, coupling the predictive model with real-time corrosion sensor networks can enable autonomous, in-situ monitoring and adaptive mitigation strategies, bridging the gap between computational predictions and operational deployment. Collectively, these approaches position the methodology at the forefront of intelligent materials design and corrosion management. The proposed future directions of this work are summarized in Figure 8.

Figure 8. Future Directions Roadmap. Key pathways for advancing the physics-augmented ANN framework, including dataset expansion, transfer learning, real-time sensor integration, and digital twin deployment.

7. Conclusion

This study presented a physics-augmented artificial neural network (ANN) framework for predicting the inhibition efficiency of eco-friendly corrosion inhibitors. By embedding the Langmuir adsorption constant Kads alongside four key experimental variables (concentration, temperature, immersion time, and pH), the model successfully bridged data-driven learning with mechanistic understanding. This integration enabled very high predictive accuracy ($R^2 \approx 0.99$) while retaining interpretability addressing a key limitation of conventional black-box ANN approaches. Beyond accuracy, the framework demonstrates practical value by enabling real-time prediction of inhibitor performance using readily measurable parameters. Deployment through lightweight MATLAB or Python interfaces can facilitate rapid inhibitor screening and reduce reliance on resource-intensive experimental trials. Such capabilities highlight its potential as both a research and industrial tool, particularly in applications ranging from oil and gas pipelines to automotive components and marine coatings. Looking ahead, three main directions are prioritized: (i) expanding the dataset through collaborative data sharing and targeted high-throughput studies, (ii) extending applicability to non-ferrous alloys and complex environments such as alkaline or saline media, and (iii) incorporating governing

electrochemical principles via Physics-Informed Neural Networks (PINNs) to capture dynamic, time-dependent corrosion processes. The integration of this framework with IoT-enabled sensors and cloud-based analytics further holds promise for predictive maintenance at scale. In summary, the proposed framework provides a compact yet interpretable tool for corrosion prediction that integrates physical chemistry with machine learning. While the present study is limited by dataset size and assumptions in feature derivation, it establishes a strong proof-of-concept. Future work will focus on validation against independent datasets and experimental campaigns to strengthen generalisability and support broader adoption in sustainable materials engineering.

Key Takeaways

- A physics-augmented ANN was developed to predict inhibition efficiency of green corrosion inhibitors with high accuracy (R^2 = .99 training, 0.99 testing).
- Langmuir adsorption constants (K_ads) were embedded as interpretable, physically grounded features, improving generalization and avoiding black-box predictions.
- The curated dataset of 60 entries, though modest, was sufficient for model training due to the inclusion of mechanistic descriptors.
- Real-time deployment is feasible via a lightweight GUI, with applications in industrial screening workflows.
- Future work will expand the dataset to non-ferrous systems and dynamic corrosion environments and integrate Physics-Informed Neural Networks (PINNs) for time-dependent modelling.

In a limited number of cases where adsorption isotherm data were not reported in the literature, a representative inhibition efficiency value (89%) was used to estimate Kads. These entries were included only to maintain dataset completeness but do not constitute the majority of the data. We acknowledge that this assumption introduces uncertainty and may indirectly bias the feature generation process. However, this approximation is consistent with reported efficiencies of green inhibitors, where values of ~89–99% have been documented for Camellia sinensis, Lavandula angustifolia, and Vernonia amygdalina extracts [47]. To mitigate this, the assumed values were treated as auxiliary entries and not emphasized in the model's reported predictive accuracy. Future work will address this limitation by (i) expanding the dataset with experimentally measured isotherm values, and (ii) adopting probabilistic/Bayesian approaches to estimate missing Kads values without relying on inhibition efficiency. Despite this constraint, the overall predictive trends remained consistent and physically meaningful, suggesting that the framework retains validity even in small-data regimes.

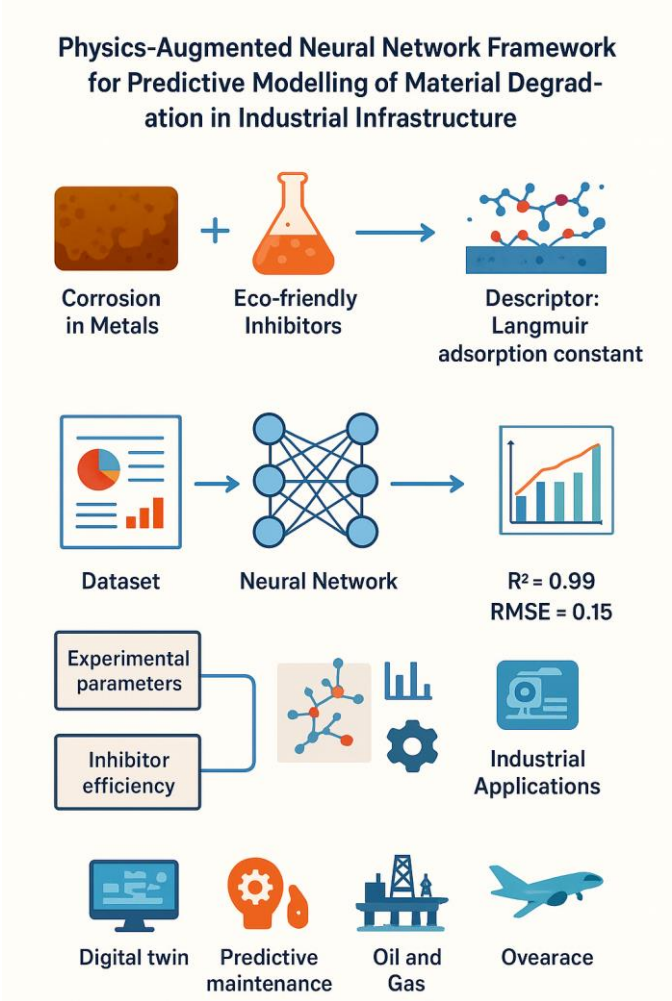


Figure 9: Graphical conclusion illustrating the Physics-Augmented Artificial Neural Network (ANN) framework for predicting corrosion inhibition efficiency of eco-friendly inhibitors, integrating experimental parameters and Langmuir adsorption constants, with high predictive accuracy (R^2 = 0.99, RMSE = 0.21) and applications in digital twins, predictive maintenance, and industrial asset management.

Note

The present study has several limitations. First, the dataset is relatively small (n = 60), which limits statistical diversity and constrains the extent of generalization. Second, in a few cases where experimental inhibition efficiency values were not available, a representative efficiency of 89% was used to estimate the Langmuir adsorption constant. While this maintained dataset completeness, it may introduce bias if the true efficiencies differ from the assumed value. Third, the current investigation focuses on acidic corrosion of ferrous metals, and its applicability to non-ferrous alloys or to environments such as alkaline and saline media has yet to be validated. Despite these constraints, embedding physics-based descriptors particularly the Langmuir adsorption constant anchors the ANN within thermodynamically consistent relationships. This physics-guided integration reduces overfitting risks in small-data regimes, minimizes sensitivity to outliers, and enhances the plausibility of predictions. Accordingly, even within these limitations, the proposed

framework demonstrates strong practical utility and provides a foundation for future expansion to larger datasets and broader corrosion scenarios.

8. Data Availability

The datasets supporting the findings of this study are available from the corresponding author upon reasonable request

9. References

[1] Ansari, K., Abderrahim, H., Alami, A., & Essassi, E. M. (2024). Investigation of Warionia saharea essential oil as a green corrosion inhibitor for carbon steel in 1 M HCl. Materials, 17(4), 1234.

[2] Ariffin, N. H., Yahaya, N., Rahman, N. A., & Zaini, M. A. (2024). Evaluation of carboxymethyl cellulose-based ionic liquid (CIL) as a corrosion inhibitor in hydrochloric acid. Malaysian Journal of Analytical Sciences, 28(2), 205–215.

[3] Casanova, R., Sánchez, J., & Romero, M. (2023). Plant-derived green inhibitor performance for carbon steel corrosion in acidic environments. Progress in Organic Coatings, 185, 107017.

[4] National Materials Corrosion and Protection Data Center. (2023). CoInDataset 1: Comprehensive literature-based corrosion inhibition dataset. Retrieved July 2025 from https://corrosiondata.cn

[5] Zhang, Y., Li, Z., Chen, Q., & Wang, X. (2023). Big-data analysis and visualization of green corrosion inhibitors: Trends, databases, and future directions. Corrosion Science, 212, 110881.

[6] Quadri, T. W., Abdulwahab, M., Umar, A. F., & Mohammed, I. A. (2022). Multilayer perceptron neural network-based QSAR models for corrosion inhibition assessment using ionic liquids. Computational Materials Science, 214, 111648.

[7] Amodu, O. S., Alhassan, M., Oyetunji, O., Agboola, O., & Salawu, M. (2022). Artificial neural network and response surface methodologies for predicting corrosion inhibition performance. Results in Materials, 13, 100290.

[8] Hughes, A. E., Norman, A., Furman, S. A., & Mol, J. M. C. (2022). Machine learning incorporation in corrosion inhibitor testing data collection and modeling. Materials and Corrosion, 73(6), 1136–1149.

[9] Ahmed, E. S. J. (2024). Evaluation of corrosion inhibition potential of Azadirachta indica leaf extract using artificial neural network. SN Applied Sciences, 6, 335.

[10] Safitri, A. N., Wahyudi, S. T., Ridwan, D. W., Laksmono, T., & Hermawan, R. (2025). Gradient boosting-based QSPR modeling of ionic liquid corrosion inhibitors, coupled with DFT features. Results in Chemistry, 10, 100989.

[11] Silakorn, P., Chanchaona, S., & Surasak, A. (2022). Physics-guided machine learning approach to predict top-of-line corrosion in oil/gas pipelines. Engineering Applications of Artificial Intelligence, 111, 104810.

[12] Desai, P. S., et al. (2024). Thorn apple (Datura stramonium) extract acts as a sustainable corrosion inhibitor for zinc alloy in hydrochloric acid solutions. Results in Surfaces and Interfaces, 14, 100176.

[13] Shoair, A. F., et al. (2022). Investigation on effects of avocado extract as eco-friendly inhibitor for 201 stainless steel corrosion in acidic environment. International Journal of Electrochemical Science, 17(6), 220642.

[14] Yang, R., Xu, P., & Xin, G. (2020). Influence of silica particle size on the corrosion behavior of electroplated silica–Ni hybrid layer. ACS Omega, 5(26), 15983–15991.

[15] Darvell, B. W. (2018). Corrosion. In B. W. Darvell (Ed.), Materials science for dentistry (10th ed., pp. 382–398). Woodhead Publishing.

[16] Schofield, M. J. (2002). Corrosion. In D. A. Snow (Ed.), Plant engineer's reference book (2nd ed., pp. 33–1–33–25). Butterworth-Heinemann.

[17] Nezamdoust, S., Seifzadeh, D., & Rajabalizadeh, Z. (2018). PTMS/OH-MWCNT sol-gel nanocomposite for corrosion protection of magnesium alloy. Surface and Coatings Technology, 335, 228–240.

[18] Bandeira, R. M., et al. (2017). Influence of the thickness and roughness of polyaniline coatings on corrosion protection of AA7075 aluminum alloy. Electrochimica Acta, 240, 215–224.

[19] Bandeira, R. M., et al. (2017). Polyaniline/polyvinyl chloride blended coatings for corrosion protection of carbon steel. Progress in Organic Coatings, 106, 50–59.

[20] Hwang, M. J., et al. (2019). Cathodic electrophoretic deposition of phenylenediamine-modified graphene oxide for anti-corrosion protection. Carbon, 142, 68–77.
[21] Conde, A., Arenas, M. A., & de Damborenea, J. J. (2011). Electrodeposition of Zn–Ni coatings as Cd replacement for high strength steel. Corrosion Science, 53(4), 1489–1497.
[22] Abedini, B., et al. (2019). Structure and corrosion behavior of Zn–Ni–Mn/ZnNi layered alloy coatings. Surface and Coatings Technology, 372, 260–267.
[23] Shalabi, K., & Nazeer, A. A. (2019). Ethoxylates nonionic surfactants as eco-friendly inhibitors for reinforcing steel in 35% NaCl. Journal of Molecular Structure, 1195, 863–876.
[24] Fang, Y., Suganthan, B., & Ramasamy, R. P. (2019). Electrochemical characterization of aromatic corrosion inhibitors from plant extracts. Journal of Electroanalytical Chemistry, 840, 74–83.
[25] Schmidt, D. P., et al. (2006). Corrosion protection assessment of sacrificial coatings as a function of exposure time. Progress in Organic Coatings, 57(4), 352–364.
[26] Yang, J., et al. (2007). Synthesis of conducting polyaniline using anionic Gemini surfactant. European Polymer Journal, 43(8), 3337–3343.
[27] Zhao, Z., et al. (2019). Polydopamine-functionalized graphene oxide nanocomposites reinforced polyurethane coatings. European Polymer Journal, 120, 109249.
[28] Møller, V. B., et al. (2017). Acid-resistant organic coatings for the chemical industry: A review. Journal of Coatings Technology and Research, 14(2), 279–306.
[29] Wilcox, G. D., & Gabe, D. R. (1993). Electrodeposited zinc alloy coatings. Corrosion Science, 35(5), 1251–1258.
[30] Yan, W., et al. (2019). Sol-gel technology for corrosion protection coating systems. Coatings, 9(1), 52.
[31] Li, L., Nie, S., Li, C., Chen, X., Qiao, Y., Ma, R., Chen, Z., Zhang, L., & Cui, J. (2024). Cavitation erosion-corrosion behavior of CoCrFeNiMoCu0.1 high entropy alloy. Ultrasonics Sonochemistry, 110, 107021.
[32] Reclaru, L., Brooks, R. A., Zuberbühler, M., Eschler, P. Y., Constantin, F., & Tomoaia, G. (2014). Evaluation of taper joints with combined fatigue and crevice corrosion testing. Materials Science and Engineering C, 34, 69–77.
[33] Royhman, D., Pourzal, R., Hall, D., Lundberg, H. J., Wimmer, M. A., Jacobs, J., Hallab, N. J., & Mathew, M. T. (2021). Fretting-corrosion in hip taper modular junctions. Journal of the Mechanical Behavior of Biomedical Materials, 118, 104443.
[34] Lekbach, Y., Liu, T., Li, Y., Moradi, M., Dou, W., Xu, D., Smith, J. A., & Lovley, D. R. (2021). Microbial corrosion of metals: The corrosion microbiome. Advances in Microbial Physiology, 78, 317–390.
[35] Kumar, P., Soni, I., Jayaprakash, G., Kumar, S., Rao, S., Flores-Moreno, R., & Swamirayachar, S. (2022). Experimental and theoretical studies of ionic liquid as corrosion inhibitor. Inorganic Chemistry Communications, 146, 110110.
[36] Olajire, A. (2018). Advances in organic coating system technologies for offshore structures. Journal of Molecular Liquids, 269, 572–606.
[37] Hayatdavoudi, H., & Rahsepar, M. (2017). Enhanced cathodic protection performance of graphene-reinforced nanocomposite coating. Journal of Alloys and Compounds, 727, 1148–1156.
[38] Xie, J., Zhang, J., You, Z., Liu, S., Guan, K., Wu, R., Wang, J., & Feng, J. (2021). Developing Mg alloys with improved strength and corrosion resistance. Journal of Magnesium and Alloys, 9, 41–56.
[39] Kadhum, A., Betti, N., Al-Adili, A., Shaker, L., & Al-Amiery, A. (2022). Organic inhibitors for corrosion of mild steel: A critical review (Part two). International Journal of Corrosion Scale Inhibition, 11, 43–63.
[40] Cowley, W. E., Robinson, F. P. A., & Kerrich, J. E. (1968). Anodic protection for corrosion fatigue. British Corrosion Journal, 3, 223–237.
[41] Cuomo, S., Di Cola, V. S., Giampaolo, F., et al. (2022). Scientific machine learning through physics-informed neural networks. Journal of Scientific Computing, 92, 88. https://doi.org/10.1007/s10915-022-01939-z
[42] Stein, M. L. (1987). Large sample properties of simulations using Latin hypercube sampling. Technometrics, 29, 143–151.
[43] Ramezanzadeh, B., Bahlakeh, G., & Quraishi, M. A. (2022). A critical review of machine learning applications in corrosion prediction and inhibitor design. Progress in Organic Coatings, 168, 106912.

[44] Stiasny, J., Misyris, G. S., & Chatzivasileiadis, S. (2021). Physics-informed neural networks for nonlinear system identification. In 2021 IEEE Madrid PowerTech (pp. 1–6). IEEE. https://doi.org/10.1109/PowerTech46648.2021.9495063
[45] Sun, S., Cao, Z., Zhu, H., et al. (2020). A survey of optimization methods from a machine learning perspective. IEEE Transactions on Cybernetics, 50(8), 3668–3681. https://doi.org/10.1109/TCYB.2019.2950779
[46] Waheed, U. B., Haghighat, E., Alkhalifah, T., et al. (2021). PINNeik: Eikonal solution using physics-informed neural networks. Computers & Geosciences, 155, 104833. https://doi.org/10.1016/j.cageo.2021.104833
[47] Galleguillos Madrid, F. M., Soliz, A., Cáceres, L., Bergendahl, M., Leiva-Guajardo, S., Portillo, C., Olivares, D., Toro, N., Jimenez-Arevalo, V., & Páez, M. (2024). Green Corrosion Inhibitors for Metal and Alloys Protection in Contact with Aqueous Saline. Materials, 17(16), 3996. https://doi.org/10.3390/ma17163996
[48] Wang, S., Sankaran, S., & Perdikaris, P. (2022). Respecting causality is all you need for training physics-informed neural networks. arXiv preprint arXiv:2203.07404.
[49] Wong, J. C., Gupta, A., & Ong, Y. S. (2021). Can transfer neuroevolution tractably solve your differential equations? IEEE Computational Intelligence Magazine, 16(2), 14–30. https://doi.org/10.1109/MCI.2021.3061854
[50] Wong, J. C., Ooi, C., Gupta, A., et al. (2022). Learning in sinusoidal spaces with physics-informed neural networks. arXiv preprint arXiv:2109.09338
[51] Legut, D., Kądzielawa, A. P., Pánek, P., Marková, K., Váňová, P., Konečná, K., & Langová, Š. (2021). Inhibition of steel corrosion with imidazolium-based compounds: Experimental and theoretical study. Applied Surface Science, 570**,** 151204. https://doi.org/10.1016/j.apsusc.2021.151204.
[52] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5
[53] Wang, Y., Chen, R., Zhang, T., & Xie, T. (2022). Incorporating physical laws into neural networks for materials science. Nature Computational Science, 2(12), 761–773
[54] Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. IEEE Access, 8, 21980–22012.
[55] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer

10. Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

11. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for profit sectors.