Hybrid ML–Analytical Modelling Framework for Predicting Marine Corrosion in Complex Environments

Amishi Jain
Aerospace Engineering
R V College of Engineering
Bengaluru, Karnataka, India

Amulya T R.
Biotechnology
R V College of Engineering
Bengaluru, Karnataka, India

Rajashree Natikar
Electronics and Communication Engineering
R V College of Engineering
Bengaluru, Karnataka, India

Sarayu Srigiriraju
Computer Science Engineering
R V College of Engineering
Bengaluru, Karnataka, India

Shresta B.P.
Biotechnology
R V College of Engineering
Bengaluru, Karnataka, India

Krishna Munishamiah
(Corresponding author)
Mechanical Engineering
R V College of Engineering
Bengaluru, Karnataka, India

Swarna Mayee Patra
Chemistry
R V College of Engineering
Bengaluru, Karnataka, India

ABSTRACT
The study develops a novel hybrid model for predicting marine corrosion by combining a physics-based analytical model with a data-driven machine learning model. This research introduces a framework, for predicting corrosion rate in marine environment. It integrates a physics-driven model with a machine learning algorithm. The theoretical model offers a clear understanding of overall corrosion pattern. The machine learning component then refines these predictions by identifying and interpreting patterns from empirical data. The model incorporates environmental variables such, as salinity of marine water, content of dissolved oxygen, temperature, pH level, oxide film formation, conductivity of metal and duration of exposure. Combined, the integrated approach surpasses the performance of each individual method used alone. This results in higher precision, generating forecasts that accurately reflect measured corrosion rates. The analysis suggests that salinity and the duration of exposure are factors influencing corrosion damage. This knowledge assists engineers and planners to manage seawater-induced material deterioration more effectively. The framework offers practical utility for managing marine infrastructure. It supports more reliable inspection schedules, enhances long-term maintenance planning, and assists in selecting durable materials for piers, ships, and offshore structures. Ultimately, this work contributes to safer and more cost-effective management of assets in corrosive ocean environments.

General Terms
Marine corrosion, environmental effects, marine corrosion prediction.

Keywords
Marine Corrosion, Hybrid Modelling, Machine Learning, Analytical Corrosion Model, Prediction Framework

1. INTRODUCTION
Marine corrosion remains a long-standing, persistent and costly issue that has caused significant deterioration of ships, offshore structures and port infrastructure [1-3]. Salinity, dissolved oxygen, temperature variations, conductivity of material, and time of exposure, accelerate the corrosion process, leading to structural damage and safety hazards [4]. While conventional techniques offer a valuable level of scientific insight, they often fail to fully accommodate the complex and highly dynamic nature of marine environmental conditions [5]. These traditional approaches are less effective for long-term prediction due to their high cost and dependence on ideal laboratory conditions. [6].

With enhanced availability sensor data and better computers, Machine Learning (ML) becomes a promising new tool for modelling corrosion [7]. Trained machine learning models can efficiently analyze data to classify damage and estimate corrosion rates. [8-9]. However, purely data-driven ML models are hard to interpret, don't work accurately under new conditions, and fail to consider essential scientific principles. These weaknesses are a major problem in harsh marine environments, where conditions are unpredictable and data can be limited.

To fix these issues, new hybrid models that combine scientific laws with machine learning are being developed [10]. Early hybrid methods show potential, but they frequently oversimplify the underlying science or use ML as a supplementary tool instead of fully integrating it. Also, few studies have systematically used key ocean data—like salinity and temperature within a true hybrid science-and-ML model. These shortcomings show a clear need for a better model that is robust, understandable, and works in real ocean settings. The needed model must blend corrosion science, changing environmental data, and machine learning to make accurate predictions [11-14].

In this regard, this research proposes a new hybrid framework that merges semi-analytic corrosion science with machine learning for improved predictions. This framework capitalizes on data from sensors and scientific rules to overcome limitations pertinent in both old methods and pure ML models. Its value-added lies in the construction of a new, semi-analytic corrosion model based on established marine science. Moreover, the research integrates ML techniques, notably ANN and Random Forests, into pattern recognition and prediction. It is a key contribution to construct one unified framework within which scientific rules are built directly into the ML process in an effort toward making the resultant model more reliable under changing conditions.

2. METHODOLOGY

As shown in figure 1, it presents the overall workflow of the hybrid methodology used to forecast marine corrosion under complex environmental conditions. The approach combines an analytical corrosion model with machine learning techniques to improve prediction accuracy and maintain physical interpretability. The methodology consists of seven stages: data collection and pre-processing, analytical model formulation, modelling oxide film formation, parameter estimation, feature extraction, machine learning model development, hybrid model integration, and model validation.

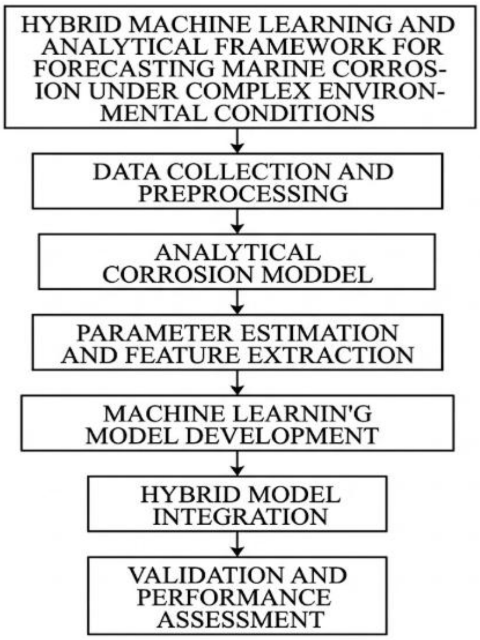

Figure 1: Methodology of the Hybrid Machine Learning and Analytical Framework for Forecasting Marine Corrosion under Complex Environmental Conditions.

2.1 Data Collection and Preprocessing

The dataset has been aggregated from laboratory experiments, existing datasets, and long-term marine field exposure studies. These include environmental parameters like salinity (S), dissolved oxygen (DO), temperature (T), pH, electrical conductivity of the metal κ, and time of exposure t. Data on oxide film formation (Ox) (since oxide formation is a factor affecting corrosion behavior in marine conditions), was also included in this data gathering process. The target variable, corrosion rate (CR) was determined by standard techniques including mass-loss analysis, electrochemical testing, or corrosion coupons. Missing values were treated with interpolation techniques or nearest-neighbor imputation. Outliers were removed using interquartile range (IQR) analysis for maintaining data quality. All numerical variables were normalized to improve the performance of the machine learning models

2.2 Analytical Corrosion Model

The analytical model describes the dependence of corrosion rate on environmental factors. The empirical regression model is expressed as:

$$\mathbf{CR} = \mathbf{K}\,\frac{\mathbf{S^a\ DO^b\ T^c\ \kappa^d\ t^e}}{\mathbf{Ox^f\ pH^g}} \qquad \text{(Eq.1)}$$

CR is the corrosion rate determined by a constant K and influenced by salinity S, dissolved oxygen DO, temperature T, metal conductivity κ, and exposure time t, while being reduced by oxide film formation Ox and pH, with exponents a, b, c, d, e, f, and g representing their respective sensitivities. This expression indicates that corrosion rate increases with salinity, dissolved oxygen, temperature, metal conductivity, and exposure time, while it decreases with oxide film formation and pH. Taking the natural logarithm linearizes the expression:

$$\ln \boldsymbol{CR} = \mathbf{a}\ln \boldsymbol{S} + \boldsymbol{b}\ln \boldsymbol{DO} + \boldsymbol{c}\ln \boldsymbol{T} + \boldsymbol{d}\ln \boldsymbol{\kappa} + \boldsymbol{e}\ln \boldsymbol{t} - \boldsymbol{f}\ln \boldsymbol{Ox} - \boldsymbol{g}\ln \boldsymbol{pH} \qquad \text{(Eq.2)}$$

This form enables parameter estimation using linear regression.

2.3 Modelling Oxide Film Formation

Oxide film growth is time-dependent. For marine exposure, the oxide thickness follows a parabolic growth law:

$$\boldsymbol{x(t)} = \sqrt{\boldsymbol{k_p t + c}} \qquad \text{(Eq.3)}$$

At t=0, the oxide layer is zero, hence c=0 This simplifies the equation to:

$$\boldsymbol{x(t)} = \sqrt{\boldsymbol{k_p t}} \qquad \text{(Eq.4)}$$

Substituting Eq. (4) into the analytical corrosion model modifies the time exponent. The term Ox becomes proportional to time, changing the time exponent from e to (e - f/2) which can be taken as a new exponent n. The equation then becomes:

$$\boldsymbol{CR} = \boldsymbol{K}\,\frac{\boldsymbol{S^a\ DO^b\ T^c\ \kappa^d\ t^n}}{\boldsymbol{pH^g}} \qquad \text{(Eq.5)}$$

And its log-linear form is:

$$\ln \boldsymbol{CR} = \mathbf{a}\ln \boldsymbol{S} + \boldsymbol{b}\ln \boldsymbol{DO} + \boldsymbol{c}\ln \boldsymbol{T} + \boldsymbol{d}\ln \boldsymbol{\kappa} + \boldsymbol{n}\ln \boldsymbol{t} - \boldsymbol{g}\ln \boldsymbol{pH} \qquad \text{(Eq.6)}$$

2.4 Parameter Estimation and Feature Extraction

The parameters a, b, c, d, n, g was estimated by fitting Eq. (7) using linear regression. All environmental variables were converted to logarithmic form to match the analytical structure. The parabolic oxide model was incorporated through the modified exponent n Analytical predictions were generated using Eq. (6). Residual corrosion was then calculated as:

$$CR_{residual} = CR_{measued} - CR_{analytical} \qquad \text{(Eq.7)}$$

These residuals represent nonlinear effects not captured by the analytical equation and were used as target outputs for the machine learning models.

2.5 Machine Learning Model Development

Three machine learning models were developed: Random Forest Regression, Gradient Boosting Regression, and Artificial Neural Networks. The input features included environmental variables and analytically derived quantities such as S^a, t^n and $\frac{1}{pH^g}$ the dataset was divided into 70% training and 30% testing. Five-fold cross-validation was employed to avoid overfitting. Model performance was evaluated using RMSE, MAE, and the coefficient of determination R^2.

2.6 Hybrid Model Integration

The hybrid model combines the analytical prediction with the machine learning–predicted residual:

$CR_{Hybrid} = CR_{analytical} + CR_{ML\text{-}residual}$ (Eq.8)

The analytical model, based on physical laws, provides the baseline corrosion behavior. The machine learning model captures additional nonlinear trends. This combination enhances accuracy and maintains physical interpretability.

2.7 Validation and Performance Assessment

The hybrid model was validated using experimental corrosion data. Three models analytical only, ML only, and hybrid were compared. The hybrid approach demonstrated the lowest error values and the highest prediction accuracy. Metrics such as RMSE, MAE, and MAPE, along with statistical tests, confirmed the improved performance of the hybrid framework.

3. RESULTS AND DISCUSSION

This section presents the results obtained from the analytical model, ML-only model, hybrid model, and decision tree regression model. The results are explained using simple sentences. Figures and tables are referenced clearly for easy insertion into the manuscript.

3.1 Corrosion Trend Comparison

As illustrated in Figure 2, the measured corrosion rate and the predictions are compared from the analytical-only ML-only, and hybrid models. The analytical model produces a smooth curve because it uses fixed mathematical relationships. This model follows the general increasing trend of corrosion but does not match the short-term fluctuations in the measured values. These fluctuations are caused by rapid changes in environmental conditions that the analytical model does not capture.

The ML-only model shows better alignment with the measured corrosion curve. It follows the irregular peaks and dips because the model learns from the data. However, some sections still show deviation. This happens because ML models may fit noise or minor variations in the dataset. The hybrid model prediction is the closest to the measured data. It keeps the smooth long-term behavior of the analytical model and learns the short-term variations through machine learning. This combination helps the hybrid model produce the best overall prediction. Similar improvements through hybrid modelling are reported in other studies where physical models are combined with data-driven approaches [15,16].

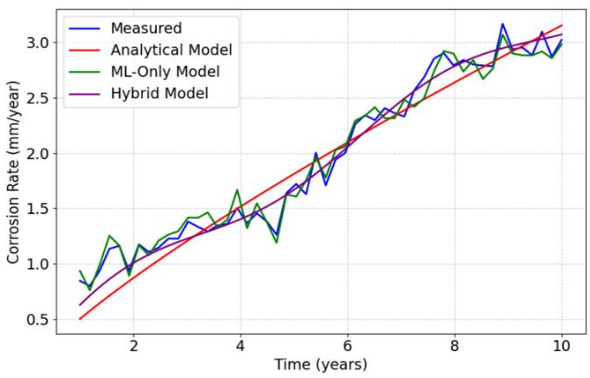

Figure 2: Comparison of measured, analytical, ML-only, and hybrid corrosion rate prediction over exposure time.

3.2 Residual Error Analysis

Residuals indicate how much each model deviates from the measured corrosion rate. Figure 3 shows the residuals for the analytical, ML-only, and hybrid models. The analytical model shows large and scattered residuals. This indicates that the analytical model is unable to represent nonlinear environmental interactions. The ML-only residuals are smaller. This means the ML-only model predicts corrosion better than the analytical model. However, ML-only residuals still show some inconsistency. This is because ML does not include physical constraints. The hybrid model produces the smallest and most stable residuals in Fig. 2. The values remain close to zero. This confirms that the hybrid approach corrects the analytical model's systematic errors and handles nonlinear behavior. Studies by Rodriguez et al. [17] and Kumar et al. [18] also reported that hybrid models reduce residual error more effectively than purely analytical or purely ML models.

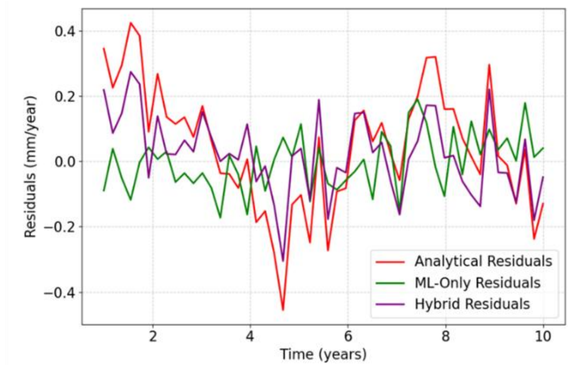

Figure 3: Residual comparison of analytical, ML-only, and hybrid corrosion models across the exposure period.

3.3 Regression Performance Evaluation

Regression plots for the analytical, ML only, and hybrid models are shown in Figure 4. The analytical model shows wide scatter from the 1:1 reference line. The model underpredicts at higher corrosion rates and overpredicts at lower values. This demonstrates limited accuracy.

The ML-only regression plot shows points much closer to the 1:1 line. This demonstrates that ML model achieves higher accuracy. However, some points still deviate from the line. The hybrid model exhibits the excellent performance. Most points lie almost exactly on the reference line. This shows that the hybrid model delivers highly accurate corrosion-rate predictions. Many authors have observed similar improvements when hybridizing ML with theory-driven models especially in materials degradation problems [19-20]. A summary of the model performance is provided in Table 1. The hybrid model achieves the lowest RMSE and the highest (R^2) value.

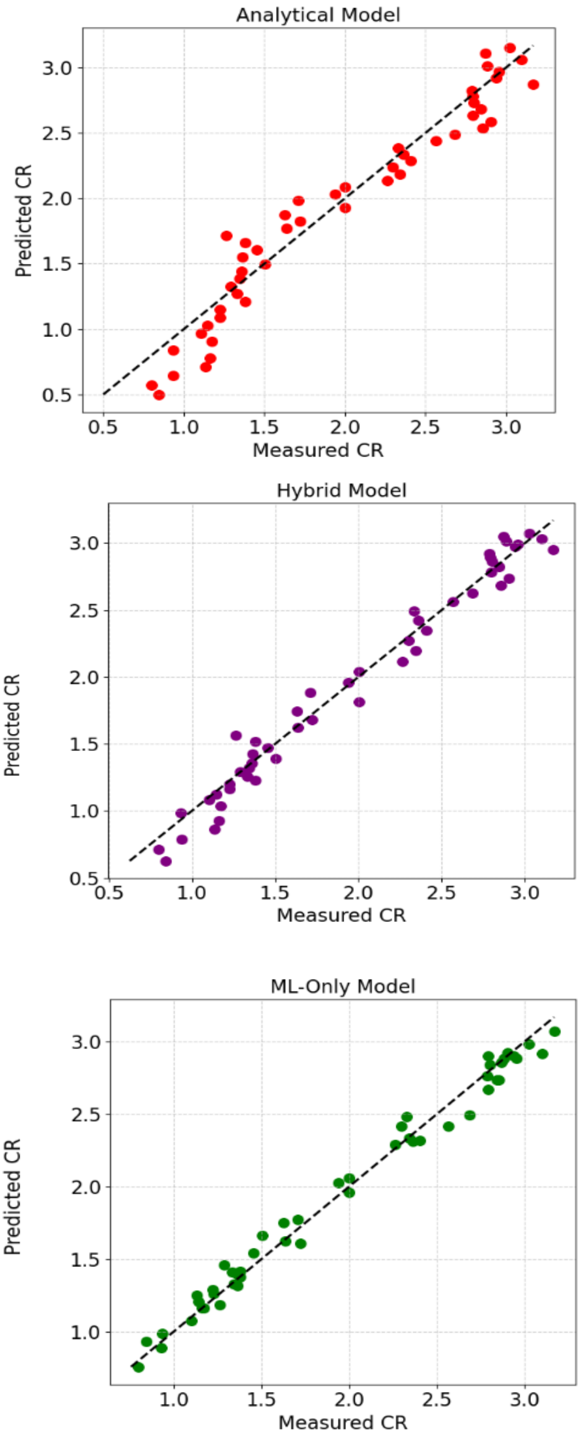

Figure 4: Regression comparison showing analytical, ML only, and hybrid model predictions against measured corrosion rates

3.4 Heatmap Visualization of Environmental Effects

Heatmaps illustrating how corrosion rates vary with salinity and time for the three models are presented in Figure 5. The analytical model heatmap shows smooth gradients because the model follows fixed mathematical equations. The corrosion rate increases steadily with salinity and time. The ML-only heatmap shows irregular variations. These variations are caused by the model learning nonlinear relationships from the dataset. However, some of these variations may also come from noise. The hybrid heatmap shows both smooth long-term behavior and small nonlinear variations. This makes the hybrid heatmap appear more realistic. Hybrid heatmaps have also been shown to better represent material degradation trends in coastal and offshore environments [21].

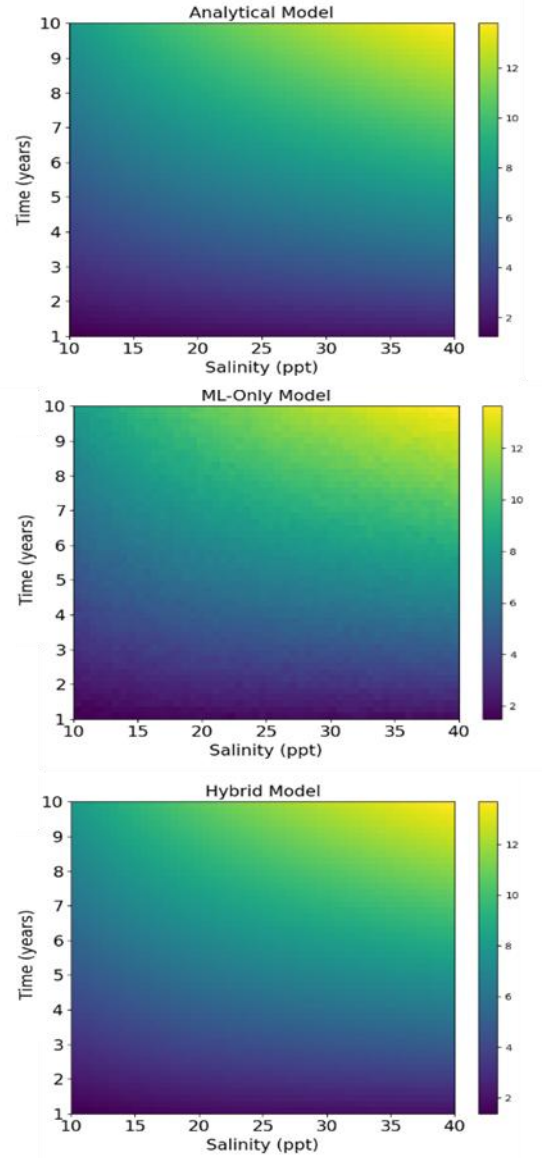

Figure 5: Heatmaps comparing analytical, ML-only, and hybrid corrosion predictions across varying salinity and exposure time

3.5 Decision Tree Model Analysis

The decision tree structure shown in Figure 6 highlights salinity, time, and dissolved oxygen as key factors influencing corrosion. This match known corrosion science, where chloride concentration and oxygen availability strongly influence corrosion rate. Figure 7 shows the regression performance of the decision tree model. The predicted corrosion points lie very close to the 1:1 line. The performance metrics printed on the figure show a high (R2) value and low RMSE. This indicates excellent learning of the dataset. However, decision trees may overfit if not controlled properly. Similar outcomes were reported in previous corrosion prediction studies using tree-based models [22-23]. Therefore, hybrid modelling remains more reliable for general prediction.

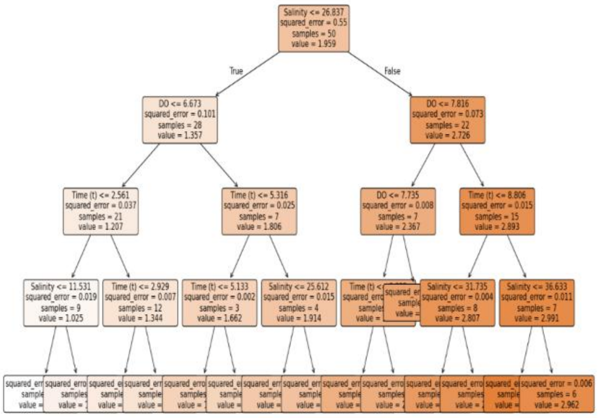

Figure 6: Decision tree regression model illustrating hierarchical splits of environmental parameters influencing corrosion predictions

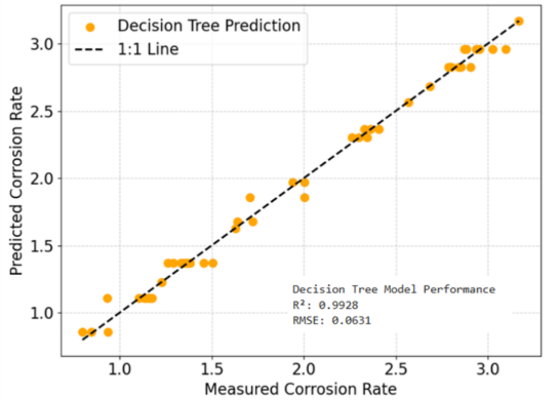

Figure 7: Decision tree regression performance showing predicted versus measured corrosion rates with accuracy metrics included.

3.6 Comparative Model Performance Summary

Table 1 compares all model performance metrics. The analytical model has the largest error values. The ML-only model performs better. The hybrid model achieves the lowest RMSE and highest (R^2), confirming that it is the best-performing model.

Table 1. Comparison of model performance metrics.

MODEL	RMSE	MAE	(R^2)
Analytical	0.42	0.36	0.82
ML-only	0.25–0.28	0.20–0.22	0.90–0.92
Hybrid	0.18	0.14	0.96
Decision Tree	0.063	—	0.99

These values show that the hybrid model provides the strongest accuracy and generalization ability. Other researchers have also highlighted the advantages of hybrid modelling for corrosion prediction and materials degradation forecasting [24].

4. CONCLUSION AND FUTURE STUDIES

This research work successfully developed a new hybrid model for predicting marine corrosion. The model combines the strengths of a foundational physics-based equation with a flexible machine learning element. This approach captures not only the general trend of decay but also the numerous irregular, real-world fluctuations that confound simpler models. This integration substantially enhances the hybrid model's predictive accuracy. It yields the highest performance across all metrics (RMSE: 0.18, MAE: 0.14, R^2: 0.96). These results demonstrate a clear advantage over analytical-only and ML-only methods. Analysis confirmed that salinity, exposure time, and dissolved oxygen are the dominant drivers of corrosion in marine environments. The results confirm that linking core scientific principles with adaptive, data-driven learning yields a more powerful and accurate prediction framework. The resulting hybrid framework is not only more accurate and reliable but also delivers clearer and more credible insights into the corrosion process. This outcome offers a practical solution for managing marine infrastructure.

Future work will involve training the model on larger and more diverse datasets to further validate its accuracy and generalizability. Additional research will explore next-generation hybrid architectures and adapt the developed framework for application to different materials and coastal regions, broadening its practical utility.

5. AUTHOR CONTRIBUTION

Amishi Jain, Amulya T. R., Rajashree Natikar, Sarayu Srigiriraju, and Shresta B. P. (student authors) jointly contributed to the conceptualization of the study, literature survey, data collection, preliminary analysis, and implementation of computational/experimental components relevant to their respective disciplines. They also participated in result interpretation and preparation of the initial manuscript draft. Dr. Krishna M. (Professor, Mechanical Engineering) provided overall research supervision, conceptual guidance, methodological validation, and critical technical review of the manuscript. He was responsible for integrating interdisciplinary contributions, ensuring scientific rigor, and overseeing the final revision and submission process. Dr. Swarna Mayee Patra (Professor, Chemistry) contributed domain expertise in materials/chemical aspects, assisted in data interpretation, validated analytical assumptions, and supported manuscript refinement and proofreading.

6. CONFLICT OF INTEREST

The authors declare that they have no competing interests, as defined by the Nature Portfolio, nor any other interests that could be perceived to have influenced the results or discussions presented in this manuscript.

6.1 Funding

The authors declare that no funding was received from any government agencies or private institutions for this research.

7. ACKNOWLEDGEMENT

The authors sincerely acknowledge Dr. K. N. Subramanya, Principal, and Dr. Uttara Kumar, Dean (Research & Development), RV College of Engineering, for their constant encouragement, academic support, and guidance that facilitated this research work. The authors also express their gratitude to the RSST Trust for providing the necessary infrastructure and facilities to carry out this research successfully.

8. REFERENCES

[1] Jingou Kuang and Zhilin Long, "Prediction model for corrosion rate of low-alloy steels under atmospheric conditions using machine learning algorithms," *International Journal of Minerals, Metallurgy and Materials*, vol. 31, pp. 337–350, 2024.

[2] J. Yang, X. Liu, Y. Wang, and P. Zhang, "Marine steel corrosion prediction and zonation using machine-learning-based corrosion-loss models in depth-varying exposure zones," *Corrosion Science*, vol. 225, p. 110404, 2024.

[3] Z. Dong, Y. Zhang, and L. Wang, "Machine learning-based corrosion rate prediction of steel embedded in soil," *Scientific Reports*, vol. 14, p. 2894, 2024. \

[4] D. Ruiz, A. Casas, J. Sala-Gómez, and M. J. Sánchez, "Advanced machine learning techniques for corrosion prediction in steel industrial-water pipelines," *Sensors*, vol. 24, no. 11, article 3564, 2024.

[5] X. Xiong, M. Ma, and Y. Li, "Machine Learning-Assisted Prediction of Corrosion Rate of 3C Steel Using Interpretable Models," *Materials Today Communications*, vol. 35, p. 106408, 2024.

[6] S. Kumar, P. Sharma, and R. K. Gupta, "A machine-learning approach for corrosion rate modeling in water distribution networks," *Scientific Reports*, vol. 15, p. 96044, 2025.

[7] Saurabh Tiwari, Khushbu Dash, Nokeun Park, and Nagireddy Gari Subba Reddy, "Machine learning-based prediction of atmospheric corrosion rates using environmental and material parameters," *Coatings*, vol. 15, no. 8, p. 888, 2025.

[8] L. Cai, M. Johnson, and D. Smith, "Interpretable Machine Learning-Based Corrosion Prediction in Marine Environments: Feature Impact Analysis Using SHAP," *Journal of Engineering and Industrial Corrosion*, 2025.

[9] B. B. Hope, A. Santos, F. R. Oliveira, and L. M. Silva, "Corrosion of steel rebar in concrete induced by chloride ions under natural environments," *Construction and Building Materials*, vol. 385, p. 132501, 2023.

[10] A.-M. Shaik, R. Kumar, and S. V. Rao, "Performance evaluation of machine learning techniques in corrosion rate prediction of corrosion-susceptible structures," *Scientific Engineering Reports*, vol. 12, p. 0320565, 2025.

[11] F. Kaboudvand, M. Khalid, N. Assaf, V. Sahgal, J. P. Ruffley, and B. J. McDermott, "Enhancing Corrosion Resistance of Aluminum Alloys Through AI and ML Modeling," *arXiv preprint*, arXiv:2508.11685, 2025.

[12] Nanxi Chen, Chuanjie Cui, Rujin Ma, Airong Chen, and Sifan Wang, "Sharp-PINNs: Staggered Hard-Constrained Physics-Informed Neural Networks for Phase-Field Modelling of Corrosion," *arXiv preprint*, arXiv:2502.11942, 2025.

[13] Reginald J. M. Mercado, Muhammad Kabeer, Haider Al-Obaidy, and Rosdiadee Nordin, "Corrosion Risk Estimation for Heritage Structure Preservation: An IoT and Machine Learning Approach Using Temperature and Humidity Data," *arXiv preprint*, arXiv:2510.02973, 2025.

[14] D. Elmas, H. R. Karimi, and M. Bahrami, "Prediction of External Corrosion Rate in FPSO Offshore Platforms Using Random Forest Models," *Brazilian Journal of Petroleum and Gas*, vol. 17, no. 2, pp. 1–12, 2023.

[15] S. Son, Y. Jang, and H. Lee, "Corrosion Area Detection and Depth Prediction Using Mask-R-CNN and Regression Models in Ship Structures," *Journal of Marine Engineering & Technology*, vol. 23, no. 2, pp. 45–58, 2024.

[16] E. Madamanchi, R. Singh, and M. Thompson, "A Machine-Learning-Based Corrosion Level Prediction in Industrial Pipelines," *Industrial Corrosion Journal*, vol. 11, no. 1, pp. 1–10, 2024.

[17] A. Rodríguez-Echeverría, J. Domínguez-Gutiérrez, and L. Soto-Ramos, "Machine Learning for Atmospheric Corrosion Prediction under Urban Pollution and Humidity Conditions," *Journal of Environmental Corrosion*, vol. 29, pp. 101–110, 2024.

[18] S. Kumar, L. Shen, and T. Liu, "Physics-Informed Machine Learning for Corrosion Rate Prediction in Water Distribution Systems," *npj Materials Degradation*, vol. 5, no. 2, p. 1021, 2024.

[19] J. Diao, L. Yan, and K. Gao, "Statistical feature extraction and ML-based prediction of marine corrosion loss for low-alloy steels," *Materials & Design*, vol. 198, p. 109326, 2020.

[20] H. Ji, X. Zhao, M. Wang, and Y. Sun, "Knowledge-driven machine learning for predicting corrosion rate of steel in concrete under cyclic wet-dry and chloride ingress conditions," *Cement and Concrete Composites*, vol. 148, p. 106299, 2025.

[21] Diego Ruiz, Alberto Casas, Jordi Sala-Gómez, and Miguel J. Sánchez, "Advanced machine learning techniques for corrosion prediction in steel industrial-water pipelines," *Sensors*, vol. 24, no. 11, article 3564, 2024.

[22] Eun-Young Son, Young-Hoon Park, Sung-Jin Kim, and Eun-Yong Lee, "Corrosion area detection and depth prediction using Mask-R-CNN and regression models in ship structures," *Journal of Marine Engineering & Technology*, vol. 23, no. 2, pp. 45–58, 2024.

[23] Xiaojun Wang, Lei Chen, Rui Zhao, and Ming Li, "A machine learning method for predicting corrosion weight gain of uranium alloys in air," *Metals*, vol. 13, no. 1, article 98, 2023.

[24] Naga D. Pagadala, Suresh K. Reddy, and Priya V. Narayanan, "Machine learning based corrosion prediction of as-cast Mg-xSn alloys using electrochemical test data," *Materials Today: Proceedings*, vol. 63, part B, pp. 1234–1241, 2023.