Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Corrosion Potential Prediction of Marine Engineering Steel Based on Machine Learning

Bin Wu
Jiangsu Ocean University
Yicong Luo
Jiangsu Ocean University
Shiwei Yu
Jiangsu Ocean University
Endian Fan
fanendian@163.com
Jiangsu Ocean University

Article

Keywords: Marine engineering steel, Corrosion potential prediction, Machine learning

Posted Date: February 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8675886/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at npj Materials Degradation on May 16th, 2026. See the published version at https://doi.org/10.1038/s41529-026-00809-5.

Corrosion Potential Prediction of Marine Engineering Steel Based on Machine Learning

Bin Wu[a], Yicong Luo[a], Shiwei Yu [a], Endian Fan[a, b, c*]

a Jiangsu Institute of Marine Resources Development, Jiangsu Ocean University, Lianyungang, 222005, China

b Co-Innovation Center of Jiangsu Marine Bio-industry Technology, Jiangsu Ocean University, Lianyungang, 222005, China

c School of Chemistry and Chemical Engineering, Nanjing University of Science and Technology, Nanjing, 210094 China

Abstract Corrosion is a major cause of failure in marine engineering steels, resulting in large economic losses worldwide. This study combines marine corrosion knowledge with machine learning techniques to predict corrosion potential. Using experimental data collected from many published studies, five machine learning models were built in Python: K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), Gradient Boosting Regressor (GBR), Stacked Generalization (Stacking), and a Weighted Average Ensemble. Model prediction accuracy was improved through feature engineering and data augmentation. The XGBoost model performed best and achieved a coefficient of determination (R^2) of 0.80 on the training set and 0.62 on the test set. Its mean absolute error (MAE) was 0.07 V and root mean square error (RMSE) was 0.09 V. The generalization gap was 0.179. Feature importance analysis revealed that Mn, Cr, and the Cr×Mo interaction are key factors influencing corrosion potential. This approach provides a accurate and interpretable technical solution to predict corrosion potential for marine engineering steels. This study offers valuable insights for optimizing steel composition and enhancing corrosion-resistant design.

Keywords: Marine engineering steel; Corrosion potential prediction; Machine learning.

1 Introduction

With the rapid development of the ocean economy, metallic materials-especially

* Corresponding author.
E-mail addresses: fanendian@163.com (Endian Fan).

various types of steel-are widely used in marine engineering equipment and related fields. Carbon steel is highly favored due to its high strength, low cost, and abundant raw materials [1]. Low-alloy steel, produced by adding small amounts of alloying elements such as Mn, Si, Cr, and Ni (with total alloy content not exceeding 5%) to carbon steel, offers improved strength, toughness, and corrosion resistance. Despite a moderately higher cost, it is also extensively employed in marine applications [2]. However, the harsh marine environment leads to severe corrosion, significantly undermining the reliability and service life of marine engineering equipment and posing a global challenge. According to a 2016 NACE International report (shown in Figure 1), the global annual economic cost of corrosion was approximately US $2.5 trillion, accounting for about 3.4% of the world's Gross Domestic Product (GDP) in 2013 [3,4]. The report further breaks down these costs by region: corrosion accounts for 2.7% of GDP in the United States, 3.8% in the European Union (including Switzerland and Norway), 4.2% in India, 5% in Arab nations, 4.2% in China, 4% in Russia, and 1% in Japan [5]. Timely and accurate assessment of the corrosion resistance of metallic materials, along with a clear understanding of the influencing factors, would greatly support the optimization of material design and protection strategies. Nevertheless, traditional research methods-such as outdoor exposure tests or laboratory simulations of corrosive environments-are often limited by long testing cycles, high costs, and the complexity of analyzing corrosion behavior.

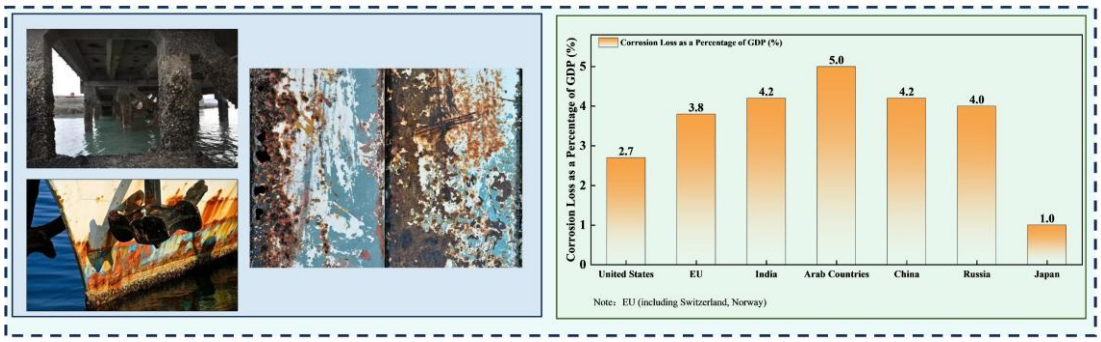


Figure 1. Corrosion cases, global economic losses from corrosion.

The advancement of big data technologies has enabled machine learning (ML) to excel in building predictive models and revealing complex, underlying patterns within large-scale datasets. The integration of ML with materials science has unlocked

powerful data-mining potential, driving progress in key areas such as compositional optimization, property prediction, fatigue life assessment, and the design of high-throughput experiments [6-11]. Supported by a rich ecosystem of modeling algorithms and accessible implementation frameworks, ML has become an indispensable tool in modern materials research and development [12]. Its impact is evident in various advanced applications, including the accelerated discovery of novel high-temperature alloys, the design of shape memory alloys with minimized thermal hysteresis, and the optimization of processes for manufacturing metallic additives [13,14].

Within the specialized domain of corrosion science and engineering, ML methodologies are gaining significant traction. Established techniques like support vector regression and artificial neural networks have been widely reported for corrosion-related predictions [15]. These data-driven approaches are valued for their exceptional predictive accuracy and capacity to model highly nonlinear relationships. For instance, Jihua Bao et al. developed and compared six ML models-including linear regression, random forest, decision tree, artificial neural network, KNN, and XGBoost-to predict the stress corrosion cracking behavior of carbon steel in simulated marine environments, ultimately optimizing the XGBoost model for superior performance [1]. Similarly, Luchun Yan et al. employed a random forest model to predict the corrosion rate of low-alloy steel in marine atmospheres and leveraged SHAP analysis to interpret the model, identifying chemical composition, chloride deposition rate, and rainfall as the most influential factors [2]. The field continues to advance with more sophisticated architectures; hybrid deep learning models such as Convolutional Gated Recurrent Units (CGRU) have been applied to forecast atmospheric corrosion [16]. Furthermore, Shihang Lu et al. effectively combined extensive corrosion datasets with the XGBoost algorithm to elucidate the specific impacts of environmental factors like temperature and humidity on the atmospheric corrosion of EH36 steel [17]. Collectively, these studies underscore the substantial and growing potential of machine learning to enhance corrosion analysis, manage complex nonlinear fittings, and extract actionable insights from vast corrosion data, paving the way for more predictive and preventive maintenance strategies in marine and industrial applications.

While machine learning applications in corrosion science show considerable promise, current methodologies reveal clear opportunities for refinement. A primary concern lies in feature engineering strategies, which often remain predominantly empirical and data-centric, rather than being guided by established corrosion electrochemistry and metallurgical principles. This limits the physical interpretability and generalizability of the resulting models. Furthermore, the research emphasis has been disproportionately placed on correlating environmental parameters-such as chloride concentration, temperature, and humidity-with corrosion rates. This comes at the expense of a more systematic investigation into the intrinsic role of alloying elements and their interactive effects, which are fundamental determinants of a material's corrosion behavior and passive film stability. Bridging this gap between domain knowledge and data-driven modeling is essential for developing robust, interpretable, and materially informative predictive tools.

The corrosion potential serves as a crucial electrochemical indicator for assessing the corrosion behavior of metallic materials, with its value determined by both alloy composition and environmental conditions. Extensive research confirms that alloying elements-including Cr, Mn, Mo, and Cu-significantly influence corrosion performance. For example, Haoping Peng et al. demonstrated through controlled experiments that increasing Cr content facilitates the formation of a protective corrosion product layer, thereby improving the corrosion resistance of low-alloy steels [18]. Similarly, Joonoh Moon et al. reported that appropriate additions of Mo or Cr promote the development of a thicker, denser, and more protective passive film. Their findings indicate that higher Mo/Cr levels combined with lower Mn content enhance pitting resistance, though excessive additions of Mo and/or Cr may prove detrimental [19]. Furthermore, Tianyi Zhang et al. observed that in a simulated harsh marine environment (3.5% NaCl, 40±1°C, 80±5% humidity), rust layers on Ni-Mo steels with varying Cu content exhibited favorable electrochemical performance. This improvement was attributed to the role of Cu in enhancing the protective quality of the rust layer through the formation of $CuFeO_2$ [20].

Building on these insights, this study integrates domain knowledge from marine

corrosion science with machine learning techniques to predict the corrosion potential of marine engineering steels. Corrosion data were collected with an emphasis on alloying element composition, supplemented by key environmental parameters. Using Python, we developed three predictive models: XGBoost, Gradient Boosting Regressor (GBR), and K-Nearest Neighbors (KNN). To better capture the nonlinear relationships between features and corrosion potential, we applied feature engineering and data augmentation techniques. Subsequently, two ensemble strategies-Stacking and Weighted Average-were implemented to further improve predictive accuracy. To interpret the underlying mechanisms of the models, feature importance analysis and SHAP (SHapley Additive exPlanations) methods were employed, thereby elucidating the "black-box" nature of the machine learning approaches and offering an intuitive understanding of how individual features influence corrosion potential. Finally, the practical feasibility of the developed models was validated through independent experimental testing.

2 Experimental Details

Figure 2 illustrates the overall experimental workflow of this study, which consists of three main stages: data preprocessing, feature engineering, and model training.The feature engineering phase involves a three-stage progressive feature selection process, feature interaction, and data augmentation. For model training, three individual models-Gradient Boosting Regressor (GBR), XGBoost, and K-Nearest Neighbors (KNN)-were employed. Furthermore, two ensemble strategies, Stacking and Weighted Average, were implemented to enhance prediction performance.

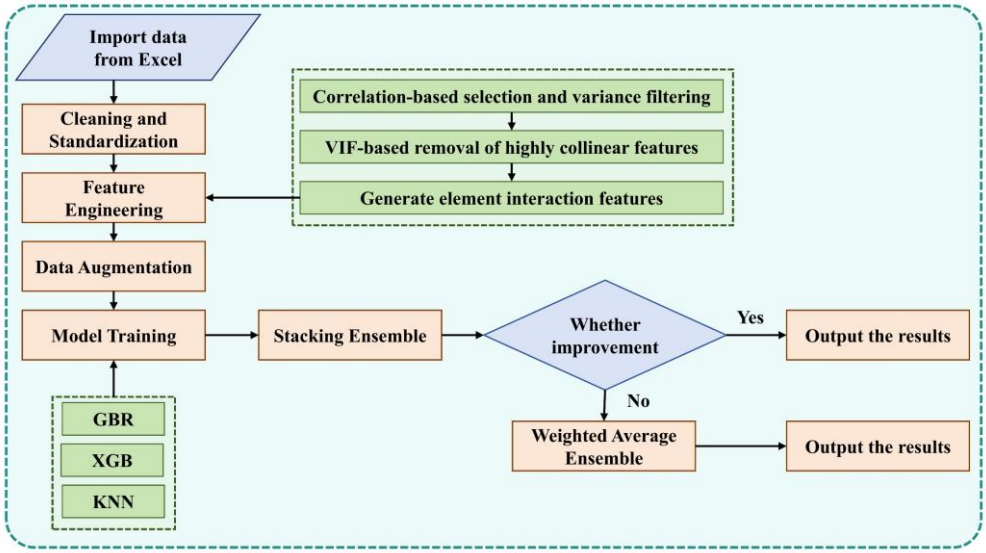


Figure 2. Experimental flow chart. All models were trained based on the scikit-learn library in Python.

2.1 Data Sources and Preprocessing

The corrosion data used in this study were collected from authoritative databases. The dataset comprises corrosion data for various carbon steels, low-alloy steels, and high-strength steels obtained from laboratory-simulated seawater environments documented in relevant literature, journals, and reports. Given the primary objective of investigating the influence of alloying elements on the corrosion potential of marine engineering steels, each alloying element was treated as an independent feature. The original dataset consists of 26 feature variables: 24 alloying element compositions (including C, Al, Mn, Mo, and S) and 2 environmental parameters (NaCl concentration and temperature). The target variable is the corrosion potential. In total, the dataset contains 264 data samples.

To eliminate scale differences and data distribution bias, preprocessing was performed on the raw dataset. Missing values for certain trace elements were replaced with 0, while missing temperature records were assumed to be room temperature (25±1°C). Data with NaCl concentrations outside the 3–5% range were excluded, and alloying elements present in only a few specific steel grades were removed. After cleaning, the final dataset retained 16 alloying elements, 2 environmental parameters, and 1 target variable, comprising 170 valid samples. To prevent the influence of varying

feature scales, all input features and the target variable were normalized to the range [-1, 1] using MinMaxScaler. The normalization formula is as follows:

$$x_{scaled} = 2 \times \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1 \quad (1)$$

Where x is the original feature value, and x_{max} and x_{min} are the maximum and minimum values of that feature across the entire dataset, respectively.

2.2 Feature Engineering

2.2.1 Feature Selection

Given the characteristics of the corrosion prediction task, a three-stage progressive feature screening process was designed. The first stage involved correlation-based screening between the input features and the target variable (corrosion potential). Pearson correlation coefficients were calculated pairwise among all input features as well as between each feature and the target variable. The Pearson correlation coefficient, widely used to measure the linear relationship between two variables, is defined as the quotient of their covariance and the product of their standard deviations. Its value ranges from -1 to 1, where 1 indicates a perfect positive linear correlation, -1 indicates a perfect negative linear correlation, and 0 indicates no linear correlation. The formula is given as follows:

$$r = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

Where x_i represents the feature value of the *i-th* sample, y_i represents the corrosion potential of the *i-th* sample, and $\bar{x}$ and $\bar{y}$ are the mean values of the feature and corrosion potential, respectively. The second stage applied low-variance filtering to remove features with variance below a set threshold, thereby eliminating variables that contributed little information. The third stage involved collinearity elimination, where the Variance Inflation Factor (VIF) was calculated. Features with a VIF greater than 10 were excluded to mitigate the risk of model overfitting caused by multicollinearity. Together, this three-stage screening process effectively helps prevent model overfitting.

2.2.2 Feature Interaction and Data Augmentation

Research indicates that interactions among alloying elements significantly influence the corrosion performance of steel. For example, Chen Xinhua et al. found that the synergistic effect of Si and Al can improve the corrosion resistance of metals in marine atmospheric environments. Similarly, Nishimura et al. studied the marine atmospheric corrosion resistance of Si- and Al-bearing steels. Through indoor accelerated corrosion tests, they observed a notable enhancement in corrosion resistance with increasing Si and Al content [21-24]. Xinyong Yan et al., while investigating the corrosion behavior of high-Mn steel in a simulated marine splash zone environment, reported that Mo oxides contribute to forming a stable passive film, which reduces the influence of chloride ions. Concurrently, Cr oxides in the outer rust layer increase its compactness, thereby slowing corrosion progression [25]. Furthermore, Dazheng Zhang et al. found that significant enrichment of Mn in the rust layer under wet-dry cyclic conditions negatively affects the corrosion resistance of steel, and manganese compounds can also diminish the protective capacity of corrosion products. There is also a view that a specific ratio of C to Mn may increase pearlite content, leading to higher corrosion current density and reduced corrosion resistance, though direct experimental evidence for this mechanism appears limited [26]. Therefore, based on the above understanding, this study constructed three interaction features: Cr×Mo, Si×Al, and C×Mn. Additionally, a boundary condition of Cr > 1.2% was set as an indicator for stainless steel, reflecting the abrupt point for passive film formation. This allows the model to distinguish the influence of Cr content on corrosion potential before reaching this threshold. Finally, VIF screening was reapplied to the newly generated interaction features (removing those with VIF > 15), missing original features were filled, and the final feature set was determined.

To enhance model robustness and mitigate overfitting, a data augmentation module was introduced prior to model training. To avoid sample distortion caused by uniform noise, differentiated random perturbations were applied to different features. Considering the sensitivity of corrosion rate to temperature [27], a small fluctuation of ±2% was applied to the temperature feature. Medium-level fluctuations of ±3% were

applied to alloying element features to account for potential compositional variations arising from manufacturing processes. For the corrosive medium feature, a fluctuation of ±5% was introduced.

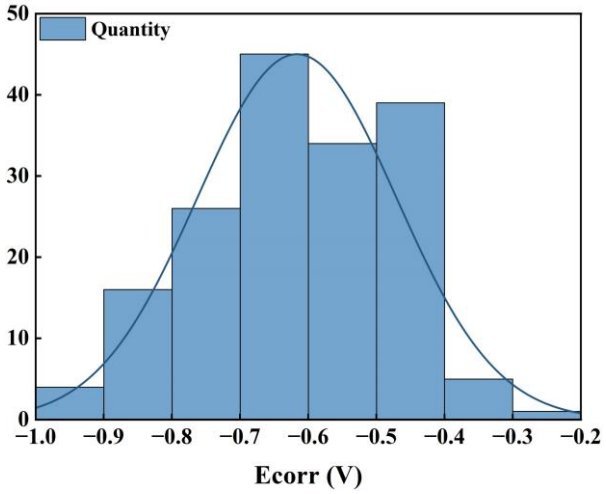


Figure 3. Corrosion potential distribution map. It is used to visualize the distribution of corrosion potential and avoid the occurrence of outliers and extreme values.

The final dataset was divided using stratified random sampling, with 80% allocated as the training set and 20% as the test set. To ensure reproducibility, the random seed was fixed at random_state = 102. The distribution of the corrosion potential in the final dataset is shown in Figure 3, where no significant outliers were observed.

2.3 Model Construction

In this study, three individual machine learning algorithms-Gradient Boosting Regressor (GBR), XGBoost, and K-Nearest Neighbors (KNN)-and two ensemble algorithms-Stacking and Weighted Average-were employed. For all models, a two-step hyperparameter tuning process was conducted. Initial default parameters were first determined via grid search. These were then iteratively refined within defined ranges to identify regions yielding higher predictive accuracy [28-31].

Gradient Boosting Regressor (GBR) is an iterative boosting algorithm based on decision trees, which is highly effective at capturing nonlinear relationships and feature interactions. It performs well in fitting continuous features, such as alloying element concentrations and temperature. Using regression trees as weak learners, GBR iteratively trains new learners to fit the residuals (differences between actual and

predicted values) from the previous iteration through gradient descent. The final prediction is obtained by summing the weighted outputs of all weak learners. Through grid search and empirical tuning, the following hyperparameters were selected: n_estimators = 120, learning_rate = 0.05, max_depth = 3, subsample = 0.8, and random_state = 42.

XGBoost is an optimized variant of Gradient Boosting Regressor (GBR). While it retains the core iterative residual training approach, XGBoost incorporates additional mechanisms such as regularization, column subsampling, and native handling of missing values. It also employs parallel computation to optimize training efficiency. Its objective function combines a loss term with a regularization term, balancing predictive accuracy and model complexity. Through grid search and empirical tuning, the following hyperparameters were selected: n_estimators = 100, learning_rate = 0.07, max_depth = 3, reg_lambda = 8, objective = "reg: squarederror", and random_state = 43.

GBR and XGBoost are both ensemble learning models from the gradient boosting family. Their shared key hyperparameters include: n_estimators (the number of weak learners, i.e., decision trees), max_depth (the maximum depth of each tree), and learning_rate (the shrinkage coefficient applied to each learner's contribution). The parameter subsample specifies the fraction of samples used to train each individual tree. A parameter specific to XGBoost is reg_lambda, which controls the weight of the L2 regularization term.Tuning these parameters aims to constrain model complexity, mitigate overfitting, and balance performance on the training set with generalization capability on unseen test data. The objective parameter defines the loss function; here, "reg: squarederror" denotes the mean squared error, which is suitable for regression tasks such as corrosion potential prediction. Compared to GBR, XGBoost generally exhibits superior generalization ability, faster training speed, more refined feature weighting, and a lower risk of overfitting on small datasets.

Unlike GBR and XGBoost, K-Nearest Neighbors (KNN) is a non-parametric, distance-based model with no explicit training phase. For prediction, it calculates the distance between a query sample and all samples in the training set, identifies the k

nearest neighbors, and outputs the mean (or weighted mean) of their target values. This approach imposes no assumptions on the underlying data distribution, making it well-suited for capturing the complex, nonlinear variations typical of corrosion potential. The following hyperparameters were selected through grid search and empirical tuning: n_neighbors = 5, weights = "uniform", p = 2 (Euclidean distance), and n_jobs = -1. Here, weights = "uniform" assigns equal weight to each of the k nearest neighbors; p = 2 defines the distance metric as Euclidean; and n_jobs = -1 enables parallel computation across all available CPU cores to accelerate distance calculations during prediction.

The core idea behind ensemble models is to combine the strengths of multiple individual models to enhance overall generalization capability [32]. Specifically, the Weighted Average Ensemble assigns a weight to each base model based on its performance on the test set; models with better predictive accuracy receive higher weights. The final prediction is the weighted sum of the predictions from all base models, as defined by the following formula:

$$\omega_i = \frac{\max(R_j^2, 0)}{\Sigma_{j=1}^{3} \max(R_j^2, 0)} \tag{3}$$

Where ω_i is the weight assigned to the *i-th* base model, and $R_i{}^2$ is the coefficient of determination (R^2) of that model on the test set.

Stacking is a two-layer nested ensemble strategy, comprising a base model layer and a meta-model layer [32-34]. In the base layer, predictions on the training set from multiple base models are generated via K-fold cross-validation; these predictions form the meta-features. Simultaneously, the base models also produce predictions on the test set, creating the corresponding test meta-features. The meta-model is then trained using these meta-features as input and the true target values as output, ultimately yielding the final prediction. This approach effectively leverages the complementary strengths of different base models while mitigating overfitting from any single model, making it theoretically one of the most accurate ensemble methods. In this study, GBR, XGBoost, and KNN served as the base models, while Ridge regression (with alpha = 0.5) was used as the meta-model. A 5-fold cross-validation scheme was applied to generate the

meta-features, further reducing the risk of overfitting.

2.4 Machine learning performance evaluation

All machine learning algorithms in this study were implemented in Python using the scikit-learn library. Model performance was evaluated using three key metrics: the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean square error (RMSE). These metrics are defined by the following equations:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - y_l)^2}{\Sigma_{i=1}^{n}(y_i - \overline{y_l})^2} \tag{4}$$

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}\left|y_i - y_l\right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - y_l)^2} \tag{6}$$

Where y_i, y_l and $\overline{y_l}$ denote the actual corrosion potential, the predicted corrosion potential, and the mean of the actual corrosion potential values, respectively.

3. Results and Discussion

3.1 Analysis of Feature Engineering Results

Figure 4 presents the Pearson correlation coefficients calculated among the original input features. The analysis reveals strong positive correlations (coefficients > 0.70) between several elemental pairs, notably C with Al, C with Mn, and Al with Mn. This high degree of collinearity indicates that these features vary in a highly synchronized manner within the dataset, which introduces a risk of multicollinearity. If unaddressed, multicollinearity can destabilize model coefficients and impair the interpretability of individual feature effects. A similarly strong correlation (0.75) is observed between Mo and S, further underscoring the prevalence of interdependent variation among composition variables. In contrast, the absolute correlation values between the target variable (corrosion potential) and most single alloying elements are notably low. This indicates that individual elements alone possess limited explanatory power for predicting corrosion potential shifts, thereby providing a clear rationale for

the subsequent construction of interactive feature terms to capture synergistic elemental effects. Based on this preliminary correlation analysis, a subset of 18 features was provisionally retained for the next stage of feature selection.

Features that contributed minimal informational value were eliminated through low-variance filtering, resulting in the removal of 8 variables at this stage. Finally, to mitigate multicollinearity, the variance inflation factor (VIF) was computed for each retained feature. Features exhibiting a VIF greater than 10-a conventional threshold signaling pronounced multicollinearity-were excluded to prevent model overfitting. After completing this three-stage feature selection process, a refined subset of 10 original features was preserved: C, Al, Mn, Mo, Si, Cr, Ni, Cu, NaCl concentration, and temperature. These were further enhanced by incorporating three engineered interaction terms (Cr×Mo, Si×Al, C×Mn) and a binary threshold feature indicating whether Cr content exceeds 1.2%. This yielded a final feature set of 14 dimensions, which was utilized for subsequent model training.

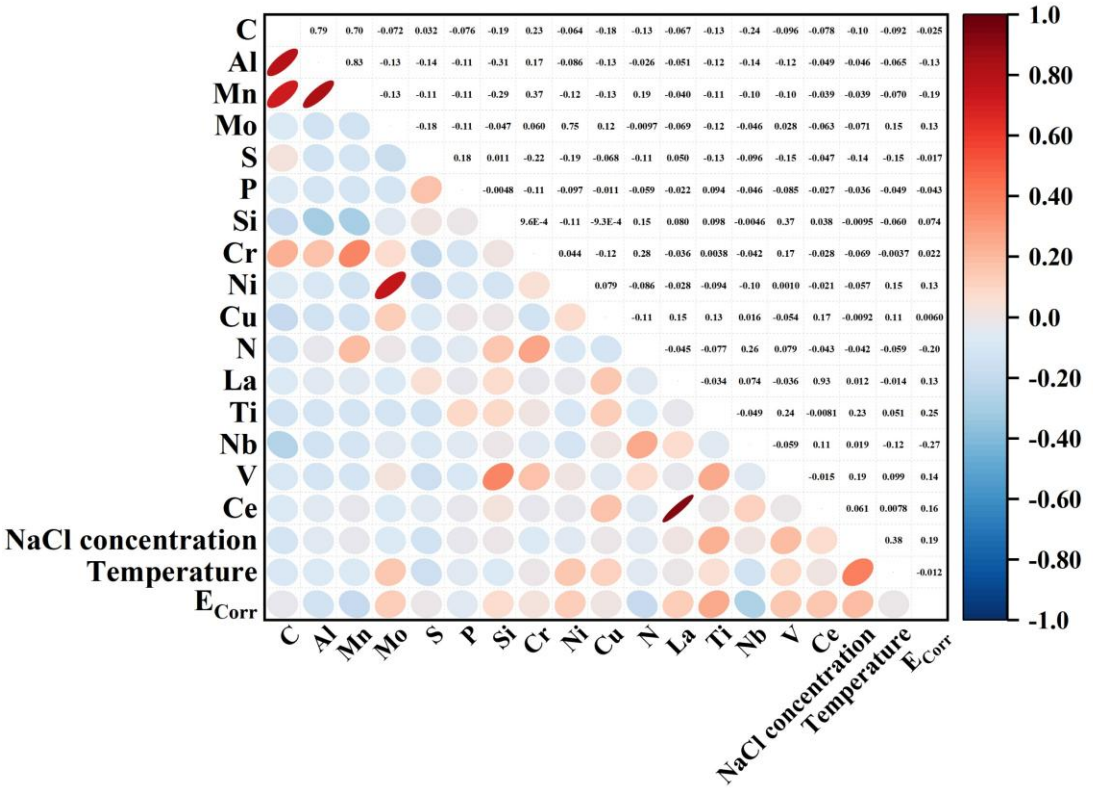


Figure 4. Pearson correlation map of the original dataset. The upper right part shows the Pearson correlation coefficients, while the lower left part uses colors to indicate

the correlation between data (red represents positive correlation and blue represents negative correlation).

3.2 Model Performance Evaluation and Comparison

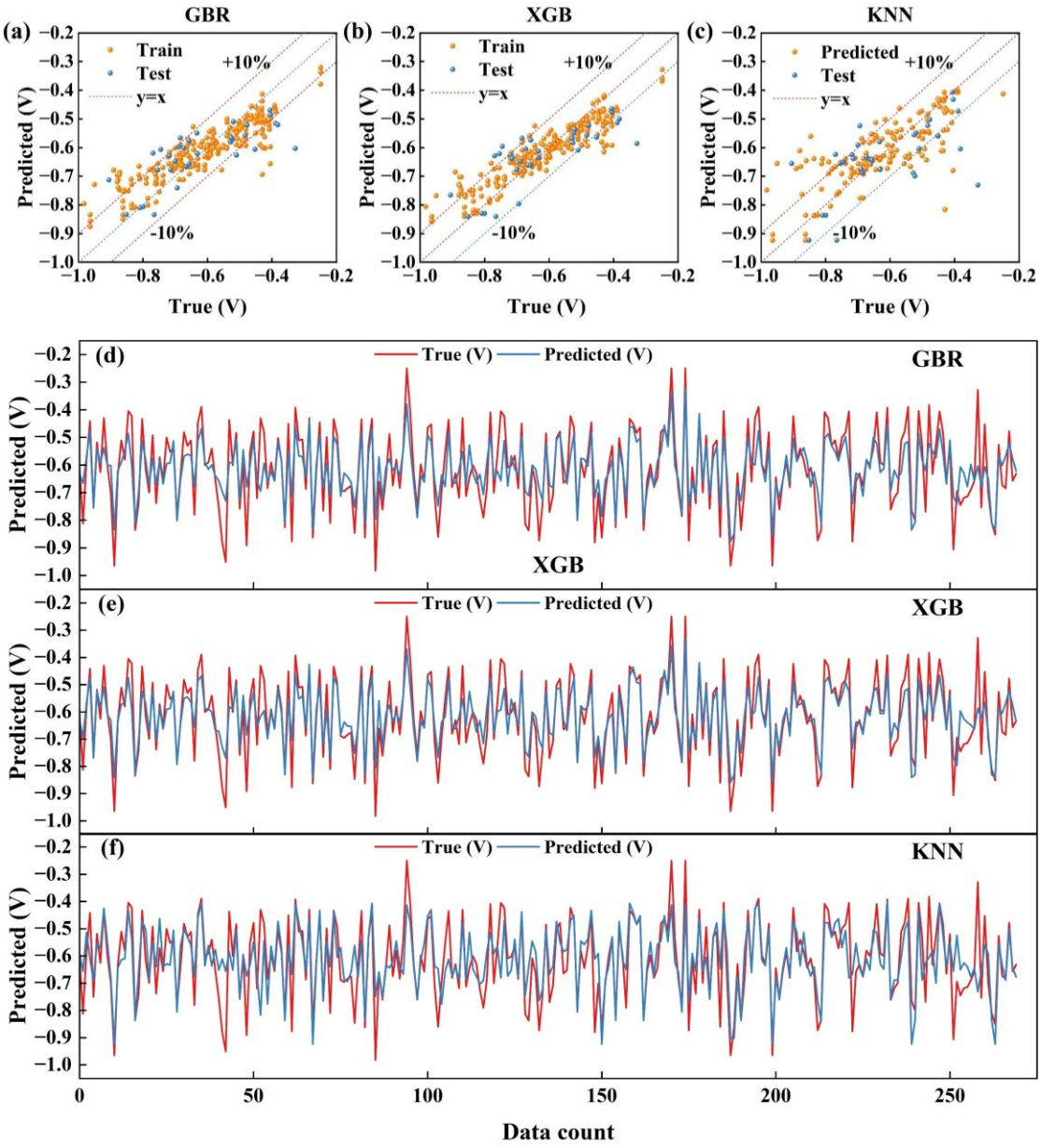


Figure 5. Linear fitting results of single models. Scatter plots: (a) GBR, (b) XGBoost, (c) KNN (the orange part represents training set samples, and the blue part represents test set samples); Line Charts: (d) GBR, (e) XGBoost, (f) KNN (the red part represents true values, and the blue part represents predicted values).

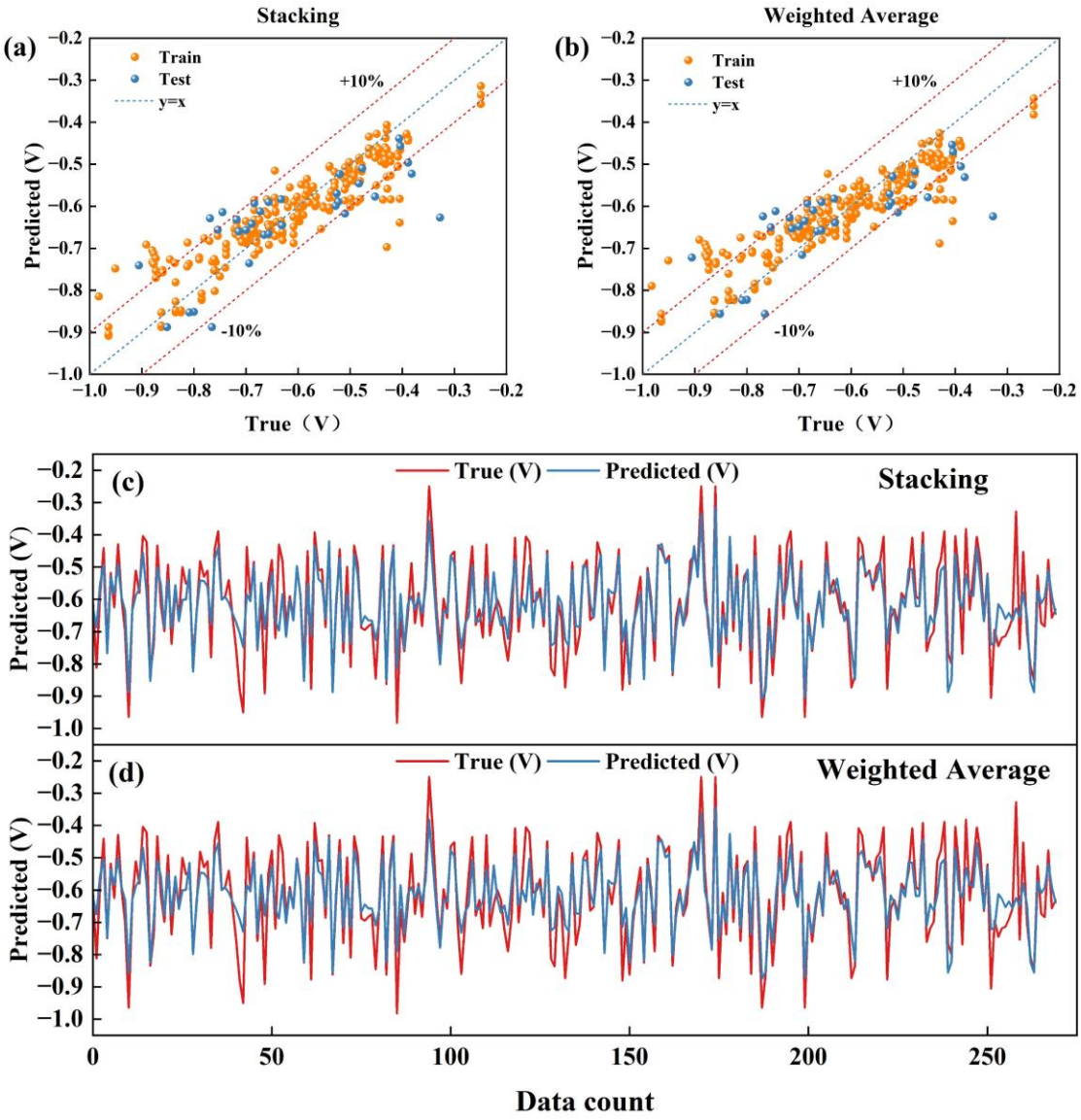


Figure 6. Linear fitting results of ensemble models. Scatter plots: (d) Stacking ensemble, (e) Weighted average ensemble (the orange part represents training set samples, and the blue part represents test set samples); Line charts: (c) Stacking ensemble, (d) Weighted average ensemble (the red part represents true values, and the blue part represents predicted values).

The predictive fitting results of each model are shown in Figures 5 and 6. Figures 5(a)-(c) present the scatter plots of the predicted versus actual corrosion potential for the three individual models: GBR, XGBoost, and KNN. In each plot, the horizontal axis represents the actual values, while the vertical axis represents the predicted values from the corresponding model. The blue dashed line indicates the line of perfect agreement (y=x), and the orange dashed lines mark the ±10% error band. Predictions falling within

this band are considered acceptable, with closer proximity to the reference line indicating higher predictive accuracy. As shown in Figure 5(a) and (b), the predictions from both the GBR and XGBoost models predominantly lie within the ±10% error margin. Specifically, the XGBoost model shows training and test data points tightly clustered around the ideal y=x line. Less than 5% of its samples fall outside the error band, and its fitted trend most closely matches the actual corrosion potential, making it the best-performing single model. The GBR model performs slightly less favorably, with data points somewhat more dispersed, yet still maintains a good overall linear correlation with the actual values. In contrast, the KNN model yields the poorest fitting results, as seen in Figure 5(c). Its data points are considerably more scattered relative to the y=x line. Most test set samples are concentrated in the -1.0 to -0.8 V range, showing a clear deviation from the reference line and a significant increase in fitting error. Figures 5(d)-(f) display line charts comparing the predicted and actual corrosion potentials across sample indices. The red curves represent the actual values, and the blue curves represent the predicted values. The XGBoost model (Figure 5(e)) demonstrates the highest degree of overlap between the two curves, indicating minimal deviation and the best trend alignment for most samples. In the GBR model (Figure 5(d)), the curves for a few samples show minor deviations, but the overall trend remains consistent with the actual values. The KNN model (Figure 5(f)) exhibits the most pronounced curve divergence, with prediction errors exceeding 0.1 V for some samples and the lowest level of trend matching.In summary, the XGBoost model demonstrates the best overall fitting accuracy and sample-wise predictive capability, establishing it as the optimal choice among the single models. The KNN model underperforms in both aspects. This may be attributed to the high-dimensional feature space, where distances between all points become less distinguishable, and the presence of low-relevance features or noise, which can degrade the prediction accuracy of distance-based algorithms like KNN.

To enhance the generalization capability of the models, two ensemble methods-Stacking and Weighted Average-were constructed. Their fitting results are presented in Figure 6. As shown in Figure 6(a) and (b), the test set predictions from both ensemble

models largely fall within the ±10% error band. However, the test set points of the Stacking ensemble model are more tightly clustered around the ideal reference line (y=x), with less than 3% of samples deviating outside the ±10% band. This indicates that the Stacking model effectively leverages complementary information from its base learners. Although the fitting performance of the Weighted Average ensemble is not as strong as that of the Stacking ensemble, the overall distribution of its test set predictions remains superior to those of the individual GBR and KNN models.

Figure 6(c) and (d) show the line charts comparing the predicted and actual values for the two ensemble models. A comparison between Figure 5(e) (XGBoost) and Figure 6(c) (Stacking) reveals that the degree of overlap between the actual and predicted curves is very similar. However, in sample intervals where the corrosion potential exhibits significant fluctuations (samples 100–150), the Stacking ensemble model's predictions follow the changing trend of the actual values more promptly and accurately, with deviations generally controlled within 0.05 V. Figure 6(d) displays the results for the Weighted Average ensemble. The overlap between its predicted and actual curves is slightly lower. This may be attributed to the model's direct use of the final predictions from individual models for weighted averaging, unlike the Stacking model, which employs a meta-learner to learn from the intermediate representations generated by the base models.

The predictive performance of each model was quantified using three standard regression metrics: the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). A comparative statistical summary of all five models is presented in the bar chart in Figure 7. In the chart, solid blue bars denote the training set R^2, while hatched blue bars represent the test set R^2. Correspondingly, solid orange bars indicate the MAE values, and hatched orange bars depict the RMSE values. These metrics collectively provide a comprehensive assessment of model accuracy, error magnitude, and generalization ability across both training and test datasets.

The coefficient of determination (R^2) reflects a model's ability to explain the variance in the target variable, with values closer to 1 indicating a better fit. As shown in Figure 7, the Stacking ensemble and XGBoost models achieved the highest training

set R^2 values of 0.81 and 0.80, respectively, indicating the most effective learning of patterns from the training data. In contrast, the KNN model had a significantly lower training set R^2 of 0.64, consistent with its limited capacity to fit complex, high-dimensional corrosion-related features. On the test set, the KNN model's R^2 was only 0.34, showing a substantial performance gap compared to other models. XGBoost achieved the highest test set R^2 among all models at 0.62. Among the three single models, XGBoost delivered the best overall results, while KNN exhibited severe overfitting, further highlighting its inadequacy in handling high-dimensional feature spaces.

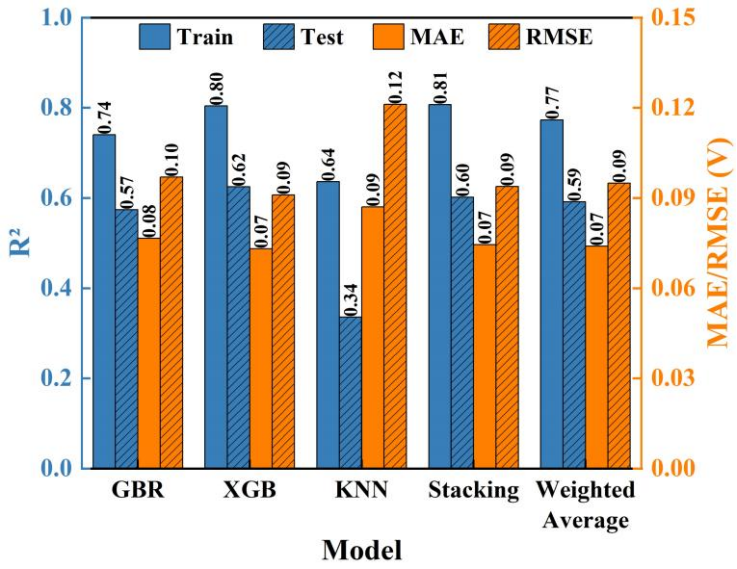


Figure 7. Comparison chart of fitting results of various models.

The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) measure the deviation between predicted and actual values, with lower values indicating higher predictive accuracy. XGBoost, the Stacking ensemble, and the Weighted Average ensemble all achieved the lowest errors, with an MAE of 0.07 V and an RMSE of 0.09 V. The GBR model showed slightly higher errors (MAE of 0.08 V, RMSE of 0.10 V) compared to XGBoost. This difference may be attributed to the absence of L1/L2 regularization in GBR, making it more susceptible to overfitting noise in the training data and resulting in larger prediction deviations.

In summary, among the individual models, XGBoost demonstrates the best fitting performance and highest predictive accuracy. Its regularization mechanism and parallel

computing architecture make it well-suited for handling the high-dimensional features involved in predicting the corrosion potential of marine engineering steels. While the Stacking ensemble achieves error metrics on the test set comparable to those of XGBoost, its degree of overfitting is slightly higher. Although the Stacking framework theoretically integrates the complementary strengths of tree-based models (XGBoost and GBR) and a distance-based model (KNN)-balancing predictive accuracy with generalization capability and potentially offering more stable predictions for extreme corrosion potential samples-the adverse influence of the weaker KNN model may have contributed to its marginally elevated overfitting. Therefore, XGBoost emerges as the optimal choice among the five evaluated models, effectively balancing model fit and predictive precision.

3.3 Feature Importance Analysis

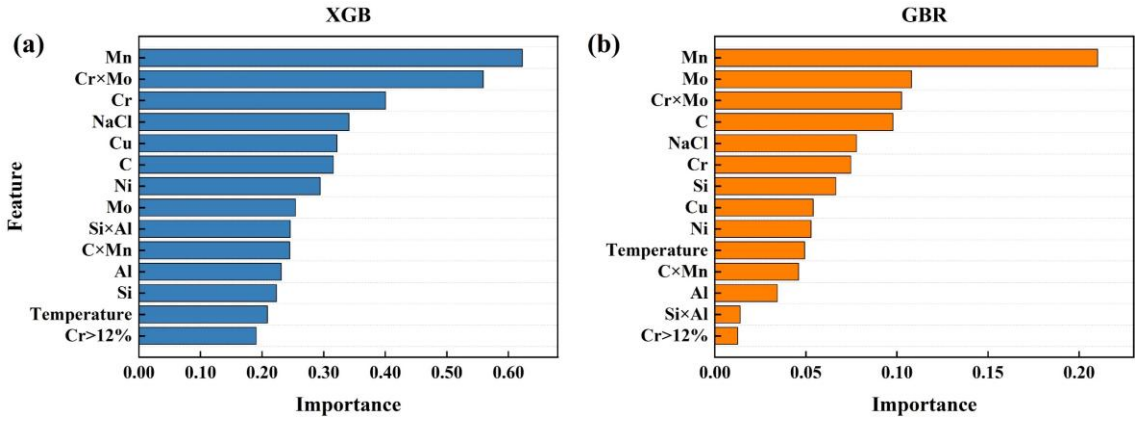


Figure 8. Feature importance plots of (a) XGBoost and (b) GBR.

Feature importance reflects the contribution of input features to a model's predictions and serves as a critical bridge between machine learning outputs and the underlying physical corrosion mechanisms. In contrast, KNN makes predictions based on global distance calculations, where all features are inherently treated as equally relevant during distance computation. The algorithm itself lacks an intrinsic mechanism to evaluate which features are more discriminative during the training phase. Therefore, this study analyzes feature importance only for XGBoost and GBR, as shown in Figure 8(a) and (b).

As shown in Figure 8(a), the top three contributing features in the XGBoost model are Mn, Cr×Mo, and Cr. In Figure 8(b), the top three for the GBR model are Mn, Mo,

and Cr×Mo. Mn is a common alloying element in steel, which not only enhances mechanical properties but also influences corrosion potential by modifying the compactness of the rust layer-a effect supported by numerous studies, explaining its highest contribution in both models. Cr is a key element for passive film formation, while Mo promotes passivation and refines the film structure to inhibit pitting, with their synergistic interaction enhancing corrosion resistance. The high importance of these features aligns well with established corrosion principles. While the core influential features are similar between the two models, their specific ranking differs due to inherent algorithmic distinctions: GBR relies on first-order gradient fitting, making it more sensitive to the local contribution of individual features, hence ranking Mo above Cr×Mo. In contrast, XGBoost employs second-order Taylor expansion and regularization, enabling more precise capture of feature interactions, which elevates the relative importance of the interaction term. These findings confirm that alloying elements and their interactions are indeed core determinants of corrosion potential. Additionally, the environmental feature NaCl concentration also ranks notably high, underscoring the regulatory role of environmental factors.

3.4 SHAP Feature Contribution Analysis

To further quantify how specific feature values influence the prediction of corrosion potential and to intuitively reveal the underlying mechanism by which each feature affects steel corrosion behavior, this study employed SHAP (Shapley Additive Explanations) analysis to interpret the predictions of the XGBoost model. The results are presented in Figure 9.

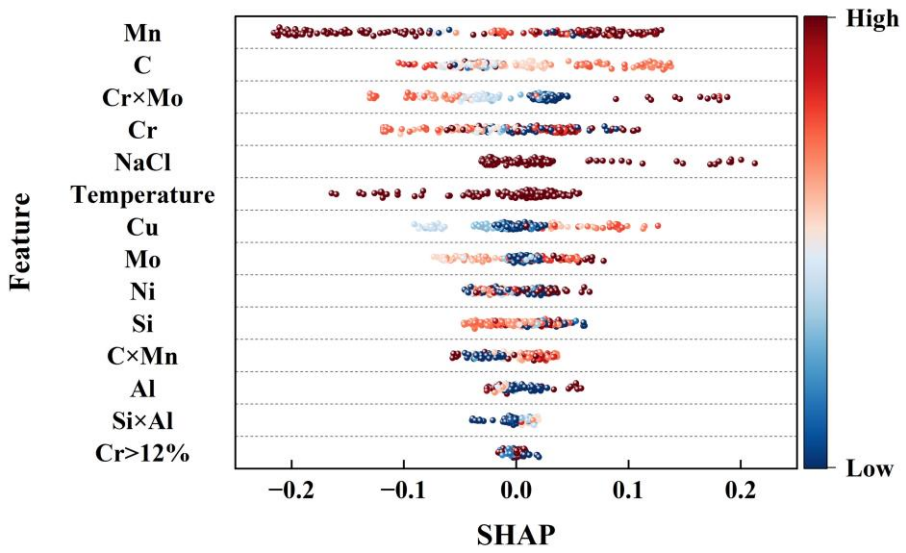


Figure 9. SHAP summary plot of the XGBoost model.

To further quantify the impact of specific feature values on the prediction of corrosion potential and to intuitively reveal the underlying mechanisms through which each feature influences steel corrosion behavior, this study employed SHAP (Shapley Additive Explanations) analysis to interpret the predictions of the XGBoost model. The results are presented in Figure 9. In the figure, the vertical axis lists the feature names, and the horizontal axis represents the SHAP value. The SHAP value quantifies the contribution of a feature to the predicted corrosion potential. A positive SHAP value indicates that the feature increases the predicted corrosion potential (enhancing corrosion resistance), while a negative value indicates a decrease (reducing corrosion resistance). The features are arranged on the vertical axis in descending order of importance, which shows a high degree of alignment with the feature importance ranking from XGBoost (Figure 8(a)). Key features such as Mn, Cr×Mo, Cr, NaCl concentration, and Cu are consistently identified as dominant factors influencing corrosion potential, mutually validating the reliability of the importance assessment. Each point in the SHAP summary plot corresponds to a sample, with its color gradient from red (high feature value) to blue (low feature value) clearly illustrating how the feature magnitude regulates the SHAP value. For example, the points for Mn are predominantly red, with SHAP values distributed between -0.3 and 0.2, indicating a bidirectional influence. Moderate to low Mn content (mostly positive SHAP values)

enhances corrosion resistance, whereas excessive Mn (negative SHAP values) diminishes it. This aligns with findings by Guanqiao Su et al., who reported that high Mn content adversely affects corrosion potential [35,36]. The interaction feature Cr×Mo, ranked second in importance, demonstrates a synergistic passivation effect in the SHAP plot: high values contribute positively, while low values contribute negatively. This reflects the positive interactive role of Cr and Mo oxides in forming a stable passive film [25]. Furthermore, when NaCl concentration is high (red points), the corresponding SHAP values tend to be positive and show greater dispersion. This suggests that high-salinity environments accelerate the penetration and adsorption of corrosive ions, thereby increasing the corrosion rate.

The uniqueness of this SHAP summary plot lies in its transformation of the "black-box prediction" of machine learning into a visually interpretable and quantitatively supported explanation of corrosion mechanisms. It quantitatively validates established theories, such as the bidirectional role of Mn and the synergistic passivation effect of Cr×Mo. These analytical results provide a quantified, data-driven technical basis for optimizing the composition of marine engineering steels and for managing corrosion risks in specific environments. They also offer a theoretical foundation for the design of subsequent corrosion protection strategies.

3.5 Experimental Validation

To validate the practical feasibility and generalization capability of the model, independently designed electrochemical experiments were conducted to obtain test data for experimental validation of the optimal XGBoost model.

3.5.1 Experimental Design

The chemical compositions of the five experimental steels (Steel-A to Steel-E) are listed in Table 1. Electrochemical tests were performed using a three-electrode system on an electrochemical workstation. A 3.5 wt% NaCl solution served as the electrolyte. Each Steel specimen was used as the working electrode, a platinum plate as the counter electrode, and a saturated calomel electrode (SCE) as the reference electrode. All experiments were conducted at room temperature (25 ± 1 °C).

Table 1 Chemical Composition of the Tested Steel.

Steel	C	Si	Mn	P	S	Al	Nb	V	Ti	Cu	Cr	Ni	Mo	B	N	Ca	Fe
Steel-A	0.06	0.26	1.64	0.01	0.0008	0.031	0.0447	0.003	0.012	0.0167	0.25	0.0167	0.01	0.000174	0.0034	0.002	Bal.
Steel-B	0.068	0.31	1.52	0.006	0.005	0	0	0	0	0.29	0.46	0.21	0.21	0	0	0	Bal.
Steel-C	0.058	0.28	1.85	0.004	0.006	0	0	0	0.016	0.26	0	0	0.26	0	0	0	Bal.
Steel-D	0.15	0.2	1	0.0058	0.0014	0.036	0	0.03	0	0.0091	0.99	1.45	0.37	0	0	0	Bal.
Steel-E	0.059	0.092	1.1	0.005	0.011	0	0	0	0	0	0	0	0.85	0	0	0	Bal.

The potentiodynamic polarization tests were scanned from -0.4 V to 1.0 V (vs. SCE). To ensure data reliability, each test was repeated at least three times. The corrosion potential was determined using the Tafel extrapolation method.The material composition and environmental parameters were compiled into a validation dataset and input into the trained XGBoost model to obtain predicted corrosion potentials. Model performance was evaluated using the absolute error and relative error between the predicted and measured values.

3.5.2 Result

The predictive performance of the XGBoost model for the corrosion potential of the tested steels is summarized in Table 2. Taking Steel-A as an example, the model output a corrosion potential of –0.77436 V (vs. SCE), while the experimentally measured value was –0.73569 V (vs. SCE). The resulting absolute error is 0.03867 V, corresponding to a relative error of 5.26%. These figures indicate a close agreement between prediction and measurement, demonstrating the model's practical predictive capability. The model's accuracy was evaluated systematically using both absolute error and relative error between predicted and measured corrosion potentials across all validation samples. The small magnitude of these errors, consistently below 0.06 V in absolute terms and largely within 10% relatively, supports the robustness of the trained XGBoost model. This performance can be attributed to the prior feature-engineering steps-such as introducing interaction terms and a threshold feature for Cr content-which allowed the model to capture key electrochemical influences identified in the training phase.

Moreover, the errors are distributed without obvious bias, suggesting that the model generalizes adequately within the compositional and environmental space covered by the training data. The largest relative error observed (10.15% for Steel-C) remains within an acceptable range for corrosion-potential prediction, further affirming the utility of the proposed machine-learning framework for rapid, data-assisted assessment of marine engineering steels. Overall, the validation results confirm that the XGBoost model, enriched with domain-knowledge-driven features, provides reliable estimates of corrosion potential, thereby offering a useful tool for preliminary material screening and composition optimization.

Table 2 Comparison of True and Predicted Corrosion Potentials of the Tested Steel.

Steel	True (V, vs. SCE)	Predicted (V, vs. SCE)	Absolute Error (V)	Relative Error (%)
Steel-A	-0.736	-0.774	0.038	5.26
Steel-B	-0.623	-0.647	0.024	3.85
Steel-C	-0.591	-0.651	0.060	10.15
Steel-D	-0.619	-0.654	0.035	5.65
Steel-E	-0.532	-0.580	0.048	9.02

4 Conclusions

This study aimed to predict the corrosion potential of marine engineering steels. Employing the scikit-learn library in Python, five machine learning models were developed. By integrating feature engineering and interpretability analysis, the research systematically investigated the influence of alloying elements and environmental factors on corrosion potential. The main conclusions are as follows:

(1) The input feature set was optimized by introducing element interaction terms such as Cr×Mo and the compositional threshold feature Cr>12%. A comparison of the GBR, XGBoost, KNN, Stacking ensemble, and Weighted Average ensemble models revealed that XGBoost achieved the best fit and highest predictive accuracy among all models. The Stacking ensemble demonstrated comparable prediction error and generalization capability but exhibited a slightly higher degree of overfitting, indicating

potential for further optimization.

(2) Feature importance and SHAP interpretability analyses identified the core features governing the corrosion potential of marine engineering steels and clarified their mechanisms of influence. Key findings include: the bidirectional influence of Mn; the synergistic effect of Cr×Mo in enhancing passive film stability; and the accelerated penetration and adsorption of corrosive ions with increasing NaCl concentration. These results quantitatively bridge data-driven predictions with corrosion electrochemical theory, validating the physical interpretability of the model.

(3) Validation experiments confirmed that the XGBoost model developed in this study can effectively predict the corrosion potential of marine engineering steels. It provides reliable guidance for alloy composition design and offers an accurate and reliable assessment of the corrosion resistance for low-alloy steels, carbon steels, and high-strength steels.

Acknowledgments

This work was financially supported by the Lianyungang Key R&D Program (No. CG2410) and the Lianyungang Postdoctoral Research Funding Program (No. LYG20230006).

Author Contributions

B.W.: Data curation, Conceptualization, Investigation, Writing-original draft, Formal analysis. Y.L.: Investigation, Writing-review & editing. W.Y.: Investigation, editing. E.F.: Supervision, Conceptualization, Investigation, Formal analysis, Writing-review &editing.

Competing Interests

The authors declare no competing interests.

References

[1] Bao J, Wu P, Zhang J, et al. Machine learning models for predicting the corrosion behavior of a plain carbon steel in simulated marine environment[J]. Journal of Industrial and Engineering Chemistry, 2025.

[2] Yan L, Diao Y, Lang Z, et al. Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach[J]. Science and technology of Advanced Materials, 2020, 21(1): 359-370.

[3] Farh H, Seghier M, Zayed T. comprehensive review of corrosion protection and control techniques for metallic pipelines[J]. Engineering Failure Analysis, 2023, 143: 106885.

[4] Wasim M, Shoaib S, Mubarak N M, et al. Factors influencing corrosion of metal pipes in soils, Environmental Chemistry Letters. 2018, 16 (3): 861-879.

[5] Koch G. Cost of corrosion[J]. Trends in oil and gas corrosion research and technologies, 2017: 3-30.

[6] Diao Y, Yan L, Gao K. Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features[J]. Materials & Design, 2021, 198: 109326.

[7] Wen C, Zhang Y, Wang C, et al. Machine learning assisted design of high entropy alloys with desired property[J]. Acta Materialia, 2019, 170: 109-117.

[8] Mu W, Rahaman M, Rios F L, et al. Predicting strain-induced martensite in austenitic steels by combining physical modelling and machine learning[J]. Materials & Design, 2021, 197: 109199.

[9] Wan H Y, Chen G F, Li C P, et al. Data-driven evaluation of fatigue performance of additive manufactured parts using miniature specimens[J]. Journal of Materials Science & Technology, 2019, 35(6): 1137-1146.

[10] Yu J, Wang C, Chen Y, et al. Accelerated design of L12-strengthened Co-base superalloys based on machine learning of experimental data[J]. Materials & Design, 2020, 195: 108996.

[11] Lu Q, Liu S, Li W, et al. Combination of thermodynamic knowledge and multilayer

feedforward neural networks for accurate prediction of MS temperature in steels[J]. Materials & Design, 2020, 192: 108696.

[12] Pruksawan S, Lambard G, Samitsu S, et al. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning[J]. Science and technology of advanced materials, 2019, 20(1): 1010-1021.

[13] Shin D, Lee S, Shyam A, et al. Petascale supercomputing to accelerate the design of high-temperature alloys[J]. Science and Technology of advanced MaTerialS, 2017, 18(1): 828-838.

[14] Xue D, Xue D, Yuan R, et al. An informatics approach to transformation temperatures of NiTi-based shape memory alloys[J]. Acta Materialia, 2017, 125: 532-541.

[15] Kamrunnahar M, Urquidi-Macdonald M. Prediction of corrosion behaviour of Alloy 22 using neural network as a data mining tool[J]. Corrosion science, 2011, 53(3): 961-967.

[16] Ben M E A S, Truong T T, Feiler C, et al. A hybrid deep learning model for predicting atmospheric corrosion in steel energy structures under maritime conditions based on time-series data[J]. Results in Engineering, 2025, 25: 104417.

[17] Lu S, He J, Xue N, et al. Data-driven corrosion behavior and prediction of EH36 steel in marine atmosphere: Integrating corrosion big data with machine learning[J]. Journal of Materials Science & Technology, 2025.

[18] Peng H, Xia F, Gu Y, et al. Investigating the effect of Cr content on the microstructure and electrochemical measurement of low alloy steel[J]. Materials Today Communications, 2024, 39: 108716.

[19] Moon J, Ha H Y, Park S J, et al. Effect of Mo and Cr additions on the microstructure, mechanical properties and pitting corrosion resistance of austenitic Fe-30Mn-10.5 Al-1.1 C lightweight steels[J]. Journal of Alloys and Compounds, 2019, 775: 1136-1146.

[20] Zhang T, Liu W, Dong B, et al. Corrosion of Cu-doped Ni–Mo low-alloy steel in a severe marine environment[J]. Journal of Physics and Chemistry of Solids, 2022, 163: 110584.

[21] Li D, Fu G, Zhu M. Effect of Si on corrosion behavior of model bridge steels by alternative wetting/drying test in an artificial medium simulated hot and humid atmosphere of marine and industrial area [J]. Journal of Chinese Society for Corrosion and Protection, 2016, 36(5): 433-440.

[22] Nishimura T. Corrosion resistance of Si–Al-bearing ultrafine-grained weathering steel[J]. Science and Technology of Advanced Materials, 2008, 9(1): 013005.

[23] Nishimura T. Rust formation and corrosion performance of Si-and Al-bearing ultrafine grained weathering steel[J]. Corrosion Science, 2008, 50(5): 1306-1312.

[24] Nishimura T. Electrochemical behaviour and structure of rust formed on Si-and Al-bearing steel after atmospheric exposure[J]. Corrosion science, 2010, 52(11): 3609-3614.

[25] Yan X, Yan L, Kang S, et al. Corrosion behavior and electrochemical corrosion of a high manganese steel in simulated marine splash zone[J]. Materials Research Express, 2021, 8(12): 126507.

[26] Zhang D, Gao X, Su G, et al. Corrosion behavior of low-C medium-Mn steel in simulated marine immersion and splash zone environment[J]. Journal of Materials Engineering and Performance, 2017, 26(6): 2599-2607.

[27] Dong L, Zhang G, Li S, et al. Insights into the stress corrosion cracking of M54 ultra-high-strength steel in the marine environments at varying temperatures by comparison of slow strain rate tensile and crack growth tests[J]. Corrosion Science, 2024, 241: 112543.

[28] Pathy A, Meher S. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods[J]. Algal Research, 2020, 50: 102006.

[29] Song Y, Huang Z, Jin M, et al. Machine learning prediction of biochar physicochemical properties based on biomass characteristics and pyrolysis conditions[J]. Journal of Analytical and Applied Pyrolysis, 2024, 181: 106596.

[30] Muja M, Lowe D G. Scalable nearest neighbor algorithms for high dimensional data[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(11): 2227-2240.

[31] Leng L, Zhou J, Zhang W, et al. Machine-learning-aided hydrochar production through hydrothermal carbonization of biomass by engineering operating parameters and/or biomass mixture recipes[J]. Energy, 2024, 288: 129854.

[32] Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one?[J]. Machine learning, 2004, 54(3): 255-273.

[33] Sattar K, Zahra S Z, Faheem M, et al. Stacked ensemble model for tropical cyclone path prediction[J]. IEEE Access, 2023, 11: 69512-69521.

[34] Kumari P, Singh K R, Kumar R. Stacking ensemble algorithm to predict re-keying in group key management[J]. Arabian Journal for Science and Engineering, 2025: 1-15.

[35] Su G, Gao X. Comparison of medium manganese steel and Q345 steel on corrosion behavior in a 3.5 wt% NaCl solution[J]. Materials, 2017, 10(8): 938.

[36] Su G, Gao X, Du L, et al. Influence of Mn on the corrosion behaviour of medium manganese steels in a simulated seawater environment[J]. International Journal of Electrochemical Science, 2016, 11(11): 9447-9461.