SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Machine Learning-Based Prediction of Corrosion Rate in Reinforced Concrete Incorporating Corrosion Inhibitors

Ranveer Sahu[1], Md Daniyal[2*], Rahat Yezdani[3]

[1]Department of Civil Engineering, Sandip University, Madhubani, India-847235, ranveersahu151095@gmail.com
[2*]Department of Civil Engineering, Sandip University, Madhubani, India-847235, daniyalzhcet@gmail.com
[3]Department of Computer Science & Engineering, Purnea College of Engineering, Purnea, India-854303, rahatyajdani123@gmail.com

ABSTRACT
Corrosion of reinforcing steel is a major durability concern in reinforced concrete, particularly under chloride-rich exposure conditions. This study develops a machine-learning framework for predicting the corrosion rate of reinforced concrete incorporating calcium nitrite and diethanolamine as corrosion inhibitors. Experimental corrosion-rate data were developed using sodium chloride concentration (NaCl), inhibitor dosage (DOI), and exposure duration (T) as input variables, with corrosion rate expressed in mm/year as the target variable. An 80:20 training–testing strategy was adopted to develop Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models. Five-fold cross-validation with grid-search optimization was employed to determine the optimum model parameters. Model performance was evaluated using the correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and scatter index (SI). The optimized RF model exhibited the highest predictive capability on the independent test dataset, achieving R = 0.999628, R^2 = 0.999208, RMSE = 0.000014 mm/year, MAE = 0.000006 mm/year, MAPE = 3.21%, and SI = 0.053698. XGBoost also demonstrated strong predictive performance, with R = 0.992982 and R^2 = 0.980223, whereas optimized SVR achieved R = 0.960408 and R^2 = 0.788177. The corresponding five-fold cross-validation values were 0.994667, 0.933424, and 0.852174 for RF, XGBoost, and SVR, respectively. Feature-importance and SHAP analyses consistently identified inhibitor dosage as the most influential predictor, followed by exposure duration and NaCl concentration. The results demonstrate that machine learning, particularly RF, can effectively predict corrosion rate within the investigated experimental domain and provide a useful data-driven approach for assessing corrosion behaviour in inhibitor-modified reinforced concrete.

Keywords: Reinforced concrete; Corrosion rate; Calcium nitrite; Diethanolamine; Random Forest; XGBoost; Support Vector Regression

1. INTRODUCTION

Reinforced concrete is one of the most widely used construction materials because of its favourable mechanical performance, versatility, durability, and relatively low maintenance requirements. Nevertheless, the long-term durability of reinforced concrete structures can be severely affected by deterioration of the embedded reinforcement, particularly under chloride-contaminated and other aggressive exposure conditions (Bertolini et al., 2013; Casanova et al., 2023; Tiwari et al., 2025). Chloride ingress can progressively disrupt the passive condition of reinforcing steel, initiating electrochemical corrosion when critical environmental and electrochemical conditions are reached. The resulting corrosion products occupy a greater volume than the original steel and can generate internal stresses within the surrounding concrete. Continued corrosion may consequently produce cracking, delamination, concrete spalling, reduction in reinforcement cross-section, deterioration of steel–concrete bond, and loss of structural capacity (Leporace-Guimil et al., 2023). Recent investigations have also demonstrated that corrosion-induced deterioration is strongly affected by environmental and material parameters, making reliable assessment of corrosion behaviour important for durability-based design and maintenance planning (Sun et al., 2022).

Chloride-induced corrosion is particularly important for reinforced concrete structures exposed to marine environments, seawater, de-icing salts, saline water, and other chloride-bearing environments. The corrosion process is governed by the combined influence of chloride concentration, moisture condition, concrete permeability, chloride transport, oxygen availability, exposure duration, reinforcement

characteristics, and the condition of the steel–concrete interface (J. Liu et al., 2024). Experimental studies have demonstrated that moisture content, chloride concentration, and reinforcement deterioration are closely associated with the measurable response of reinforced concrete subjected to chloride exposure. Because these parameters interact in a nonlinear manner, corrosion behaviour cannot always be represented adequately using simple empirical relationships. Furthermore, conventional experimental evaluation of corrosion behaviour is generally time-consuming and requires controlled testing under specific combinations of material and environmental conditions. Consequently, predictive techniques capable of complementing experimental investigations are increasingly relevant for reinforced-concrete durability assessment (Tesic et al., 2023).

Corrosion-inhibiting admixtures represent one of the important strategies for reducing the susceptibility of embedded reinforcement to corrosion. These admixtures can act by maintaining or enhancing steel passivation, modifying the electrochemical condition at the steel–concrete interface, or reducing the effects of aggressive ions. Recent reviews have classified corrosion inhibitors for reinforced concrete into inorganic and organic systems and have highlighted the continuing interest in nitrite-based inhibitors, amines, alkanolamines, carboxylates, and other organic and inorganic compounds. The effectiveness of an inhibitor, however, depends on its chemical composition, dosage, concrete characteristics, chloride exposure, and stage of application. Therefore, evaluation of corrosion inhibitors under controlled exposure conditions remains an important research area (Angst, 2018; Bertolini et al., 2013; J. Huang et al., 2022; Wang et al., 2023).

Calcium nitrite (CN) is one of the most extensively investigated inorganic corrosion inhibitors for reinforced concrete. Earlier experimental investigations established that calcium-nitrite-based inhibitors can reduce the corrosion rate of embedded steel and increase the chloride threshold level associated with corrosion initiation (Ann et al., 2006). More recent electrochemical investigations have similarly reported improved corrosion resistance with increasing calcium nitrite concentration, including increased charge-transfer resistance and modification of the protective film formed on the reinforcement surface (Jinsong et al., 2022). The effectiveness of nitrite-based inhibitors is generally associated with their ability to support the formation and stabilization of a protective oxide film on steel. In addition to their electrochemical influence, nitrite ions can interact with the cementitious system and affect chloride binding and related physicochemical processes. Molecular-scale investigations have further examined the transport of nitrite species through cementitious nanopores, providing additional insight into the behaviour of nitrite-based inhibitors in concrete systems. These findings establish calcium nitrite as a technically important reference inhibitor for chloride-induced reinforcement corrosion (Balonis & Glasser, 2011).

Organic corrosion inhibitors provide another approach to mitigating reinforcement corrosion. Recent studies have investigated organic, migratory, hydrophobic, and amino-alcohol-based inhibitor systems for reinforced concrete exposed to chloride and coupled aggressive environments. In particular, amino-alcohol compounds have attracted attention because of their ability to interact with the steel–concrete interface and influence the electrochemical processes associated with corrosion. Experimental investigations of amino-alcohol migratory inhibitors have reported improvements in corrosion resistance under chloride exposure, although the effectiveness can depend strongly on the timing of inhibitor application and the extent of prior chloride deterioration. Diethanolamine (DEA), an alkanolamine compound, therefore represents an important organic inhibitor for investigation within this broader class of corrosion-control materials. However, compared with calcium nitrite and several commercially investigated migratory inhibitors, the predictive relationship between DEA dosage, chloride exposure, exposure duration, and experimentally measured corrosion rate remains insufficiently explored (Hu et al., 2021; Ormellese et al., 2009; Xu et al., 2016).

The effectiveness of corrosion inhibitors can also vary with exposure severity and deterioration stage. For example, recent studies have shown that migratory corrosion inhibitors can improve the corrosion resistance of reinforced concrete subjected to chloride attack, while their effectiveness may decrease when applied after substantial chloride deterioration has already occurred. Other investigations have examined inhibitor performance under coupled carbonation and chloride exposure and have demonstrated the importance of environmental conditions in determining corrosion-inhibition efficiency (Wang et al., 2023). Studies of organic corrosion inhibitors have likewise shown that changes in surface condition, hydrophobicity, and inhibitor–steel interaction can influence the resulting corrosion resistance (J. Huang et al., 2022). These observations indicate that corrosion-inhibitor performance cannot be evaluated solely on the basis of inhibitor type; dosage and exposure history must also be considered.

Experimental electrochemical techniques, including potentiodynamic polarization and electrochemical impedance spectroscopy, provide valuable information for quantifying reinforcement corrosion and evaluating inhibitor performance (Dhouibi et al., 2002; Shen et al., 2020). Nevertheless, experimental programmes generally generate discrete observations corresponding to particular combinations of inhibitor dosage, chloride concentration, and exposure duration. Extending such experiments to a large number of combinations can require substantial time, materials, and laboratory resources. In addition, corrosion behaviour is affected by multiple interacting variables, and the resulting relationships may be

nonlinear and difficult to describe using conventional regression equations. These limitations provide a strong motivation for developing data-driven predictive models that can learn relationships directly from experimental observations.

Machine learning (ML) has increasingly been applied to durability and deterioration problems in cement-based materials. Recent studies have demonstrated that ML models can predict chloride migration and diffusion characteristics using concrete mixture composition, curing, and environmental variables (Hosseinzadeh et al., 2023). For example, machine-learning models incorporating Random Forest, Support Vector Regression, XGBoost, and artificial neural networks have been successfully applied to predict chloride diffusion coefficients and chloride resistance of concrete (W. Liu et al., 2024; Zheng & Cai, 2024). Large experimental databases have also been used to investigate chloride transport using ensemble learning and explainable modelling techniques (Li et al., 2024). Such studies demonstrate the potential of ML to provide rapid predictions of durability-related properties that would otherwise require extensive experimental evaluation.

The application of ML to reinforcement corrosion is also developing rapidly. Recent research has investigated data-driven prediction of corrosion levels, corrosion-related deterioration, corrosion-induced structural response, and corrosion detection in reinforced concrete (Gao et al., 2024). For example, interpretable ML approaches have been used to predict corrosion levels of rust-damaged reinforced concrete and to quantify the contribution of individual input variables using SHAP analysis (T. Huang et al., 2025). Other recent studies have applied ML to corrosion prediction in marine reinforced-concrete structures and to the assessment of the deteriorated response of corroded reinforced-concrete beams (Selvaprasanth et al., 2025). Machine-learning techniques have also been combined with experimental sensing methods, including laser-induced breakdown spectroscopy, to predict corrosion-related parameters in reinforced concrete (Matthews et al., 2024). These developments indicate that ML can provide a useful complementary approach to experimental corrosion assessment.

Among supervised-learning algorithms, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) are particularly suitable for nonlinear regression problems. RF is an ensemble learning technique that combines multiple decision trees and aggregates their predictions, providing robustness and the ability to represent nonlinear relationships and interactions among variables. XGBoost extends the gradient-boosting framework by constructing decision trees sequentially to minimize prediction errors and incorporating regularization to control model complexity. SVR employs a kernel-based formulation to determine a regression function within a specified error tolerance and is particularly useful for nonlinear relationships and relatively small experimental datasets. These three approaches therefore provide complementary modelling characteristics and enable a meaningful comparison between bagging-based ensemble learning, boosting-based ensemble learning, and kernel-based regression.

The increasing use of ML in concrete durability research also highlights the importance of model interpretation. High predictive accuracy alone does not necessarily explain why a model produces a particular prediction or which experimental variables exert the greatest influence. Explainable artificial intelligence techniques, particularly SHAP (SHapley Additive exPlanations), provide a systematic means of quantifying the contribution of individual features to model predictions. Recent concrete durability studies have demonstrated the usefulness of SHAP and related feature-analysis approaches for identifying the variables governing chloride transport and corrosion-related deterioration. Incorporating feature-importance and SHAP analyses into corrosion prediction can therefore provide additional insight beyond conventional statistical performance indicators.

Despite these developments, an important research gap remains in the direct prediction of experimentally measured corrosion rate in reinforced concrete containing different dosages of specific corrosion inhibitors. Existing ML investigations have largely focused on chloride diffusion, chloride permeability, corrosion detection, corrosion level, structural deterioration, or broader durability indicators. Comparatively limited attention has been given to establishing a direct data-driven relationship between inhibitor dosage, chloride concentration, exposure duration, and corrosion rate. In particular, the comparative predictive performance of RF, XGBoost, and SVR for corrosion-rate estimation in reinforced concrete incorporating both an inorganic inhibitor such as calcium nitrite and an organic alkanolamine such as diethanolamine has not been adequately established.

In the present study, an experimental dataset comprising 162 observations is employed to develop predictive models for the corrosion rate of reinforced concrete incorporating calcium nitrite and diethanolamine. Three independent variables are considered: sodium chloride concentration, expressed as NaCl (%); corrosion-inhibitor dosage, expressed as DOI (%); and exposure duration, expressed as T (days). The experimentally measured corrosion rate, expressed in mm/year, is adopted as the target variable. The dataset therefore establishes a direct relationship between environmental exposure, inhibitor dosage, exposure duration, and measured reinforcement corrosion behaviour.

The principal objective of this study is to develop and comparatively evaluate Random Forest, XGBoost, and Support Vector Regression models for predicting the corrosion rate of inhibitor-modified reinforced concrete. An 80:20 training–testing strategy is employed, while five-fold cross-validation and systematic

grid-search hyperparameter optimization are used to improve model reliability. Model performance is assessed using the correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and scatter index (SI). In addition to predictive accuracy, feature-importance and SHAP analyses are employed to determine the relative contribution of NaCl concentration, inhibitor dosage, and exposure duration to the developed predictions.

The novelty of the present investigation lies in integrating experimentally measured corrosion behaviour of calcium-nitrite- and diethanolamine-modified reinforced concrete with comparative and interpretable machine-learning prediction. Unlike studies that focus primarily on corrosion detection, chloride transport, or structural deterioration after corrosion, the present work directly predicts experimentally measured corrosion rate from inhibitor dosage, chloride concentration, and exposure duration. The combined use of RF, XGBoost, and SVR, together with cross-validation, hyperparameter optimization, feature-importance analysis, and SHAP interpretation, provides a comprehensive data-driven framework for evaluating corrosion behaviour. The resulting framework is intended to complement experimental assessment and provide a rapid predictive tool for evaluating corrosion-control strategies within the investigated experimental domain.

2. Experimental Program

2.1 Research Framework

The experimental program was designed to investigate the influence of calcium nitrite (CN) and diethanolamine (DEA) on the mechanical and corrosion behaviour of reinforced concrete and subsequently develop machine-learning models for predicting corrosion rate. M30-grade concrete was selected as the reference concrete, while CN and DEA were incorporated at different dosages. The experimental investigation consisted of concrete production, fresh and hardened concrete characterization, exposure of reinforced concrete specimens to aggressive environments, and electrochemical assessment of reinforcement corrosion. The experimentally obtained corrosion-rate data were subsequently used to develop Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models.

The overall experimental framework comprised five major stages: (i) preparation of control and inhibitor-modified concrete mixtures, (ii) accelerated corrosion exposure and electrochemical measurements, and (iii) development of an experimental database for machine-learning-based prediction.

2.2 Materials

Ordinary Portland cement (OPC) was used as the primary cementitious material for producing M30 concrete. Fine aggregate and coarse aggregate conforming to the relevant Indian Standard requirements were used for preparing the concrete mixtures. Potable water was used for concrete mixing and initial curing. A polycarboxylate ether (PCE)-based superplasticizer was incorporated to achieve the required workability without substantially increasing the water-to-cement ratio.

Calcium nitrite was selected as the inorganic corrosion inhibitor, whereas diethanolamine was used as the organic corrosion inhibitor. Both inhibitors were incorporated into the concrete at controlled dosages expressed as a percentage of the cement mass. The selected inhibitor concentrations were 1%, 2%, 3%, 4%, and 5%. A concrete mixture without corrosion inhibitor was used as the control mixture (CM).

2.3 Concrete Mix Proportioning

M30-grade concrete was prepared using a water-to-cement ratio of 0.45. The control mixture contained approximately 400 kg/m^3 of cement, 670 kg/m^3 of fine aggregate, 1180 kg/m^3 of coarse aggregate, and 180 L/m^3 of mixing water. Approximately 4 kg/m^3 of PCE-based superplasticizer was used to maintain suitable workability.

For the inhibitor-modified mixtures, calcium nitrite and diethanolamine were incorporated at 1%, 2%, 3%, 4%, and 5% of the cement mass. The corresponding mixtures were designated as CN1, CN2, CN3, CN4, and CN5 for calcium nitrite and DEA1, DEA2, DEA3, DEA4, and DEA5 for diethanolamine. The control mixture was designated CM. Except for the inhibitor dosage, the principal mixture constituents and water-to-cement ratio were maintained consistently to facilitate comparison among the mixtures.

2.4 Reinforced Concrete Specimens for Corrosion Testing

Accelerated corrosion experiments were performed using cylindrical reinforced concrete specimens measuring 80 mm in diameter and 160 mm in height. A steel reinforcement bar was positioned centrally within each specimen, providing a nominal concrete cover of approximately 35 mm. The reinforcement bar was extended approximately 30 mm beyond the top surface of the specimen to provide an electrical connection for electrochemical measurements.

After casting and demoulding, the specimens were subjected to potable water curing for 28 days then exposed to saline environments (0%, 2% and 4% NaCl). To maintain relatively consistent exposure conditions, the solutions were replaced at two-week intervals throughout the exposure period. Care was taken to maintain consistent reinforcement positioning, concrete cover, specimen dimensions, and exposure conditions so that variations in measured corrosion behaviour could primarily be attributed to

the inhibitor type, inhibitor dosage, and exposure environments.

2.5 Accelerated Corrosion Testing

The corrosion behaviour of the embedded reinforcement was evaluated using an electrochemical polarization technique. A Tafel potentiodynamic polarization method was employed to determine the electrochemical corrosion characteristics of the reinforcement. An ACM corrosion monitoring instrument was used for electrochemical measurements.

The exposed reinforcement served as the working electrode, while appropriate counter and reference electrodes were employed to complete the electrochemical measurement system. The polarization response was recorded after stabilization of the specimen under the specified exposure condition. The corrosion potential and corrosion current density were obtained from the polarization response, and the corrosion rate was subsequently determined from the electrochemical parameters.

The corrosion rate, expressed in mm/year, was considered the principal response variable for the machine-learning investigation. Lower corrosion-rate values indicated greater corrosion protection provided by the corresponding inhibitor system.

2.6 Experimental Variables Used for Machine-Learning Prediction

The electrochemical investigation generated a database containing 162 observations of corrosion rate. For machine-learning modelling, the experimental parameters were organized into predictor and response variables. Three principal predictor variables such as NaCl concentration (%), Corrosion-inhibitor dosage (%) and Exposure duration (days) were considered. The corresponding experimentally measured corrosion rate (mm/year) was selected as the target variable.

The inclusion of these three predictors was based on their direct relevance to reinforcement corrosion. Chloride concentration represents the severity of the chloride exposure, inhibitor dosage represents the level of corrosion protection, and exposure duration accounts for the temporal development of corrosion. Their combined consideration enables the machine-learning models to capture nonlinear interactions among environmental severity, inhibitor concentration, and exposure time.

2.7 Experimental Database

The final database consisted of 162 experimental observations obtained from the corrosion investigation. Each observation contained the corresponding exposure condition, inhibitor dosage, exposure duration, and experimentally determined corrosion rate. The database incorporated measurements associated with both calcium nitrite and diethanolamine systems and their respective dosage levels.

Prior to model development, the dataset was examined for completeness, consistency, and physically meaningful trends. The data were subsequently divided into training and testing subsets. The training data were used to establish the relationship between the predictor variables and corrosion rate, whereas the independent testing data were reserved for evaluating the predictive capability of the developed models.

3. Dataset Development and Machine-Learning Methodology

3.1 Development of the Dataset

The experimental corrosion database developed in the present investigation was used to establish machine-learning models for predicting the corrosion rate of reinforced concrete containing calcium nitrite (CN) and diethanolamine (DEA). The database comprised 162 experimental observations obtained under different combinations of chloride exposure, corrosion-inhibitor dosage, and exposure duration. The experimentally measured corrosion rate, expressed in mm/year, was considered the dependent variable.

Three parameters were selected as predictor variables based on their physical relevance to reinforcement corrosion: NaCl concentration (%), corrosion-inhibitor dosage (%), and exposure duration (days). NaCl concentration represents the severity of chloride exposure, inhibitor dosage represents the level of corrosion protection, and exposure duration accounts for the progressive development of corrosion.

The input matrix was therefore defined as:

$$X = [X1, X2, X3]$$

where (X1) is NaCl concentration (%), (X2) is inhibitor dosage (%), and (X3) is exposure duration (days). The target variable was defined as:

$$Y = CR$$

where (CR) represents the experimentally measured corrosion rate in mm/year.

The selected variables were used to establish a nonlinear relationship between environmental severity, inhibitor concentration, exposure duration, and corrosion rate. This formulation also allows the developed models to be used for estimating corrosion behaviors for combinations of input parameters within the range represented by the experimental dataset.

3.2 Data Preparation and Computational Framework

The machine-learning analysis was performed in a cloud-based Python environment using open-source libraries. NumPy and pandas were used for numerical computation and data management, scikit-learn for preprocessing, model development, cross-validation, hyperparameter optimization, and performance evaluation, XGBoost for gradient-boosting regression, and Matplotlib for visualization. A fixed random state

of 42 was used to ensure reproducibility.

The dataset comprising 162 observations was initially examined for missing values, duplicate records, inconsistent entries, and potential data errors. Descriptive statistics and Pearson correlation analysis were performed to characterize the variables and assess their relationships. Exploratory plots were also generated to identify data patterns and potential nonlinear relationships.

The dataset was divided into 80% training and 20% testing subsets using random state 42. The training subset was used for model fitting, five-fold cross-validation, and hyperparameter optimization, while the testing subset was retained exclusively for final performance evaluation. This procedure ensured an independent assessment of model generalization to unseen observations.

3.3 Machine-Learning Models

Three supervised machine-learning regression algorithms—Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR)—were employed to predict the corrosion rate of reinforced concrete incorporating calcium nitrite and diethanolamine. The models were developed using NaCl concentration (%), inhibitor dosage (%), and exposure duration (days) as input variables, while the experimentally measured corrosion rate (mm/year) was considered the target variable. The three algorithms were selected to enable comparison between ensemble tree-based and kernel-based regression approaches for modelling the nonlinear relationship between the exposure conditions, inhibitor dosage, and corrosion response.

Random Forest Regression

Random Forest (RF) is an ensemble-learning algorithm that combines the predictions of multiple decision trees to obtain a robust regression output. Each tree is developed using a randomly selected subset of the training observations and predictor variables, while the final prediction is obtained by averaging the individual tree predictions:

$$\hat{P} = \frac{1}{N}\sum_{j=1}^{N} T_j\,(X)$$

where $\hat{P}$ is the predicted corrosion rate, N is the number of decision trees, T_j (X) represents the prediction of the tree, and X denotes the input feature vector. RF was selected because of its ability to capture nonlinear relationships and interactions among NaCl concentration, inhibitor dosage, and exposure duration without requiring a predefined functional relationship.

The RF model was optimized by varying the number of estimators, maximum tree depth, maximum number of features considered for splitting, minimum samples required for a split, and minimum samples required at a leaf node. The optimized RF configuration obtained through five-fold cross-validation consisted of 600 estimators, unlimited maximum depth, maximum feature fraction of 1.0, minimum samples per leaf of 1, and minimum samples per split of 2.

Extreme Gradient Boosting Regression

Extreme Gradient Boosting (XGBoost) is an ensemble regression technique based on sequential construction of decision trees. Each successive tree is developed to reduce the residual errors produced by the existing ensemble, thereby progressively improving prediction accuracy. The general objective function can be expressed as:

$$L = \sum_{i=1}^{n} l(O_i, P_i) + \sum_{j} \Omega(f_j)$$

where O_i, and P_i represent the observed and predicted corrosion rates, respectively, $l(O_i, P_i)$ denotes the loss function, and $\Omega(f_j)$ represents the regularization term associated with the j^{th} tree. The regularization component controls model complexity and helps reduce overfitting.

The XGBoost model was optimized with respect to the number of estimators, learning rate, maximum tree depth, minimum child weight, subsampling ratio, and column-sampling ratio. The optimized configuration consisted of 200 estimators, a learning rate of 0.1, maximum tree depth of 2, minimum child weight of 1, subsampling ratio of 0.8, and column-sampling ratio of 1.0.

Support Vector Regression

Support Vector Regression (SVR) is a kernel-based regression method that determines an optimal regression function while controlling model complexity and prediction error. In the present study, an RBF kernel was adopted to represent the nonlinear relationship between the input variables and corrosion rate:

$$K(x_i, x_j) = exp(-\gamma|x_i - x_j|^2)$$

where $K(x_i, x_j)$ denotes the similarity between the i^{th} and j^{th} input observations, x_i and x_j represent their corresponding input feature vectors, γ is the kernel coefficient, and $|x_i - x_j|^2$ represents the squared Euclidean distance between the two observations.

The SVR model uses an ϵ-insensitive loss formulation, allowing prediction errors within the specified tolerance to be disregarded during optimization. The principal SVR hyperparameters were the regularization parameter C, insensitive loss parameter ϵ, and RBF kernel coefficient γ. The input variables were standardized before SVR modelling. The optimized SVR configuration obtained from five-fold cross-

validation was C=1000, ϵ=1×10-6, and γ=0.01.

3.4 Hyperparameter Optimization and Cross-Validation

The predictive performance of machine-learning models is strongly influenced by the selection of appropriate hyperparameters. Therefore, systematic hyperparameter optimization was performed for the Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models rather than relying on their default configurations. The optimization was conducted using GridSearchCV within the training dataset, with five-fold cross-validation and a fixed random state of 42.

In five-fold cross-validation, the training dataset was divided into five approximately equal subsets. During each iteration, four subsets were used for model training and the remaining subset was used for validation. This process was repeated five times so that each subset was used once for validation. The cross-validation performance was subsequently averaged to obtain a more robust estimate of model performance.

The hyperparameter search spaces were selected according to the characteristics of each algorithm. For RF, the principal parameters included the number of estimators, maximum tree depth, maximum number of features, minimum samples required for splitting, and minimum samples required at a leaf node. For XGBoost, the parameters included the number of estimators, learning rate, maximum tree depth, minimum child weight, subsampling ratio, and column-sampling ratio. For SVR, the regularization parameter (C), insensitive loss parameter (ϵ), and RBF kernel coefficient (γ) were optimized.

3.5 Model Performance Evaluation

The predictive performance of the developed Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) models was evaluated using widely accepted statistical performance indicators. These metrics were selected to assess the correlation, goodness of fit, magnitude of prediction errors, relative prediction error, and normalized error between the experimentally observed and model-predicted corrosion rates. The evaluation metrics included the Correlation Coefficient (R), Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Scatter Index (SI).

Let Oi represent the experimentally observed corrosion rate, Pi represent the model-predicted corrosion rate, O represent the mean observed corrosion rate, P represent the mean predicted corrosion rate, and n denote the total number of observations.

Correlation Coefficient (R)

The correlation coefficient (R) measures the strength of the linear association between the experimentally observed and predicted corrosion rates. Values approaching unity indicate a strong positive correlation and, consequently, better agreement between the model predictions and experimental observations.

$$R = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}}$$

Coefficient of Determination (R^2)

The coefficient of determination (R^2) represents the proportion of variation in the experimentally observed corrosion rate explained by the predictive model. Values approaching unity indicate a high degree of agreement between observed and predicted values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$$

Root Mean Square Error (RMSE)

RMSE quantifies the magnitude of prediction errors and gives greater weight to larger deviations because the errors are squared before averaging. Lower RMSE values indicate better predictive accuracy.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}}$$

Mean Absolute Error (MAE)

MAE represents the average absolute difference between the observed and predicted corrosion rates. Lower MAE values indicate closer agreement between the model predictions and experimental measurements.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i|$$

Mean Absolute Percentage Error (MAPE)

MAPE expresses the prediction error relative to the observed corrosion rate in percentage terms. Lower MAPE values indicate better predictive accuracy.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|O_i - P_i|}{|O_i|}$$

Scatter Index (SI)

The Scatter Index (SI) is a normalized error measure obtained by relating RMSE to the mean observed corrosion rate. Lower SI values indicate better predictive performance.

$$SI = \frac{RMSE}{\bar{O}}$$

3.6 Feature Importance and Model Interpretability

Feature-importance and model-interpretability analyses were performed to quantify the contribution of the input variables to the predicted corrosion rate. The three input variables considered were NaCl concentration (NaCl, %), inhibitor dosage (DOI, %), and exposure duration (T, days).

For the Random Forest (RF) and XGBoost models, the inherent feature-importance measures provided by the respective algorithms were extracted from the optimized models. These measures quantify the relative contribution of each predictor to the model's decision-making process.

For the Support Vector Regression (SVR) model, which does not provide an intrinsic feature-importance measure, permutation importance was employed. In this approach, the values of one feature are randomly permuted while the remaining features are kept unchanged. The resulting deterioration in model performance is used to quantify the importance of that feature. A greater reduction in model performance indicates a greater contribution of the corresponding predictor.

In addition, SHapley Additive exPlanations (SHAP) analysis was applied to the optimized RF and XGBoost models to provide a more detailed interpretation of their predictions. SHAP is based on Shapley values from cooperative game theory and quantifies the contribution of individual features to a model prediction relative to its baseline output. The mean absolute SHAP value was used to determine the overall importance of each feature, with larger values indicating a greater contribution to the model output. The feature-importance and SHAP analyses were subsequently used to establish the relative influence of NaCl concentration, inhibitor dosage, and exposure duration on corrosion-rate prediction.

4. Results and Discussion

4.1 Dataset Characteristics and Correlation Analysis

The developed dataset comprised 162 experimental observations representing the corrosion behaviour of reinforced concrete under different combinations of chloride concentration, inhibitor dosage, and exposure duration. The three predictor variables considered in the machine-learning models were NaCl concentration (NaCl, %), inhibitor dosage (DOI, %), and exposure duration (T, days), while the experimentally measured corrosion rate (CR, mm/year) was considered the response variable. The dataset was subsequently divided into training and testing subsets in an 80:20 ratio for machine-learning development and independent validation.

A Pearson correlation analysis was initially performed to examine the linear relationships among the input variables and corrosion rate. The resulting correlation matrix is shown in Fig. 1.

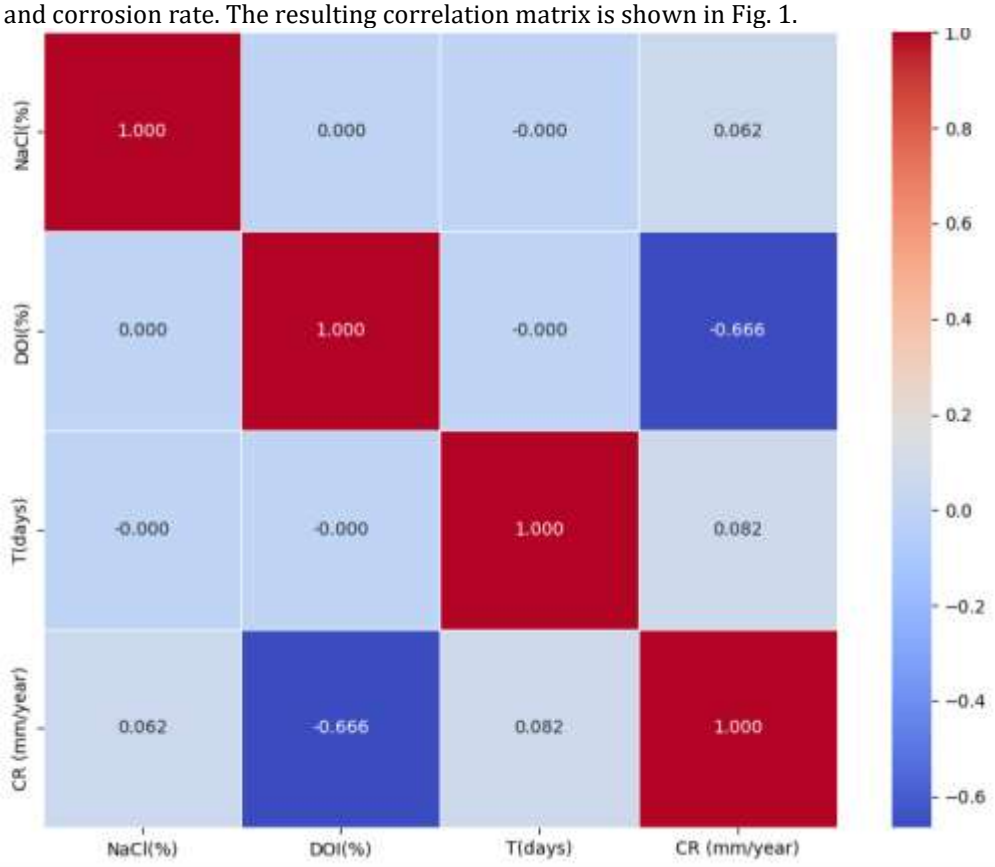

Figure 1. Pearson correlation matrix of NaCl concentration, inhibitor dosage, exposure duration, and corrosion rate

The correlation analysis reveals that inhibitor dosage exhibits the strongest linear association with corrosion rate, with a correlation coefficient of −0.6656. The negative correlation indicates that increasing inhibitor dosage is generally associated with a reduction in the measured corrosion rate. This trend is consistent with the expected corrosion-inhibiting action of calcium nitrite and diethanolamine, which can reduce the electrochemical activity of reinforcing steel and consequently decrease the corrosion rate.

In contrast, the linear correlation between NaCl concentration and corrosion rate was relatively weak. Similarly, exposure duration exhibited a weak positive correlation with corrosion rate. These relatively low correlation coefficients should not necessarily be interpreted as indicating that NaCl concentration and exposure duration have no influence on corrosion behaviour. Rather, they indicate that their individual linear relationships with corrosion rate are weak within the investigated dataset.

The correlation coefficients among the three predictor variables were approximately zero, with NaCl concentration–DOI, NaCl concentration–exposure duration, and DOI–exposure duration correlations of approximately 0.0000, −0.0000, and −0.0000, respectively. This indicates that the predictor variables were effectively uncorrelated with one another in the developed dataset. Consequently, the dataset does not exhibit an apparent linear multicollinearity problem among the selected input variables.

An important observation from the correlation analysis is the difference between the relatively weak individual correlations of NaCl concentration and exposure duration with corrosion rate and the strong predictive performance subsequently obtained from the nonlinear machine-learning models. This suggests that the corrosion response may involve nonlinear relationships and interactions among chloride concentration, inhibitor dosage, and exposure duration that cannot be adequately represented using simple pairwise linear correlations. Therefore, the use of RF, XGBoost, and SVR is justified for capturing such complex relationships.

4.2 Baseline Model Performance

The initial predictive performance of the three machine-learning algorithms was evaluated using their baseline configurations before hyperparameter optimization. Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) were trained using the 80% training dataset and subsequently evaluated on both the training and independent testing datasets. The results obtained from the baseline models are presented in Table 1. The correlation coefficient for the baseline SVR was not reported because the corresponding calculation returned an undefined value.

The agreement between experimentally measured and model-predicted corrosion rates was further examined using measured-versus-predicted plots. The plots for the baseline RF, XGBoost, and SVR models are presented in Fig. 2, Fig. 3, and Fig. 4, respectively. The ideal 1:1 line represents perfect agreement between the measured and predicted corrosion rates.

Table 1. Baseline performance of RF, XGBoost, and SVR

Model	Dataset	R	R^2	RMSE	MAE	MAPE	SI
RF	Train	0.999763	0.999499	0.000014	0.000006	1.357878	0.038299
RF	Test	0.999583	0.999163	0.000015	0.000006	3.273934	0.055198
XGBoost	Train	0.979758	0.942402	0.000153	0.000086	58.35019	0.410658
XGBoost	Test	0.991287	0.975748	0.00008	0.000057	60.74501	0.297134
SVR	Train	—	−1.779531	0.001064	0.001024	1271.21	2.852735
SVR	Test	—	−3.490352	0.001084	0.001055	1383.447	4.043099

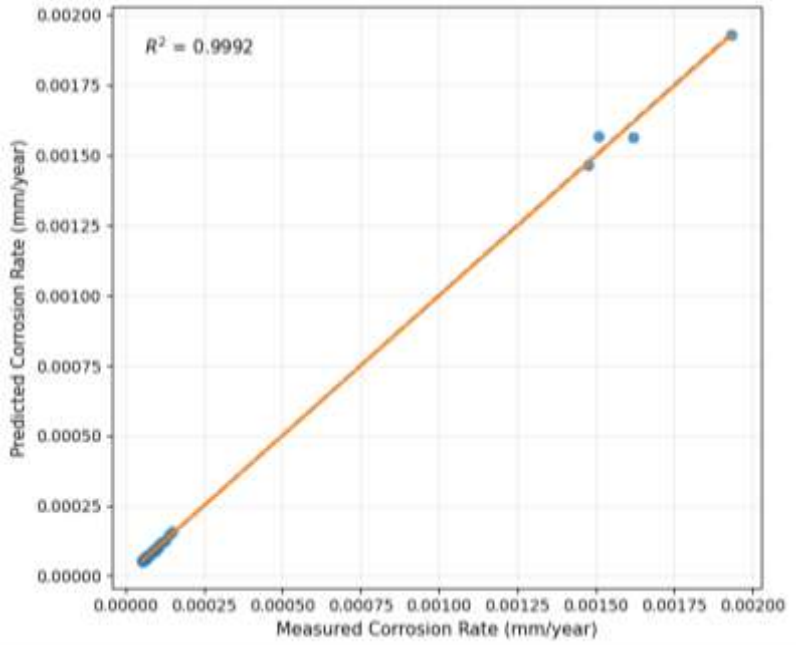

Figure 2. Measured versus predicted corrosion rate for the baseline RF model

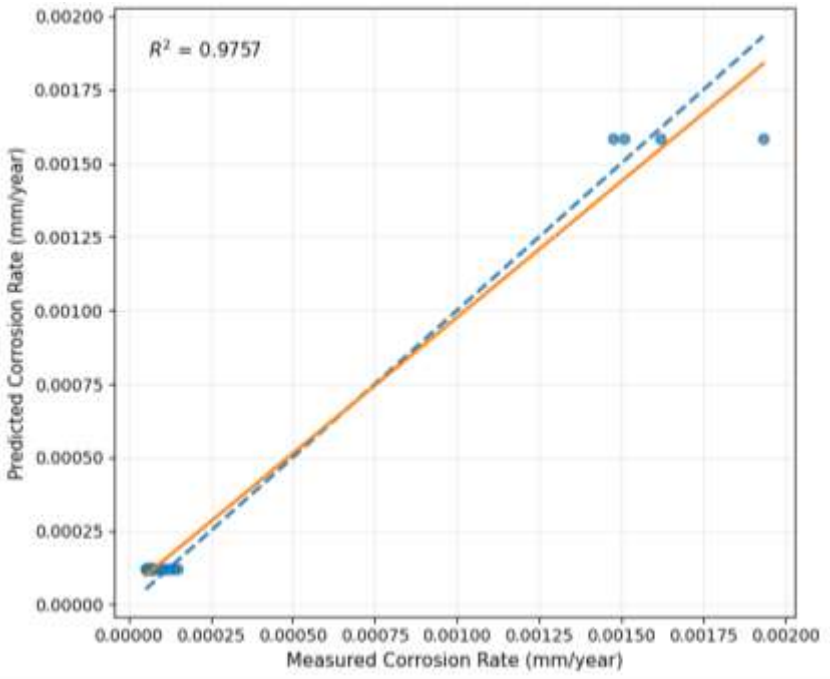

Figure 3. Measured versus predicted corrosion rate for the baseline XGBoost model

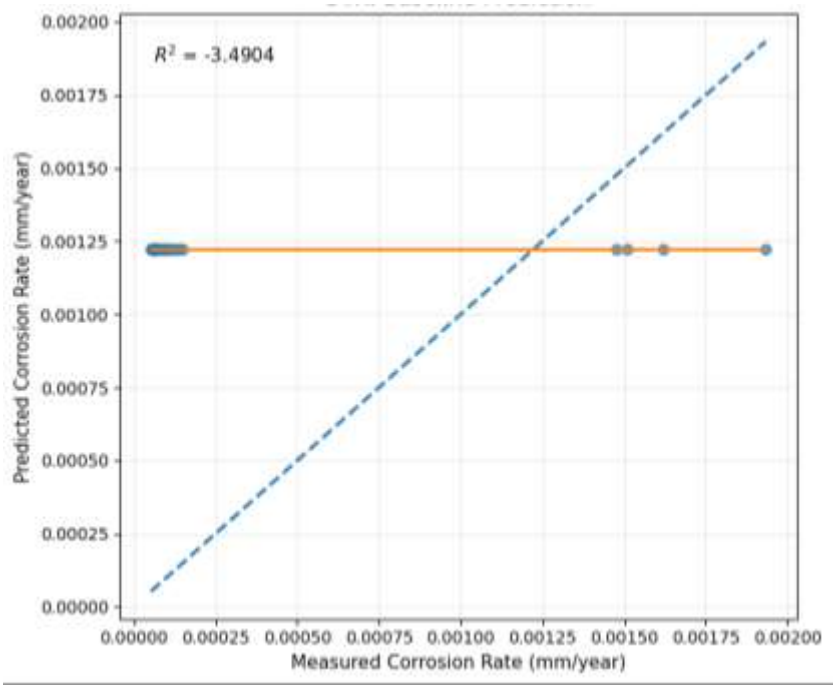

Figure 4. Measured versus predicted corrosion rate for the baseline SVR model

The RF model exhibited the strongest baseline performance, with a test-set R of 0.999583 and R^2 of 0.999163. The corresponding RMSE, MAE, and SI were 0.000015 mm/year, 0.000006 mm/year, and 0.055198, respectively. The measured-versus-predicted plot (Fig. 2) showed that the RF predictions were closely distributed around the ideal 1:1 line, confirming its high predictive accuracy and ability to capture the nonlinear relationship between the input variables and corrosion rate. This performance can be attributed to the ensemble-tree structure of RF, which combines multiple decision trees and effectively captures nonlinear relationships and interactions among the input parameters without requiring an explicitly defined functional relationship.

The XGBoost model also showed strong predictive capability, achieving a test-set R of 0.991287 and R^2 of 0.975748, with RMSE, MAE, and SI values of 0.000080 mm/year, 0.000057 mm/year, and 0.297134, respectively. The measured-versus-predicted plot (Fig. 3) indicates a generally close agreement with the ideal 1:1 line, although greater scatter was observed compared with RF. The MAPE of 60.75% should be interpreted cautiously because the very small corrosion-rate values can cause relatively small absolute deviations to produce large percentage errors. The results indicate that XGBoost effectively captured the nonlinear corrosion behaviour, but its prediction errors were higher than those of RF for the present dataset.

In contrast, the baseline SVR model exhibited considerably weaker predictive performance, with a test-set R^2 of −3.490352, RMSE of 0.001084 mm/year, MAE of 0.001055 mm/year, and SI of 4.043099. The negative indicates that the model performed substantially worse than a predictor based on the mean observed corrosion rate. The measured-versus-predicted plot (Fig. 4) further confirms the poor agreement between the predicted and experimental values. The MAPE of 1383.45% should again be interpreted cautiously because of the very small magnitude of the corrosion-rate values. Similar behaviour was observed for the training dataset, which yielded an R^2 of −1.779531. The poor baseline performance does not necessarily indicate that SVR is unsuitable for corrosion-rate prediction, as its performance is highly dependent on the selection of the kernel and associated hyperparameters.

The baseline comparison therefore establishes a clear initial performance hierarchy of RF > XGBoost > SVR. The exceptionally high RF performance can be attributed to its ensemble tree structure, which is well suited to capturing nonlinear relationships within the corrosion dataset. XGBoost also provided a strong baseline prediction, whereas the default SVR configuration was inadequate for the present dataset and required more extensive parameter optimization.

The baseline results also demonstrate the importance of hyperparameter optimization, particularly for SVR. Although RF already provided excellent predictive accuracy before optimization, XGBoost and especially SVR had greater scope for improvement through systematic parameter tuning.

4.3 Hyperparameter Optimization

To improve the predictive capability of the baseline models, systematic hyperparameter optimization was performed for Random Forest (RF), XGBoost, and Support Vector Regression (SVR) using GridSearchCV with five-fold cross-validation. The optimization was conducted exclusively on the training dataset, while the independent test dataset was retained for subsequent final evaluation. The optimization results are

summarized in Table 2.

Table 2. Optimized hyperparameters and cross-validation performance

Model	Optimized hyperparameters	Best CV (R^2)	Optimization time (s)
Random Forest	n_estimators=600; max_depth=None; max_features=1.0; min_samples_leaf=1; min_samples_split=2	0.994667	741.58
XGBoost	n_estimators=200; learning_rate=0.1; max_depth=2; min_child_weight=1; subsample=0.8; colsample_bytree=1.0	0.933424	41.93
SVR	C=1000; ϵ =1×10-6; γ =0.01	0.852174	11.63

The optimized RF model achieved the highest cross-validation performance, with an of 0.994667. The selected configuration employed 600 decision trees with no imposed maximum tree depth, while the complete set of available features was considered during splitting. The minimum samples per leaf and minimum samples required for splitting were set to 1 and 2, respectively. The high cross-validation R^2 indicates that RF maintained excellent predictive capability across the five validation folds.

The XGBoost model achieved a cross-validation R^2 of 0.933424. The optimized configuration consisted of 200 estimators, a learning rate of 0.1, and a maximum tree depth of 2. The relatively shallow tree depth, together with a subsampling ratio of 0.8, provided a controlled model complexity while retaining the ability to capture nonlinear relationships. Although its cross-validation performance was lower than RF, the optimized XGBoost configuration provided a substantial basis for improving its baseline performance.

The SVR model obtained a cross-validation R^2 of 0.852174 using C=1000, ϵ =1×10-6 and γ =0.01. The substantial increase in the regularization parameter C and the use of a small ϵ value indicate that the optimized model was substantially different from an untuned default configuration. This optimization was particularly important for SVR because its baseline model produced a negative test R^2.

The comparison of the cross-validation results shows the performance order of RF > XGBoost > SVR, with corresponding R^2 values of 0.994667, 0.933424, and 0.852174, respectively. The optimization required approximately 741.58 s for RF, 41.93 s for XGBoost, and 11.63 s for SVR. The longer RF optimization time resulted from the larger hyperparameter search involving 1080 model fits, compared with 1440 fits for XGBoost and 900 fits for SVR, although the computational time also depends on the training characteristics and algorithmic implementation. The cross-validation results demonstrate that hyperparameter tuning was particularly important for improving the generalization capability of the models. However, cross-validation performance alone was not used to determine the final predictive superiority of the algorithms.

4.4 Optimized Model Performance

Following hyperparameter optimization, the RF, XGBoost, and SVR models were retrained using the complete training dataset with their respective optimum parameter configurations. Their predictive performance was then assessed using the independent test dataset. The results are presented in Table 3, while the corresponding measured-versus-predicted relationships are illustrated in Fig. 5–Fig. 7.

Table 3. Performance of optimized RF, XGBoost, and SVR models

Model	Dataset	R	R^2	RMSE	MAE	MAPE	SI
RF	Train	0.999812	0.999562	0.000013	0.000005	1.266345	0.035813
RF	Test	0.999628	0.999208	0.000014	0.000006	3.206927	0.053698
XGBoost	Train	0.978001	0.946083	0.000148	0.000081	51.95191	0.397319
XGBoost	Test	0.992982	0.980223	0.000072	0.00005	53.23092	0.268324
SVR	Train	0.96871	0.863494	0.000236	0.0002	214.5607	0.632196
SVR	Test	0.960408	0.788177	0.000235	0.000197	250.1525	0.878133

The optimized RF model demonstrated the highest predictive accuracy, achieving a test R of 0.999628 and R^2 of 0.999208. The model produced the lowest RMSE (0.000014 mm/year), MAE (0.000006 mm/year), and SI (0.053698) among the three models. Its MAPE was 3.21%, indicating a close agreement between the predicted and experimentally measured corrosion rates. The similarity between the training and testing R^2 values (0.999562 and 0.999208, respectively) further indicates consistent predictive performance.

The measured-versus-predicted relationship for RF is presented in Fig. 5. The predictions are closely distributed around the ideal 1:1 line for both training and testing observations, with only minor deviations.

This graphical agreement corresponds well with the very low RMSE, MAE, and SI values and confirms the superior predictive capability of RF for the present dataset.

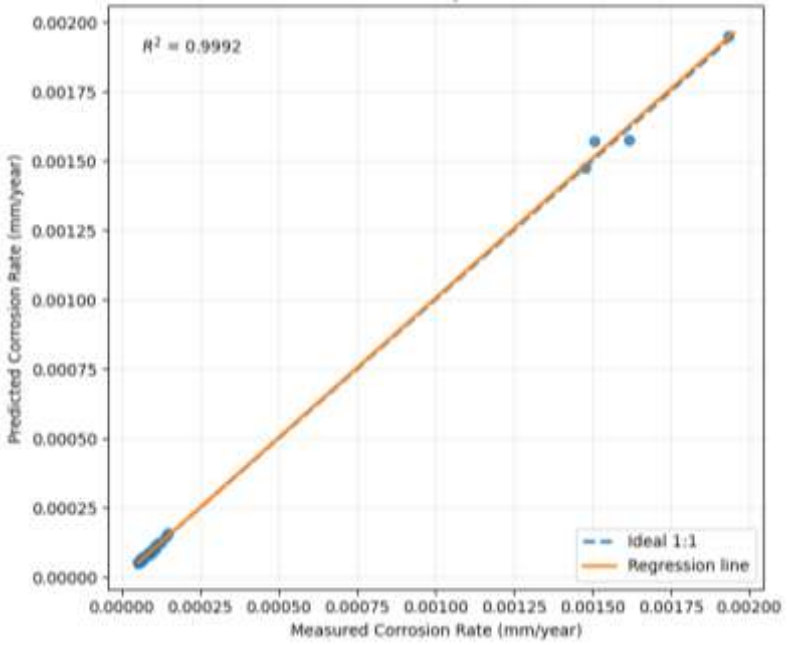

Fig. 5. Measured versus predicted corrosion rate for the optimized Random Forest model

The optimized XGBoost model also demonstrated strong predictive performance, obtaining a test R of 0.992982 and R^2 of 0.980223. The corresponding RMSE, MAE, and SI values were 0.000072 mm/year, 0.000050 mm/year, and 0.268324, respectively. Its MAPE was 53.23%. Although the XGBoost predictions showed good agreement with the measured values, the larger RMSE, MAE, and SI compared with RF indicate greater prediction dispersion. The improvement in test R^2 from 0.975748 for the baseline model to 0.980223 after optimization demonstrates the positive effect of hyperparameter tuning.

The measured-versus-predicted results for XGBoost are shown in Fig. 6. The predicted values generally follow the ideal 1:1 relationship, although the scatter is greater than that observed for RF. This observation is consistent with the comparatively higher error values obtained for XGBoost.

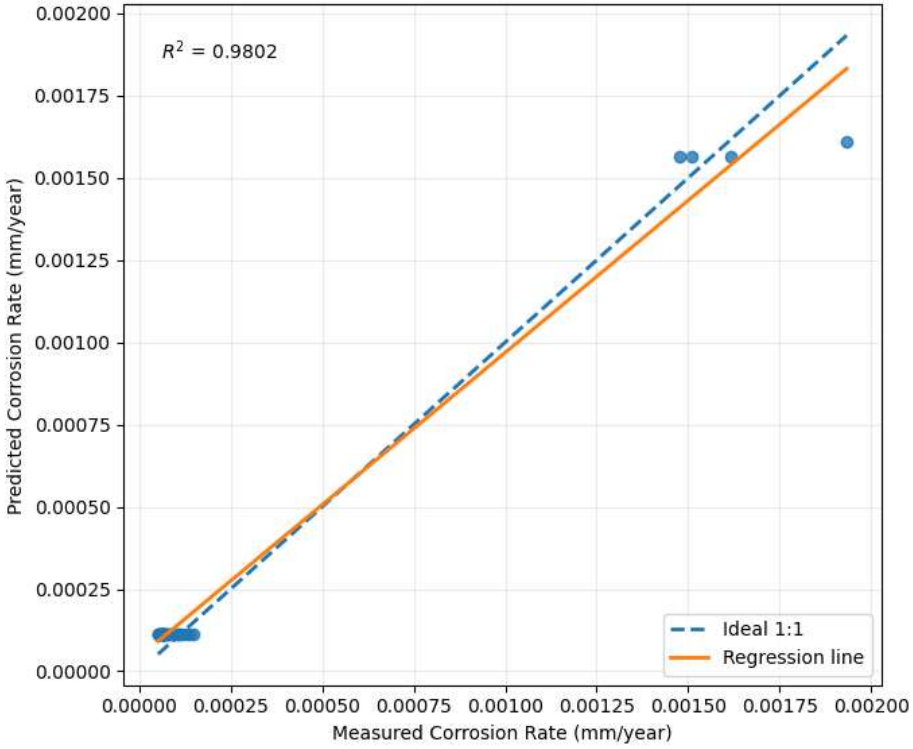

Fig. 6. Measured versus predicted corrosion rate for the optimized XGBoost model

The optimized SVR model showed substantial improvement over its baseline configuration. Its test R^2 increased from −3.490352 to 0.788177 following hyperparameter optimization. The optimized model

achieved a test of 0.960408, with RMSE, MAE, and SI values of 0.000235 mm/year, 0.000197 mm/year, and 0.878133, respectively. However, its MAPE remained relatively high at 250.15%, and the larger error values indicate lower predictive accuracy than RF and XGBoost.

The measured-versus-predicted relationship for SVR is presented in Fig. 7. Greater deviation from the ideal 1:1 line is observed compared with the two tree-based models, which agrees with its comparatively higher RMSE, MAE, MAPE, and SI values.

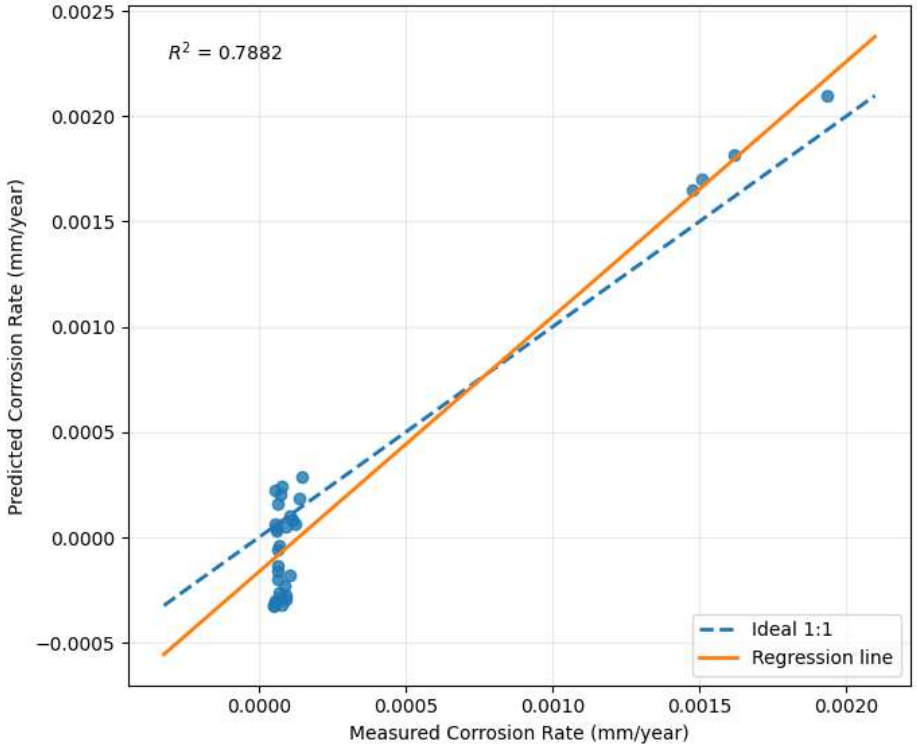

Fig. 7. Measured versus predicted corrosion rate for the optimized Support Vector Regression model

4.5 Feature Importance Analysis

Feature importance and SHAP analyses were performed to interpret the contribution of the three input variables—NaCl concentration, inhibitor dosage (DOI), and exposure duration (T)—to corrosion-rate prediction. The results obtained from RF, XGBoost, and SVR are presented in Table 4 and illustrated in Figs. 8–10.

Table 4. Feature importance obtained from RF, XGBoost, and SVR models

Model	Feature	Importance
RF	DOI (%)	95.04%
	T (days)	2.93%
	NaCl (%)	2.02%
XGBoost	DOI (%)	81.69%
	T (days)	9.55%
	NaCl (%)	8.76%
SVR	DOI (%)	2.732466
	T (days)	0.101111
	NaCl (%)	0.007771

The RF feature-importance analysis identified inhibitor dosage (DOI) as the dominant predictor, contributing 95.04% of the total feature importance. Exposure duration and NaCl concentration contributed only 2.93% and 2.02%, respectively. This pronounced dominance indicates that changes in inhibitor dosage were associated with the largest variation in the corrosion-rate predictions generated by the RF model.

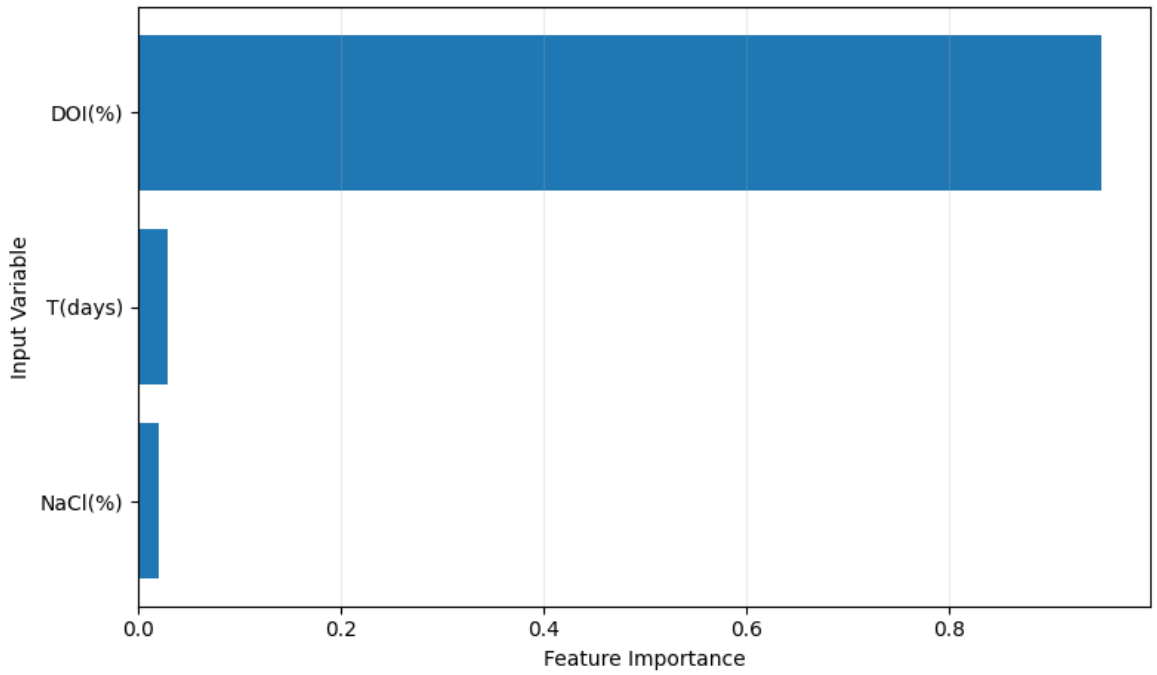

Fig. 8. Feature importance of NaCl concentration, inhibitor dosage, and exposure duration obtained from the RF model.

The XGBoost model produced a similar feature ranking. DOI accounted for 81.69% of the feature importance, followed by exposure duration (9.55%) and NaCl concentration (8.76%). Although DOI remained the most influential variable, XGBoost assigned comparatively greater contributions to exposure duration and NaCl concentration than RF. The consistency in the ranking of DOI across both tree-based models indicates that inhibitor dosage was the principal predictor within the investigated dataset.

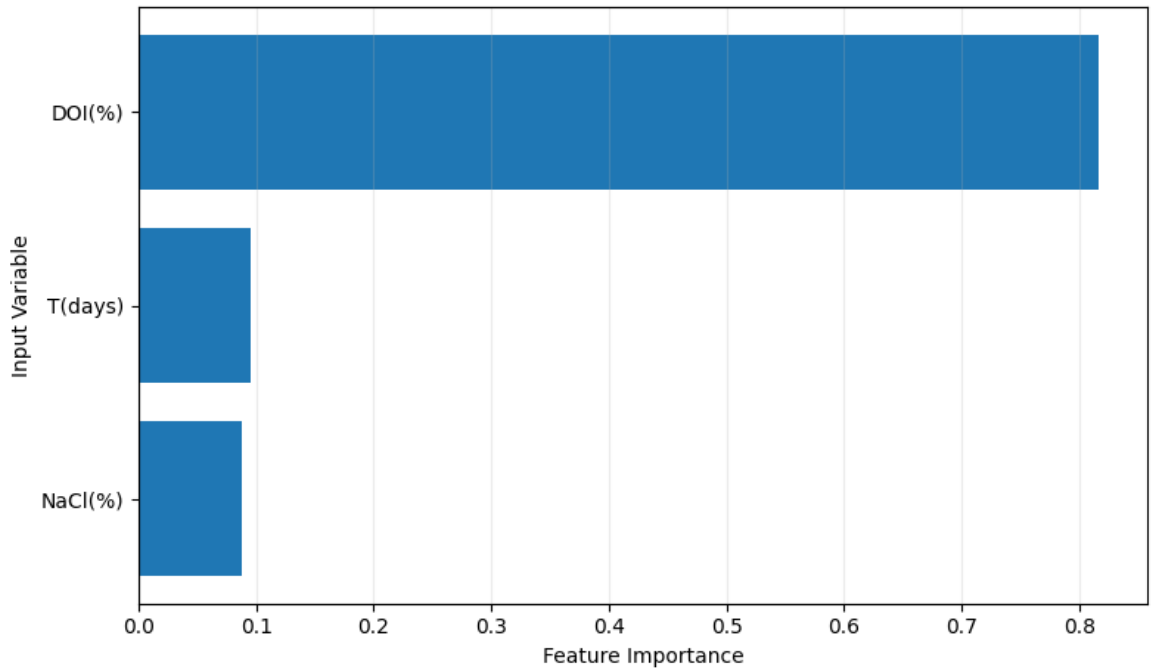

Fig. 9. Feature importance of NaCl concentration, inhibitor dosage, and exposure duration obtained from the XGBoost model.

For SVR, permutation importance also identified DOI as the most influential variable, with a mean importance of 2.732466, compared with 0.101111 for exposure duration and 0.007771 for NaCl concentration. Thus, despite differences in the magnitude of importance values arising from the different interpretation methods, all three models identified DOI as the most influential input variable.

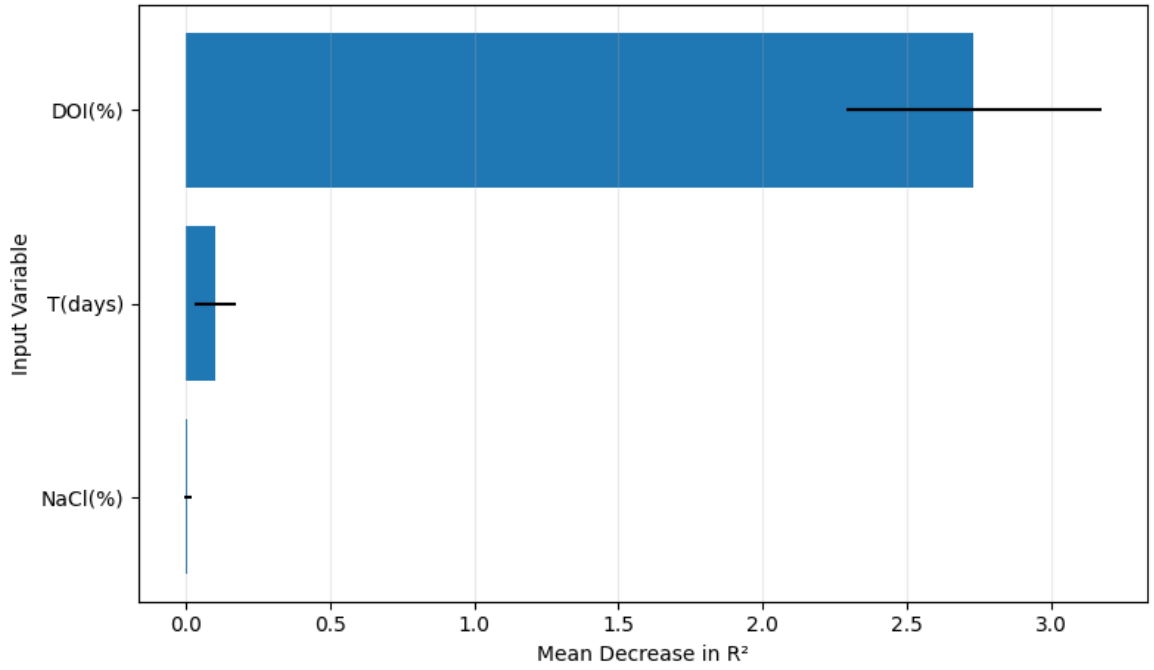

Fig. 10. Permutation importance of NaCl concentration, inhibitor dosage, and exposure duration obtained from the SVR model.

4.6 SHAP-Based Explainability

SHAP analysis was additionally performed for the RF and XGBoost models to provide a model-independent interpretation of the contribution of individual input variables. The mean absolute SHAP values are presented in Table 5. SHAP feature-importance analysis for RF and XGBoost model is shown in Fig. 11 and Fig. 12 respectively.

Table 5. Mean absolute SHAP importance for optimized RF and XGBoost

Feature	RF Mean SHAP	XGBoost Mean SHAP
DOI (%)	0.000412	0.000378
T (days)	0.000043	0.000004
NaCl (%)	0.000025	0.000004

The SHAP analysis provided results consistent with the feature-importance analysis. For RF, DOI exhibited the highest mean absolute SHAP value (0.000412), followed by exposure duration (0.000043) and NaCl concentration (0.000025). Similarly, XGBoost assigned the highest SHAP contribution to DOI (0.000378), whereas both exposure duration and NaCl concentration showed considerably smaller values (0.000004 each).

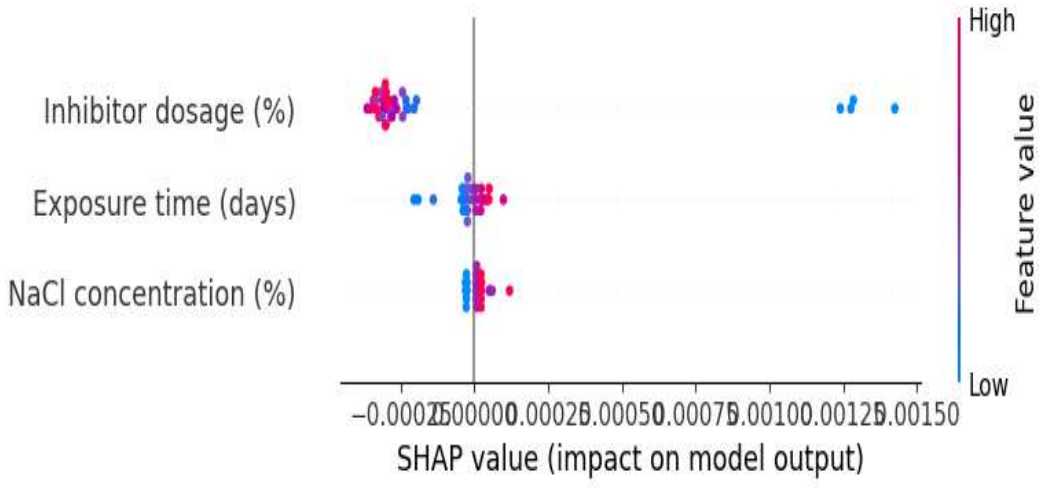

Fig. 11. SHAP feature-importance analysis for the RF model.

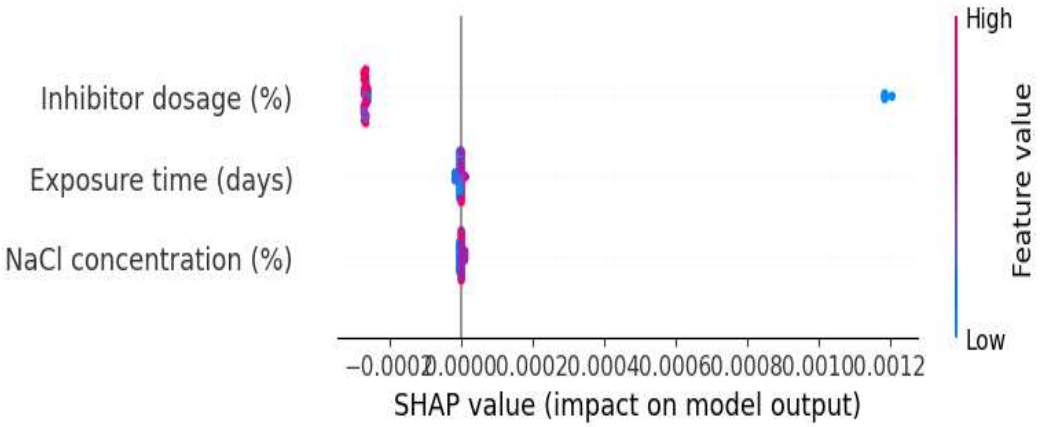

Fig. 12. SHAP feature-importance analysis for the XGBoost model.

The agreement between the conventional feature-importance and SHAP analyses strengthens the interpretation that inhibitor dosage was the dominant predictive variable for corrosion rate in the developed models. The consistent ranking across different modelling approaches suggests that the observed predictive behaviour was not solely dependent on the internal structure of a particular algorithm. The relatively smaller contributions of NaCl concentration and exposure duration indicate that, within the ranges represented in the dataset, variations in inhibitor dosage had a stronger influence on the modelled corrosion-rate response.

5. Conclusions

The present study investigated the applicability of machine-learning techniques for predicting the corrosion rate of reinforced concrete incorporating corrosion inhibitors under different chloride concentrations and exposure durations. A dataset comprising 162 observations was developed using NaCl concentration, inhibitor dosage, and exposure duration as input variables, while the experimentally measured corrosion rate was considered the target output. Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) were developed and systematically optimized using five-fold cross-validation. The conclusions are as follows:

i. RF demonstrated the highest predictive accuracy. The optimized RF model achieved a test of 0.999628 and of 0.999208, with RMSE, MAE, MAPE, and SI values of 0.000014 mm/year, 0.000006 mm/year, 3.21%, and 0.053698, respectively.
ii. XGBoost also provided strong predictive performance, achieving a test of 0.992982 and of 0.980223. Its RMSE, MAE, MAPE, and SI were 0.000072 mm/year, 0.000050 mm/year, 53.23%, and 0.268324, respectively.
iii. SVR showed substantial improvement following optimization. Its test increased from −3.490352 for the baseline model to 0.788177 for the optimized model. Nevertheless, its performance remained lower than that of RF and XGBoost, with a test of 0.960408, RMSE of 0.000235 mm/year, MAE of 0.000197 mm/year, MAPE of 250.15%, and SI of 0.878133.
iv. Hyperparameter optimization was particularly beneficial for SVR, while RF required only a marginal improvement because its baseline performance was already exceptionally high. The final predictive ranking was RF > XGBoost > SVR, which was consistent with the cross-validation results.
v. (v)Inhibitor dosage was identified as the dominant predictive variable. RF assigned an importance of 95.04% to inhibitor dosage, while XGBoost assigned 81.69%. SVR permutation importance and SHAP analysis also identified inhibitor dosage as the most influential input.
vi. SHAP analysis further confirmed the consistency of the feature-importance results. The mean absolute SHAP values for inhibitor dosage were 0.000412 for RF and 0.000378 for XGBoost, substantially higher than those associated with NaCl concentration and exposure duration.
vii. The measured-versus-predicted analyses demonstrated that the RF predictions were concentrated most closely around the ideal 1:1 relationship, supporting its superior statistical performance and generalization capability.

Declarations

Availability of Data and Materials

The data generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

The authors received no specific financial support or funding for the research, authorship, or publication of this article.

Authors' Contributions

Ranveer Sahu: Experimental investigation, data collection, machine-learning modelling, data analysis, visualization, and manuscript preparation. Md Daniyal: Conceptualization, research supervision, methodology, interpretation of results. Rahat Yezdani: Critical review, and revision of the manuscript.

Acknowledgements

The authors gratefully acknowledge the Department of Civil Engineering, Sandip University, Madhubani, India, for providing the necessary academic and laboratory facilities to conduct this research.

References

1. Angst, U. M. (2018). Challenges and opportunities in corrosion of steel in concrete. Materials and Structures 2018 51:1, 51(1), 4-. https://doi.org/10.1617/S11527-017-1131-6
2. Ann, K. Y., Jung, H. S., Kim, H. S., Kim, S. S., & Moon, H. Y. (2006). Effect of calcium nitrite-based corrosion inhibitor in preventing corrosion of embedded steel in concrete. Cement and Concrete Research, 36(3), 530–535. https://doi.org/10.1016/J.CEMCONRES.2005.09.003
3. Balonis, M., & Glasser, F. P. (2011). Calcium nitrite corrosion inhibitor in portland cement: Influence of nitrite on chloride binding and mineralogy. Journal of the American Ceramic Society, 94(7), 2230–2241. https://doi.org/10.1111/J.1551-2916.2010.04362.X
4. Bertolini, L., Elsener, B., Pedeferri, P., Redaelli, E., & Polder, R. (2013). Corrosion of Steel in Concrete: Prevention, Diagnosis, Repair. Corrosion of Steel in Concrete: Prevention, Diagnosis, Repair, 1–414. https://doi.org/10.1002/9783527651696
5. Casanova, L., Ceriani, F., Messinese, E., Paterlini, L., Beretta, S., Bolzoni, F. M., Brenna, A., Diamanti, M. V., Ormellese, M., & Pedeferri, M. P. (2023). Recent Advances in the Use of Green Corrosion Inhibitors to Prevent Chloride-Induced Corrosion in Reinforced Concrete. Materials 2023, Vol. 16, Page 7462, 16(23), 7462. https://doi.org/10.3390/MA16237462
6. Dhouibi, L., Triki, E., & Raharinaivo, A. (2002). The application of electrochemical impedance spectroscopy to determine the long-term effectiveness of corrosion inhibitors for steel in concrete. Cement and Concrete Composites, 24(1), 35–43. https://doi.org/10.1016/S0958-9465(01)00062-2
7. Gao, P., Song, Y., Wang, J., Yang, Z., Wang, K., & Yuan, Y. (2024). Prediction Model for the Chloride Ion Permeability Resistance of Recycled Aggregate Concrete Based on Machine Learning. Buildings 2024, Vol. 14, Page 3608, 14(11), 3608. https://doi.org/10.3390/BUILDINGS14113608
8. Hosseinzadeh, M., Mousavi, S. S., Hosseinzadeh, A., & Dehestani, M. (2023). An efficient machine learning approach for predicting concrete chloride resistance using a comprehensive dataset. Scientific Reports 2023 13:1, 13(1), 15024-. https://doi.org/10.1038/s41598-023-42270-3
9. Hu, J., Zhu, Y., Hang, J., Zhang, Z., Ma, Y., Huang, H., Yu, Q., & Wei, J. (2021). The effect of organic core–shell corrosion inhibitors on corrosion performance of the reinforcement in simulated concrete pore solution. Construction and Building Materials, 267. https://doi.org/10.1016/j.conbuildmat.2020.121011
10. Huang, J., Hu, J., Cai, J., Huang, H., Wei, J., & Yu, Q. (2022). Inhibition Effect of Hydrophobic Functional Organic Corrosion Inhibitor in Reinforced Concrete. Materials 2022, Vol. 15, Page 7124, 15(20), 7124. https://doi.org/10.3390/MA15207124
11. Huang, T., Wan, C., Liu, T., Ai, Y., & Miao, C. (2025). Data-driven corrosion level prediction of rust-damaged reinforced concrete: An interpretable machine learning approach. Materials Today Communications, 48, 113400. https://doi.org/10.1016/J.MTCOMM.2025.113400
12. Jinsong, L., Qian, G., Jing, L., & Lin, F. (2022). Influence of Calcium Nitrite on Electrochemical Corrosion Behavior of API X120 Carbon Steel and 316LN Stainless Steel Reinforced Concrete in Marine Environment. International Journal of Electrochemical Science, 17(2), 22027. https://doi.org/10.20964/2022.02.35
13. Leporace-Guimil, B., Russo, N., Lollini, F., & Conforti, A. (2023). Morphological and mechanical characterization of reinforcement in cracked elements exposed to chloride-induced corrosion. Construction and Building Materials, 364. https://doi.org/10.1016/J.CONBUILDMAT.2022.129822
14. Li, L., Su, L., Guo, B., Cai, R., Wang, X., & Zhang, T. (2024). Prediction and prevention of concrete chloride penetration: machine learning and MICP techniques. Frontiers in Materials, 11, 1445547. https://doi.org/10.3389/FMATS.2024.1445547/TEXT
15. Liu, J., Hang, M., Jiang, M., Song, H., & Zhou, X. (2024). Performance and mechanism of triethanolamine-triisopropanolamine corrosion inhibitor to deterioration of reinforced concrete under the coupling effect of freeze-thaw cycles and chloride attack. Journal of Building Engineering, 98, 111066. https://doi.org/10.1016/J.JOBE.2024.111066
16. Liu, W., Liu, G., & Zhu, X. (2024). Applicability of machine learning algorithms in predicting chloride diffusion in concrete: Modeling, evaluation, and feature analysis. Case Studies in Construction

Materials, 21, e03573. https://doi.org/10.1016/J.CSCM.2024.E03573
17. Matthews, B., Palermo, A., Logan, T., & Scott, A. (2024). Database and optimized machine learning prediction of the deteriorated response of corroded reinforced concrete beams. Developments in the Built Environment, 19, 100527. https://doi.org/10.1016/J.DIBE.2024.100527
18. Ormellese, M., Lazzari, L., Goidanich, S., Fumagalli, G., & Brenna, A. (2009). A study of organic substances as inhibitors for chloride-induced corrosion in concrete. Corrosion Science, 51(12), 2959–2968. https://doi.org/10.1016/j.corsci.2009.08.018
19. Selvaprasanth, P., Malathy, R., & Priyadharshini, R. (2025). Machine learning based corrosion prediction in reinforced concrete structures exposed to marine environment. Journal of Structural Engineering, 52(2), 89–112. https://jose.serc.res.in/index.php/jose/article/view/2699
20. Shen, L., Jiang, H., Cao, J., Zhang, M., & Zhang, H. (2020). A comparison study of the performance of three electro-migrating corrosion inhibitors in improving the concrete durability and rehabilitating decayed reinforced concrete. Construction and Building Materials, 238. https://doi.org/10.1016/j.conbuildmat.2019.117673
21. Sun, C., Sun, M., Liu, J., Dong, Z., Fan, L., & Duan, J. (2022). Anti-Corrosion Performance of Migratory Corrosion Inhibitors on Reinforced Concrete Exposed to Varying Degrees of Chloride Erosion. Materials 2022, Vol. 15, Page 5138, 15(15), 5138. https://doi.org/10.3390/MA15155138
22. Tesic, K., Baricevic, A., Serdar, M., & Gucunski, N. (2023). Quantifying the impact of parameters of chloride-induced reinforcement corrosion on the GPR signal. Construction and Building Materials, 399, 132594. https://doi.org/10.1016/J.CONBUILDMAT.2023.132594
23. Tiwari, A. K., Dogra, P., Goyal, S., & Luxami, V. (2025). Evaluation of Different Migratory Corrosion Inhibitors as Repair Strategy for Reinforced Concrete. Materials Journal, 122(4), 29–44. https://doi.org/10.14359/51746804
24. Wang, H., Jiang, M., Hang, M., Yang, Y., Zhou, X., Liu, X., & Xu, G. (2023). Inhibition resistance and mechanism of migrating corrosion inhibitor on reinforced concrete under coupled carbonation and chloride attack. Journal of Building Engineering, 76, 107398. https://doi.org/10.1016/J.JOBE.2023.107398
25. Xu, C., Jin, W. L., Wang, H. L., Wu, H. T., Huang, N., Li, Z. Y., & Mao, J. H. (2016). Organic corrosion inhibitor of triethylenetetramine into chloride contamination concrete by eletro-injection method. Construction and Building Materials, 115, 602–617. https://doi.org/10.1016/j.conbuildmat.2016.04.076
26. Zheng, W., & Cai, J. (2024). A optimum prediction model of chloride ion diffusion coefficient of machine-made sand concrete based on different machine learning methods. Construction and Building Materials, 411, 134414. https://doi.org/10.1016/J.CONBUILDMAT.2023.134414