OPTIMIZED PREDICTION MODELING FOR CHLORINATED MARINE CONCRETE USING DECISION TREE

[a] Muhammad Luqman* and [b] Majid Khan

a: Barrick Gold Corporation, Pakistan.18pwciv5026@uetpeshawar.edu.pk
b: Department of Civil Engineering, COMSATS University, Abbottabad, Pakistan. 18pwciv4988@uetpeshawar.edu.pk
* Corresponding author

Abstract- Chloride concentration (Cs) at the surface of concrete is an essential metric for designing resilience and estimating the lifespan of concrete structures in aquatic settings. Consequently, due to chlorine action, many reinforced concrete constructions cannot reach their intended or planned lifespan and experience early degradation. This study utilizes the independent machine learning technique Decision Tree (DT) to forecast concrete's surface chloride concentration (Cs). A comprehensive database consisting of 642 observations of Cs exposure data in the marine field, including the applicable mixture quantity of constraints, conditions of the environment, and exposure time, has been created through a thorough investigation of relevant literature. Diverse statistical criteria evaluated the model's accuracy and suitability. During the validation process, the DT model demonstrated enhanced accuracy with correlation coefficients (R) of 0.95 for training and 0.96 for validation and mean absolute errors (MAE) of 0.009. The results indicate that by including more diverse datasets and considering new variables, the predicted accuracy of standard models may be improved. The DT machine learning model, trained on a vast database, can effectively include 13 key characteristics that pose challenges for conventional models. However, to lessen the problem of overfitting, it is advisable to use a more extensive dataset, including synthetic or genuine experimental data.

Keywords- Regression, K-Fold Method, Semi exponential Model, Particle Swarm Optimizer (PSO)

1 Introduction

The erection of sea-crossing bridges, dock piers, coastal highways, and buildings is among the many applications of reinforced concrete (RC) structures in marine and coastal environments. Broadly, a passive oxide deposit can protect reinforcing steel in RC structures from rusting, as it is resilient in the high pH micro-environment that results from the concrete pore solution [1], [2]. Chloride ions present in saltwater and airborne aerosols may adhere to and infiltrate the exterior of reinforced concrete structures in coastal and aquatic settings. Suppose the chloride ions infiltrate and capture inside the matrix of concrete encasing the steel reinforcement. In that case, they have the potential to destabilize the protective layer and commence and expedite corrosion of the steel, resulting in the formation of cracks and spalling of the concrete, as well as a reduction of the load-bearing capacity of reinforced concrete structures [3]. Researchers recently developed mathematical frameworks in laboratory settings that generate chloride profiles from the underlying structure of wet concrete [4]. In 2023, Reichert et al. [5] proposed a semiempirical double exponential model designed explicitly for the penetration of chlorine in diffusion and convection zones, which does not encompass different practical conditions.

Corrosion in RC structures may cause crashes, safety hazards, and significant economic losses, affecting the normal functioning of the engineered environment. Consequently, many reinforced concrete constructions cannot reach their intended or planned lifespan and experience early degradation [6]. Hence, the problem of designing concrete buildings with longevity or accurately predicting their lifespan has become a significant undertaking in current design practices. This effort requires using solid models that can effectively include several influencing elements and provide a more precise description of the processes responsible for degradation [7]. Marine ecosystems are often categorized into four zones based

on the types of chloride exposure they experience: atmospheric, tidal, splash, and submerged zones. Fick's second law of diffusion describes chloride infiltration into concrete regardless of the zone [8]. Eq (1) provides the mathematical solution for Fick's second law, often used in maritime service life design of RC structures.

$$C(x,T) = C_S.\,ERF * \left(\frac{x}{2 * (D_e * T)^{0.5}}\right) \quad (1)$$

The function *C(x, T)* represents the concentration of chloride ions, [Cl-], at a certain distance *x* into the concrete, measured from the exposed surface, after a certain amount of time *T*. The term D_e refers to the effective diffusion coefficient. If Eq (1) is written with the assumption that De is independent of *x* and T, then the concentration of chloride ions ([Cl-]) at the surface of the concrete is supposed to be constant throughout time. The Gaussian error function (ERF) is used in this context. Additionally, at the onset of corrosion, *C(x, T)* is $C_{T,}$ and *x* represents the thickness of the rebar cover. Unlike D_e, *Cs* is a multidimensional characteristic that includes material characteristics, time, and external factors. Additional study is required to enhance the accuracy of predicting the advancement of Cs and chloride influx.

The work aims to predict Cs, accounting for all critical parameters that previous research lacks, to help design durable RC structures in maritime conditions. Using nonlinear independent variables, ML has been employed in Material Engineering and corrosion prediction [9], [10]. In order to address the shortcomings of traditional approaches, this study employs ML models (Supervised) to construct an accurate and practical model for concrete chloride intrusion prediction. This ML model will solve the material's complex composition degrees of flexibility and the multivariate link between the mixture in dependent variables and attributes.

2 Research Methodology

2.1 Theoretical Background of Decision Tree

Decision trees are adequate for both regression and classification tasks in machine learning. Resembling flowcharts, decision trees can make repeated practical decisions and handle unstructured information for predictions. One challenge in machine learning is communicating model results, but decision trees make this more straightforward due to their clear decision paths [11]. A decision tree starts with a root node and branches out, with each node representing a decision based on a specific attribute and each leaf node representing an outcome. Decision nodes lead to further branches, while leaf nodes provide the final decision. The k-fold cross-validation (CV) method, specifically the 12-fold CV, helps find the best hyper-parameters. The training data is split into 12 equal parts. Nine folds (482 observations) train the model, while three folds (160 observations) are used for testing. This process repeats nine times with different fold combinations to train and test the model. CV errors are calculated at each cycle, and model parameters are adjusted accordingly [12].

2.2 Data Collection

To develop reliable machine learning models to forecast the Cs (Surface Chloride Concentration as Wt. of Concrete) relationship to changes in concrete mix designs, environmental variables, and exposure duration, 642 experimental and fitted analytical data points were gathered from recently published research. The developed model utilizes evident exterior chloride concentration in marine concrete from field measurements on RC structures exposed to sea splash conditions tidal, and submerged zones. The mix design of concrete has ten variables. The environmental settings are also characterized by T (mean annual temperature, °C) and Cl^{-} (seawater chloride concentration, g/L). Exposure time (ETM) is a single variable representing the duration of exposure. A supplementary input was included to account for the environmental conditions exposed to the concrete. This input was assigned values of 0 for the tide zone, 1 for the splash zone, and 2 for the submerged zone. All these data attributes are concise in Table 1. This demonstrates a minimal likelihood of multicollinearity, as the correlation values are lower than the indicated threshold. Table 1 presents the dataset's statistical characteristics, including indicators such as the mean, standard deviation, skewness, and kurtosis. The variables demonstrate skewness and kurtosis values within the prescribed ranges of ±3 and ±10, respectively. Moreover, it is crucial to note that CSB has the strongest positive correlation (+0.97) with Cs compared to all other input parameters (Cs). The following variable is the ET, with a correlation value of +0.50. We have the FA and W, with a correlation coefficient of + 0.38 and + 0.48, respectively. Lastly, the exposure time has the lowest correlation coefficient of -0.01.

2.3 Model Development

Thoroughly analyzing various DT setup configurations is essential for creating a robust model in the field of ML modeling. These generated parameters are recommended based on several trial runs since they consider the essential population size

parameter directly affecting the number of programs produced. While the convergence process may be time-consuming, a more intricate and precise model benefits from a larger population. Overfitting may occur when the size of the population surpasses a particular range. The fundamental operations in the paradigm are addition, division, multiplication, and subtraction. Prior to the completion of the method, the generation numeral establishes the desired degree of precision for the model. Multiple iterations of algorithms have contributed to developing a simulation model with minimal imperfections. The present technique experimented with several combinations of parameters to identify the optimal model and successfully created one with the lowest error levels. An inherent challenge in machine learning modeling is overfitting, which occurs when the model demonstrates high accuracy on the original dataset but fails to generalize effectively to new, concealed data. To address this issue and get insight into the model's ability to generalize, evaluating its performance on data that has not been previously examined is recommended. In order to mitigate the issue of overfitting, the data is divided into two groups, allocating 75% for training purposes and 25% for validation. The algorithm's performance is evaluated using an independent validation set that was not used during the construction of the model.

Table 1 Statistical description of input variables in the database for CS

Parameter	Unit	Mean	SD[1]	Min	Max	Kurtosis	SKS[2]	OPN[3]
OPC	Kg/m^3	370.70	75.60	110.00	519.00	0.60	-0.65	Input
FAH	Kg/m^3	33.97	59.88	0.00	239.00	3.46	1.99	Input
GGBS	Kg/m^3	11.41	45.45	0.00	292.50	15.49	4.03	Input
SF	Kg/m^3	5.41	12.85	0.00	50.00	4.11	2.29	Input
S	Kg/m^3	1.47	1.99	0.00	10.20	4.67	2.09	Input
W	Kg/m^3	187.54	44.08	38.50	311.00	2.80	0.82	Input
FA	Kg/m^3	765.77	116.46	552.00	1232.00	4.74	1.46	Input
CA	Kg/m^3	993.81	135.44	410.00	1305.00	7.56	-1.85	Input
w/b	-	0.46	0.08	0.30	0.75	1.04	0.96	Input
ETM	Years	4.24	6.28	0.08	48.65	24.71	4.45	Input
T	°C	17.78	9.38	7.00	50.00	-0.85	0.42	Input
CSB	%	3.67	2.09	0.14	13.58	1.89	1.02	Input
Cl^{--}	g/L	18.99	2.79	13.00	27.37	0.72	0.52	Input
ET	0,1,2	N/A	N/A	N/A	N/A	N/A	N/A	Input
C_s	%	0.66	0.39	0.02	1.95	-0.18	0.68	Output

[1]Standard Deviation, [2]Skewness, [3]Operation

3 Results - Model Performance

The correctness of the created model is assessed by examining the gradient of the regression line derived from concrete values shown on the x-axis against projected values on the y-axis, as shown in Fig. 1

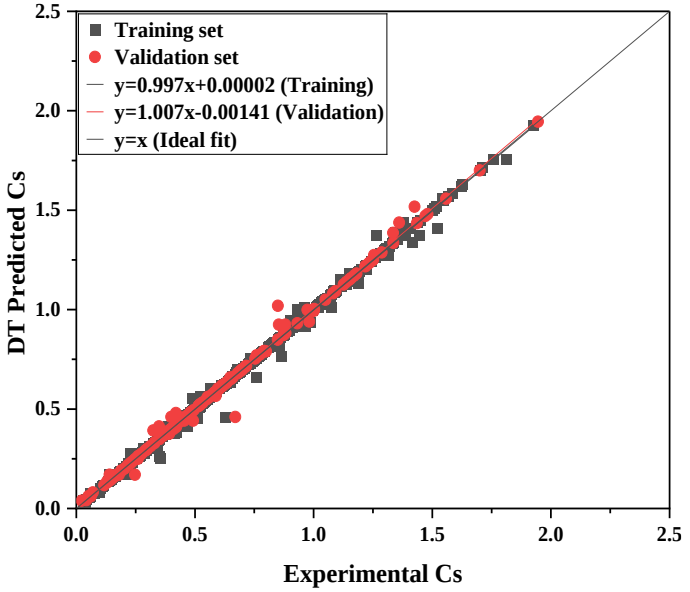

Fig. 1 Regression Slope Analysis of Model

Table 2 Performance Assessment of the ML Models

Subset	RMSE	MAE	R	a_{10} Index	a_{20} Index
Training	0.020	0.007	0.95	0.120	0.230
Validation	0.028	0.009	0.96	0.081	0.144

Researchers often use this methodology to evaluate the precision of machine learning models [13], [14], [15]. Predictions were then compared for both training and testing datasets to assess the predictive capabilities of the models. Fig. 1 and Table 2 display the statistical assessment of the model with five statistical parameters and their corresponding findings for this specific objective. The DT model exhibits slopes of 0.997 and 1.007 for the training and validation datasets. Moreover, the DT model showed R-values of 0.95 for training and 0.96 for validation. The MAE values for the DT model are shallow: 0.007 for training and 0.009 for validation. Predictions were then compared for both training and testing datasets to assess the predictive capabilities of the models.

4 Conclusions

This study employs a machine learning technique, the Decision Tree (DT), to forecast concrete's surface chloride concentration (Cs). A comprehensive database of 642 observations of Cs exposure data in the marine field—including mixture quantities, environmental conditions, and exposure times—was compiled from relevant literature. The model's accuracy and suitability were evaluated using diverse statistical criteria. The DT model demonstrated high accuracy during validation, with correlation coefficients (R) of 0.95 for training and 0.96 for validation and mean absolute errors (MAE) of 0.009. For future studies, hybrid ML models might be established, which may increase predicted accuracy. This study employs Statistical means to comprehend the model, but future investigations may benefit from model-agnostic tools like LIME, PDP, and SHAP. Understanding the study's limitations is crucial. Current literature shows experimental setup discrepancies across research. Further research should focus on controlled experimental testing to increase model robustness and dependability using a single, dependable source in the same context to collect data.

5 References

[1] O. E. Gjørv, Durability design of concrete structures in severe environments, vol. NA, no. NA. 2009. doi: NA.
[2] S. W. Tang, Y. Yao, C. Andrade, and Z. J. Li, "Recent durability studies on concrete structure," Cement and Concrete Research, vol. 78. 2015. doi: 10.1016/j.cemconres.2015.05.021.
[3] C. E. T. R. Balestra Thiago Alessi; Savaris Gustavo, "Contribution for durability studies based on chloride profiles analysis of real marine structures in different marine aggressive zones," Constr Build Mater, vol. 206, no. NA, pp. 140–150, 2019, doi: 10.1016/j.conbuildmat.2019.02.067.
[4] T. A. Reichert, C. E. T. Balestra, D. A. O. Balestra, and R. A. de Medeiros-Junior, "Laboratory procedure for obtaining chloride profiles from concrete structures cores: a mathematical approach," Journal of Building Pathology and Rehabilitation, vol. 8, no. 1, 2023, doi: 10.1007/s41024-023-00286-2.
[5] T. A. Reichert, W. A. Pansera, C. E. T. Balestra, and R. A. Medeiros-Junior, "New semiempirical temporal model to predict chloride profiles considering convection and diffusion zones," Constr Build Mater, vol. 367, 2023, doi: 10.1016/j.conbuildmat.2022.130284.
[6] L. F. C. Yang Rong; Yu Bo, "Investigation of computational model for surface chloride concentration of concrete in marine atmosphere zone," Ocean Engineering, vol. 138, no. NA, pp. 105–111, 2017, doi: 10.1016/j.oceaneng.2017.04.024.
[7] A. F. B. A. Costa Julio, "Chloride penetration into concrete in marine environment-Part II: Prediction of long term chloride penetration," Mater Struct, vol. 32, no. 5, pp. 354–359, 1999, doi: 10.1007/bf02479627.
[8] M. M. Collepardi Aldo; Turriziani Renato, "Penetration of Chloride Ions into Cement Pastes and Concretes," Journal of the American Ceramic Society, vol. 55, no. 10, pp. 534–535, 1972, doi: 10.1111/j.1151-2916.1972.tb13424.x.
[9] J.-S. N. Chou Ngoc-Tri; Chong Wai Kiong, "The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate," Eng Appl Artif Intell, vol. 65, no. NA, pp. 471–483, 2017, doi: 10.1016/j.engappai.2016.09.008.
[10] W. Z. S. Taffese Esko, "Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions," Autom Constr, vol. 77, no. NA, pp. 1–14, 2017, doi: 10.1016/j.autcon.2017.01.016.
[11] H. I. Erdal, "Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction," Eng Appl Artif Intell, vol. 26, no. 7, pp. 1689–1697, Jan. 2013, doi: 10.1016/j.engappai.2013.03.014.
[12] A. Karbassi, B. Mohebi, S. Rezaee, and P. Lestuzzi, "Damage prediction for regular reinforced concrete buildings using the decision tree algorithm," Comput Struct, vol. 130, pp. 46–56, Jan. 2014, doi: 10.1016/j.compstruc.2013.10.006.
[13] M. I. Khan et al., "Effective use of recycled waste PET in cementitious grouts for developing sustainable semi-flexible pavement surfacing using artificial neural network (ANN)," J Clean Prod, vol. 340, p. 130840, Jan. 2022, doi: 10.1016/j.jclepro.2022.130840.
[14] M. Iqbal, Q. Zhao, D. Zhang, F. E. Jalal, and A. Jamal, "Evaluation of tensile strength degradation of GFRP rebars in harsh alkaline conditions using nonlinear genetic-based models," Materials and Structures/Materiaux et Constructions, vol. 54, no. 5, 2021, doi: 10.1617/s11527-021-01783-x.
[15] H. Alabduljabbar, M. Khan, H. H. Awan, S. M. Eldin, R. Alyousef, and A. M. Mohamed, "Predicting ultra-high-performance concrete compressive strength using gene expression programming method," Case Studies in Construction Materials, vol. 18, p. e02074, Jan. 2023, doi: 10.1016/j.cscm.2023.e02074.