

(RESEARCH ARTICLE)



MACHINE LEARNING PREDICTION OF THE BASELINE COMPRESSIVE STRENGTH AND COMPRESSIVE STRENGTH OF SCM CONCRETE SUBJECTED TO CHLORIDE INGRESSION

Osasere Destiny Ikponmwosa ^{1,*}, Adetola Aderinsoro Falade ² and Saheed Adedokun ²

¹ Department of Structural Engineering, Faculty of Engineering, University of Benin, Nigeria.

² Department of Civil Engineering, Faculty of Engineering, University of Benin, Nigeria.

* Corresponding Author

ORCID Details

Osasere Destiny Ikponmwosa: <https://orcid.org/0009-0000-9727-0928>

Adetola Aderinsoro Falade: <https://orcid.org/0009-0000-5118-1946>

Saheed Adedokun: <https://orcid.org/0009-0001-6610-8538>

Global Journal of Engineering and Technology Advances, 2026, 29(01), 108–120

Publication history: Received on 20 August 2026; revised on 01 October 2026; accepted on 03 October 2026

Article DOI: <https://doi.org/10.30574/gjeta.2026.29.1.0278>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

This study explores the potential advantages of using multiple machine learning (ML) models to predict the baseline compressive strength (CS) and CS of SCM concrete subjected to chloride ingress. Four ML models, including Random Forest (RF), Gradient Boosting (GradBoost), Extreme Gradient Boosting (XgBoost), and Adaptive Boosting (AdaBoost) algorithms were employed to construct forward prediction models for predicting the baseline CS and CS of SCM concrete with combinations of fly-ash (FA), silica fume (SF), and metakaolin subjected to chloride ingress. The performance of the models was compared and assessed using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE). The performance of the model was further validated using the k-fold cross-validation approach. Compared to the other employed approaches, the GradBoost model was more effective in predicting results, with an R^2 value of 0.8612. To further enhance model interpretability, Shapley Additive Explanation (SHAP) was employed to assess the relative importances of input variables. The use of ML techniques to predict the mechanical properties of SCM concrete subjected to chloride ingress will be beneficial to the field of civil engineering because it will save time, effort, and resources compared to carrying out the laboratory experiments. The proposed techniques pave the way towards the intelligent design of concrete elements and structures.

Keywords: Machine Learning, Sustainable Concrete, Supplementary Cementitious Materials, Compressive Strength, Chloride Ingression

1 INTRODUCTION

It is evident that concrete is the one of the most used construction materials around the world, However, manufacturing one ton of Ordinary Portland Cement (OPC), which is essential component of concrete, contributes significantly to environmental degradation, releases approximately one ton of CO₂ [1]. In recent times, partially replacing OPC with

supplementary cementitious materials (SCM) not only reduce material cost and energy consumption, but it is one of the most promising methods to reduce carbon emissions [2]. The most popular SCMs for the manufacturing of sustainable concrete are industrial waste, such as fly ash (FA), and silica fume (SF) in which both chemical and physical properties of SCMs can vary significantly [3]. Another notable SCM is Metakaolin which is formed from the calcination of kaolin clay transforms, resulting in the elimination of water ions present in the kaolinite clay and the formation of 40% to 45% alumina and 50% to 55% amorphous silica, the two primary chemical components present in pozzolans [4]. Concrete is a porous material that allows ions to pass through its pore solution to balance concentration differences, including chloride ions on the surface, resulting in chloride ingress. Because of their chloride permeability, reinforced concrete structures are vulnerable to deterioration in harsh settings, which can lead to the structure's failure [5]. Corrosion is one of the most common deteriorations in reinforced concrete caused by chloride ingress [6], which reduces the cross-sectional area of steel bars, lowering the load-carrying capacity of structures [7] and jeopardizing structural safety, serviceability, and durability [8]. Thus, understanding the chloride transport and effect within concrete structures is critical for assuring their long-term durability and service life. According to research, FA-modified concrete has reduced permeability, improving reinforced structure resilience to corrosion and chloride intrusion [9], SF also prolongs the life of reinforced concrete structures by lowering carbonation, sulfate attack, and chloride ion penetration [10]. Concrete containing Metakaolin can improve the resistance to chloride ingress and freeze-thaw damage compared to concrete without SCMs [11].

Several studies have used ML for the prediction of the compressive strength of SCM-Modified OPC Concretes. The use of supervised machine learning methods, such as bagging, adaptive boosting (AdaBoost), gene expression programming, and decision tree (DT) models, to forecast the compressive strength of concrete containing FA and blast furnace slag was examined by [12]. The study showed that ML approaches can help the intelligent design and optimization of concrete structures and offer an effective and resource-efficient method for estimating the strength of SCM-based concrete. A machine learning-based method for forecasting the compressive strength of concrete with binary SCMs at various curing ages was created by [13]. The findings showed that the suggested ML models could reliably forecast compressive strength for a range of SCM combinations and ages. Liu et al., [3] investigated the use of five ML models including artificial neural network (ANN), random forest (RF), extra trees (ET), extreme gradient boosting (XgBoost), and AdaBoost to predict the 28-day compressive strength of concrete containing SCMs, including FA, SF, and GGBS. The study demonstrates the potential of ML, particularly XGB, for accurately predicting the compressive strength of SCM-based concrete while providing interpretable and reliable predictions. ML have also been applied for the investigation of the influence of FA dosage on the chloride resistance of concrete [14], prediction of chloride migration coefficient and penetration resistance of concrete [15]. Li et al., [16] investigated the application of ensemble machine learning models for predicting the chloride migration coefficient (CMC) of concrete as an alternative to conventional laboratory testing. These studies have demonstrated the potential of ML techniques for accurately evaluating concrete durability and predicting chloride transport properties. Studies have also employed ML algorithms to predict concrete strength using models such as ANN, support vector machines (SVM), RF, and AdaBoost. Some studies developed ML models using very small datasets, with sample sizes ranging from 19 to 56 data points [17, 18, 19], while others employed relatively larger datasets containing more than 1,000 data points [20, 21]. These studies demonstrate the potential of ML algorithms to achieve satisfactory prediction performance even when the available datasets are relatively limited. However, the predictive capability of ML models is influenced not only by dataset size but also by data quality, representativeness, input-variable selection, model complexity, and appropriate model validation. In this context, this study utilizes a dataset comprising 54 data points obtained from literatures to develop and evaluate ML models prediction for baseline CS and CS of SCM concrete subjected to chloride ingress.

Extensive experimental study has been conducted to assess the mechanical properties of SCM concrete exposed to chloride ingress. However, casting and testing specimens in the lab takes time, effort, and money. This study explores the potential advantages of using multiple ML models to predict the baseline compressive strength (CS) and CS of SCM concrete subjected to chloride ingress. Conventional experimental approaches for assessing the baseline CS and CS of SCM concrete subjected to chloride ingress results are time-intensive and costly. In this study, a sufficient dataset which included a single and binary SCM-concrete blends were employed. Four ML models, including RF, Gradient Boosting (GradBoost), XgBoost, and AdaBoost were utilized to construct forward prediction models for predicting the baseline CS and CS of SCM concretes subjected to chloride ingress with combinations of FA, SF, and metakaolin. To further enhance model interpretability, an explainable artificial intelligence technique, Shapley Additive Explanation (SHAP) was employed to assess the relative importances of input variables.

2 METHODOLOGY

2.1 Data Description

The data used in this study was obtained from the literature [22, 23]. A total of 12 factors were used as inputs for the models which included water, cement, fine aggregate, coarse aggregate, FA, SF, metakaolin, superplasticizer (SP),

Hydrogen Chloride (HCL), Sodium Chloride (NaCl), exposure age, and total curing age. This study utilized a total of 54 data points for the prediction of the baseline CS and CS of the concrete subjected to chloride ingress. The dataset was randomly divided into 80% training and 20% testing sets, as is frequently advocated in ML studies [24] for small to medium data sizes, with shuffling enabled and a fixed random state used to ensure reproducibility between runs. A descriptive statistical analysis for all input parameters is shown in Table 1. Figure 1. Illustrates the framework of this study, which introduces a systematic ML methodology for predictive analysis.

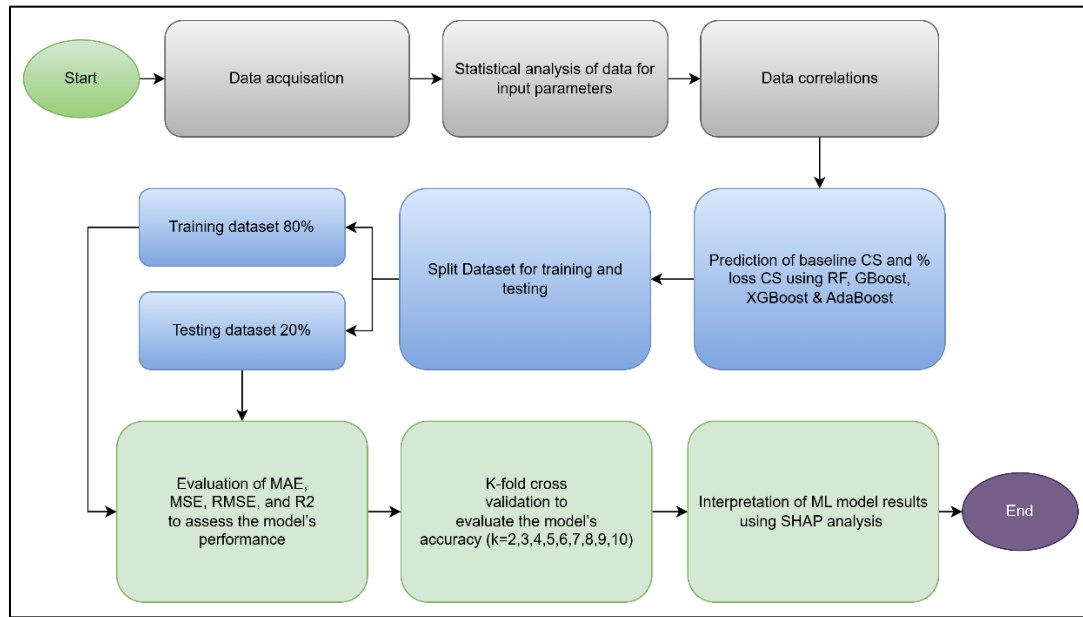


Figure 1: Framework and implementation procedure

2.2 Machine Learning Models

This study exploits four ML algorithms including RF, GradBoost, XgBoost, and AdaBoost to predict the baseline CS and CS of the SCM concrete subjected to chloride ingress. Each algorithm was chosen for its unique capabilities in the predictive analysis of CS of concrete based on previous structural engineering studies.

2.2.1 Random Forest

This machine learning approach is widely utilized in both regression and classification applications, and it has been described as one of the most successful machine learning methods [25]. It is also an effective ensemble learning method [26]. RF creates numerous decision trees from randomly sampled subsets of data and combines their outputs by averaging or majority voting [27]. RF improves predictive accuracy and decreases the risk of overfitting by creating many DT and aggregating their outputs. RF is resistant to outliers and irrelevant characteristics, and it works effectively with high-dimensional, nonlinear data without requiring extensive preprocessing. A simple schematic diagram of the RF algorithm is shown in Figure 2(a).

2.2.2 Gradient Boosting

Boosting algorithms iteratively combine weak learners to produce strong learners. Gradient boosting is a boosting-style regression technique [28]. The purpose of gradient boosting when given a dataset is to obtain an approximation of the function that translates input instances to output values by minimizing the expected value of a specified loss function.

2.2.3 Extreme Gradient Boosting

Extreme gradient boost is an improved version of the gradient boosting system. XgBoost, like gradient boosting, generates an additive extension of the objective function by minimizing a loss function. It enhances prediction accuracy by training decision trees consecutively, with each successive tree addressing the errors made by the previous one, as seen in Figure 2(e). XgBoost features some improvements that increase efficiency, and its regularizing mechanisms minimize overfitting and improve computation speed.

Table 1: Descriptive analysis of input features

	cement	water	Fine Agg.	Coarse Agg.	Fly Ash	Silica Fumes	metakaolin	exposure Age	Total Age	SP	NaCl	HCl	baseline CS	CS
count	54	54	54	54	54	54	54	54	54	54	54	54	54	54
mean	342.59	188.89	735.56	1356.11	5.19	11.11	30.00	48.52	71.33	0.56	0.69	2.92	28.18	25.13
std	32.47	12.54	72.73	275.24	12.55	12.54	29.28	25.70	31.59	0.69	1.13	2.10	1.55	3.22
min	297.5	175	655	1050	0	0	0	28	28	0	0	0	25.2	17.29
25%	315	175	655	1050	0	0	0	28	56	0	0	0	27.1	23.4
50%	340	200	800	1600	0	0	20	30	58	0	0	2.5	28.1	26.25
75%	360	200	800	1600	0	25	52.5	62	90	1.26	2.5	5	29.28	27.45
max	400	200	800	1600	35	25	80	90	118	1.79	2.5	5	31.4	29.6

2.2.4 Adaptative Boosting

Adaptive Boosting is an ensemble method for enhancing accuracy by combining multiple weak models to create one highly accurate prediction model. It prioritizes misclassified occurrences in each iteration, giving more weight to ambiguous samples [24]. The technique starts by creating a base model from the complete dataset and then evaluating errors to increase the impact of misclassified occurrences. This means that future models devote more of their attention to addressing prior flaws. This adaptive strategy enables AdaBoost to build a highly accurate model using modest learners, successfully handling complex patterns while avoiding overfitting.

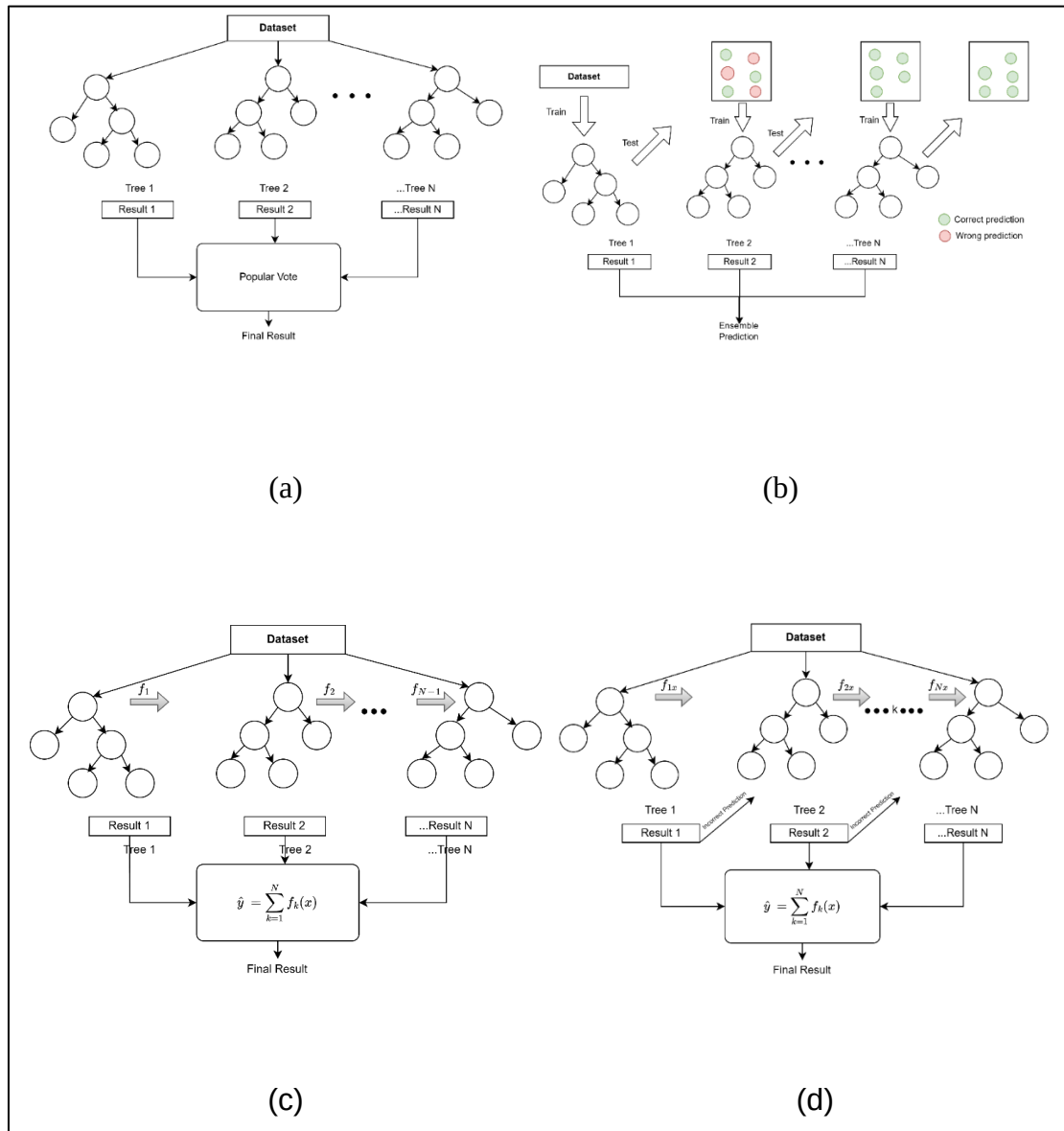


Figure 2: Diagrammatic representation of the ML models (a) Random Forest (b) Gradient Boosting (c) Extreme Gradient Boosting (d) Adaptive Boosting

2.3 Performance Evaluation of Models

Machine learning performance measures are quantitative techniques for objectively comparing models and identifying areas for improvement. The model's performance was assessed by comparing experimental baseline CS and percentage loss of CS to predicted baseline CS and CS of SCM concrete subjected to chloride ingress using the following three metric indexes: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean squared error (MSE), whose expressions are given below:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (2.6)$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (2.7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.8)$$

where y_i and $\hat{y}_i$ denote the actual and predicted values, respectively; $\bar{y}_i$ represents the mean value; and n is the number of samples in the database.

3 RESULTS AND DISCUSSION

3.1 Performance of ML Models

3.1.1 Random Forest

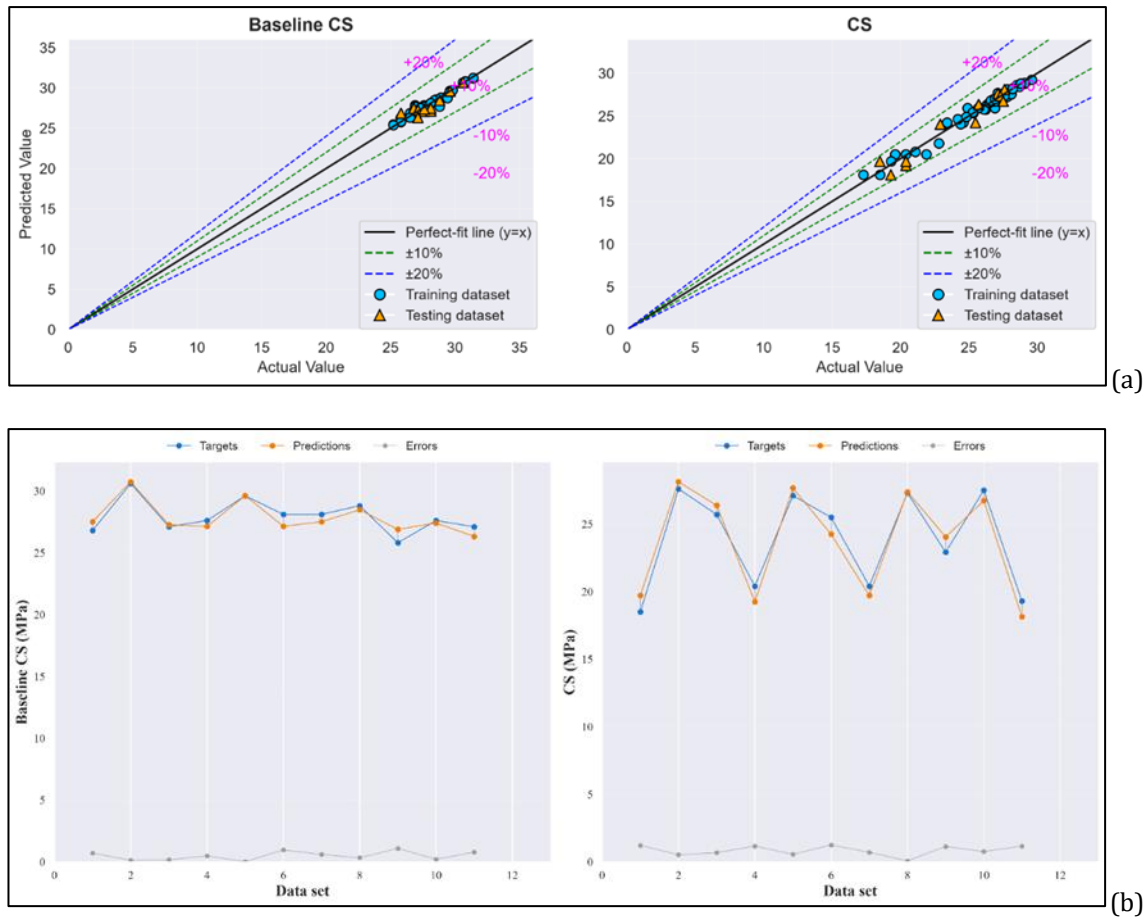


Figure 3: RF model (a) Relationship between the actual and predicted results (b) Distribution of actual and predicted values along with the errors.

Figure 4(a) shows the statistical analysis interpretation of the actual and expected results for the baseline CS and CS of SCM-based concrete subjected to chloride ingress for the RF model, with light-blue circles denoting training data and orange triangles representing test data. The solid black diagonal line represents the ideal prediction line ($y=x$), whereas the dotted lines show $\pm 10\%$ and $\pm 20\%$ error margins. This visual configuration serves as a benchmark for assessing model correctness, allowing for the clear detection of overestimation and underestimating trends. The RF model generated an output with high precision and little divergence between real and anticipated values. The RF model gave an R^2 value of 0.8547 for the testing dataset, indicating that the model is within an acceptable range in terms of result prediction. Figure 3(b) depicts the dispersion of experimental data, anticipated values, and errors in the RF model. The greatest, lowest, and average values in the distribution for baseline CS were 30.73, 26.32, and 27.81 MPa, respectively, while for the CS of the concrete subjected to chloride ingress were 28.12, 18.13, and 23.76 MPa. However, the

majority of the error data for the baseline CS values were less than or approximately 1 MPa, and for the error data of the CS values of the SCM concrete subjected to chloride were well less than 1.5 MPa.

3.1.2 Gradient Boosting

Figure 4(a) shows the statistical analysis interpretation of the actual and expected results for the baseline CS and CS of SCM-based concrete subjected to chloride ingress for the GradBoost model. The gradient boost model generated an output with high precision and little divergence between real and anticipated values. The gradient boost model gave an R^2 value of 0.8612 for the testing dataset, indicating that the model is within an acceptable range in terms of result prediction, and superior performance compared to the RF model. Figure 4(b) depicts the dispersion of experimental data, anticipated values, and errors in the RF model. The greatest, lowest, and average values in the distribution for baseline CS were 30.61, 26.17, and 27.79 MPa, respectively, while for the CS the concrete subjected to chloride ingress were 27.64, 17.89, and 23.54 MPa. 100% of the error data for the baseline CS values were less than or equivalent to 1 MPa, and for the error data of the CS values of the SCM concrete subjected to chloride ingress were less or approximately 1.5 MPa.

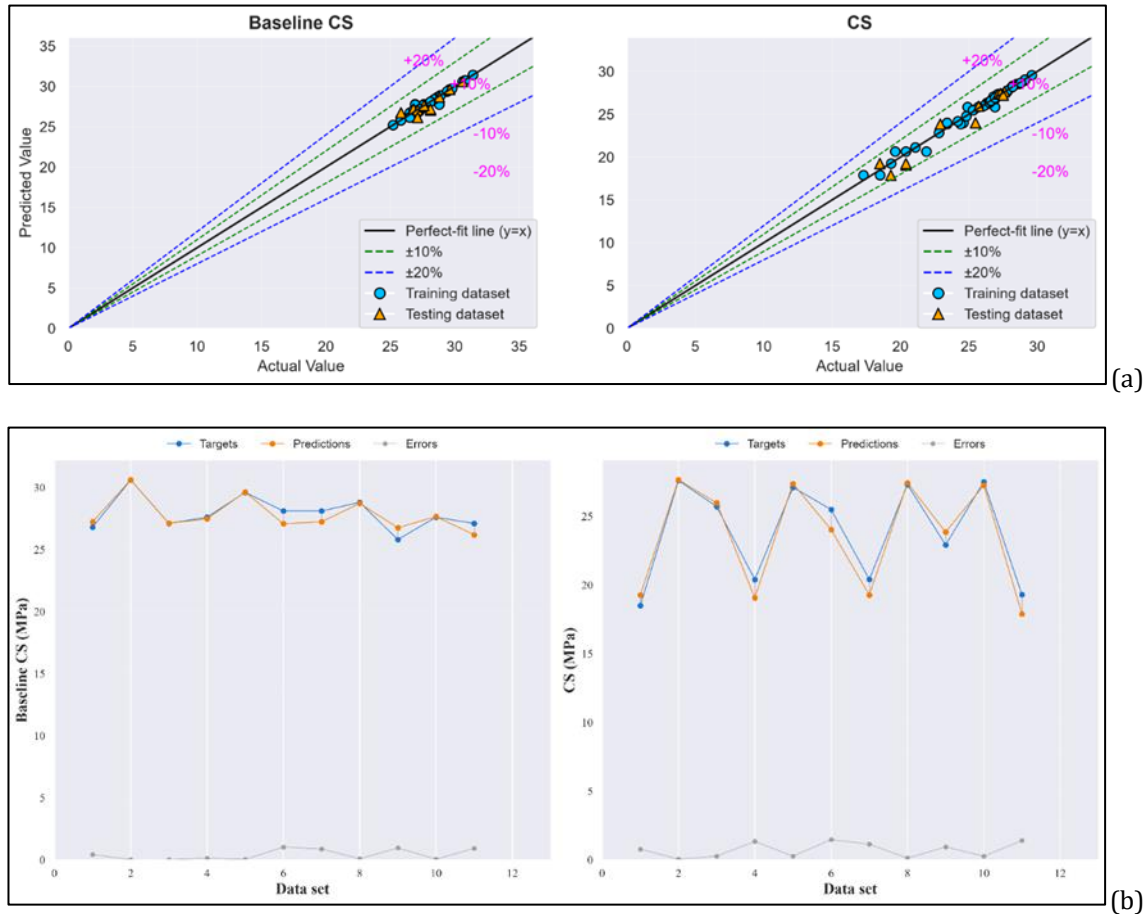


Figure 4: GradBoost model (a) Relationship between the actual and predicted results (b) Distribution of actual and predicted values along with the errors.

3.1.3 Extreme Gradient Boosting

Figure 5(a) shows the statistical analysis interpretation of the actual and expected results for the baseline CS and CS of SCM concrete subjected to chloride ingress for the XgBoost model. The XgBoost model generated an output with high precision and little divergence between real and anticipated values. The XgBoost model gave an R^2 value of 0.8593 for the testing dataset, indicating that the model is within an acceptable range in terms of result prediction. While it outperformed the RF model, it fell slightly short of the GradBoost model. Figure 5(b) shows the dispersion of experimental data, anticipated values, and errors in the RF model. The greatest, lowest, and average values in the distribution for baseline CS were 30.58, 26.15, and 27.79 MPa, respectively, while for the percentage CS loss were 27.70, 17.89, and 23.61 MPa. Majority of the error data for the baseline CS values were less than 1 MPa, and for the error data of the CS values of the SCM concrete subjected to chloride ingress were less than 1.5 MPa.

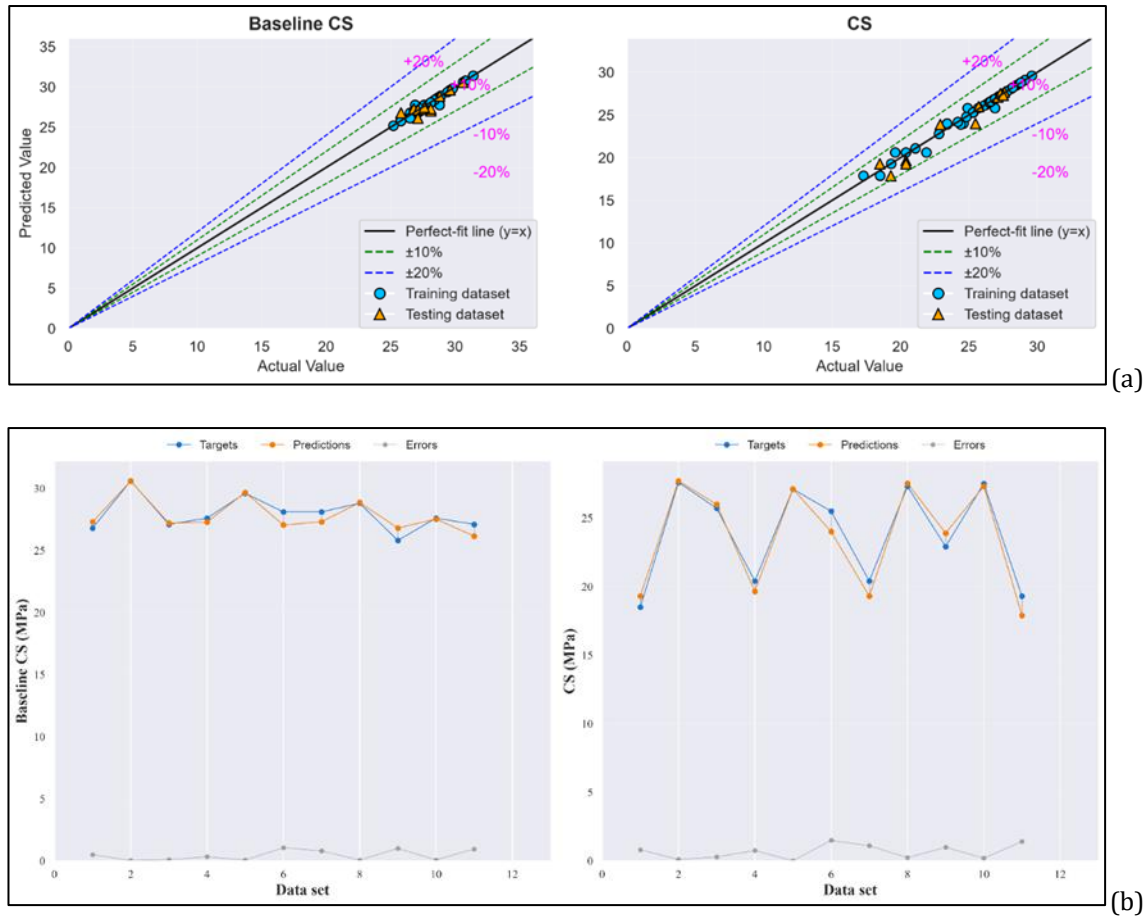


Figure 5: XgBoost model (a) Relationship between the actual and predicted results (b) Distribution of actual and predicted values along with the errors.

3.1.4 Adaptive Boosting

Figure 6(a) shows the statistical analysis interpretation of the actual and expected results for the baseline CS and CS of SCM concrete subjected to chloride ingress for the AdaBoost model. The AdaBoost model generated an output with high precision and little divergence between real and anticipated values. The AdaBoost model gave an R^2 value of 0.7938 for the testing dataset, indicating that the model is within an acceptable range in terms of result prediction. The AdaBoost model demonstrated a significantly lower R^2 value relative to the RF, GradBoost, and XgBoost models. Figure 6(b) depicts the dispersion of experimental data, anticipated values, and errors in the RF model. The greatest, lowest, and average values in the distribution for baseline CS were 30.6, 26.15, and 27.63 MPa, respectively, while for the CS of the SCM concrete subjected to chloride ingress were 27.8, 17.69, and 23.59 MPa. 100% of the error data for the baseline CS values were less than or approximately 1MPa. While for the error data, the CS values for the SCM concrete subjected to chloride ingress were less than 2.5 MPa.

Table 2: Statistical evaluations of the models.

ML Models	R^2	RMSE	MAE	MSE
Random Forest	0.8547	0.7726	0.6659	0.5969
Gradient Boost	0.8612	0.7600	0.5701	0.5776
Extreme Gradient Boost	0.8593	0.8593	0.5584	0.5354
Adaptive Boost	0.7938	0.9764	0.7398	0.9534

Overall, the GradBoost and XgBoost model delivered the most accurate and consistent predictions. In comparison, GradBoost performed the best having an R^2 value of 0.8612, and the lowest RMSE of 0.76. Which shows that the GradBoost model performs very well and makes accurate predictions. The RMSE, MAE and MSE value for all the ML models is shown in Table 2.

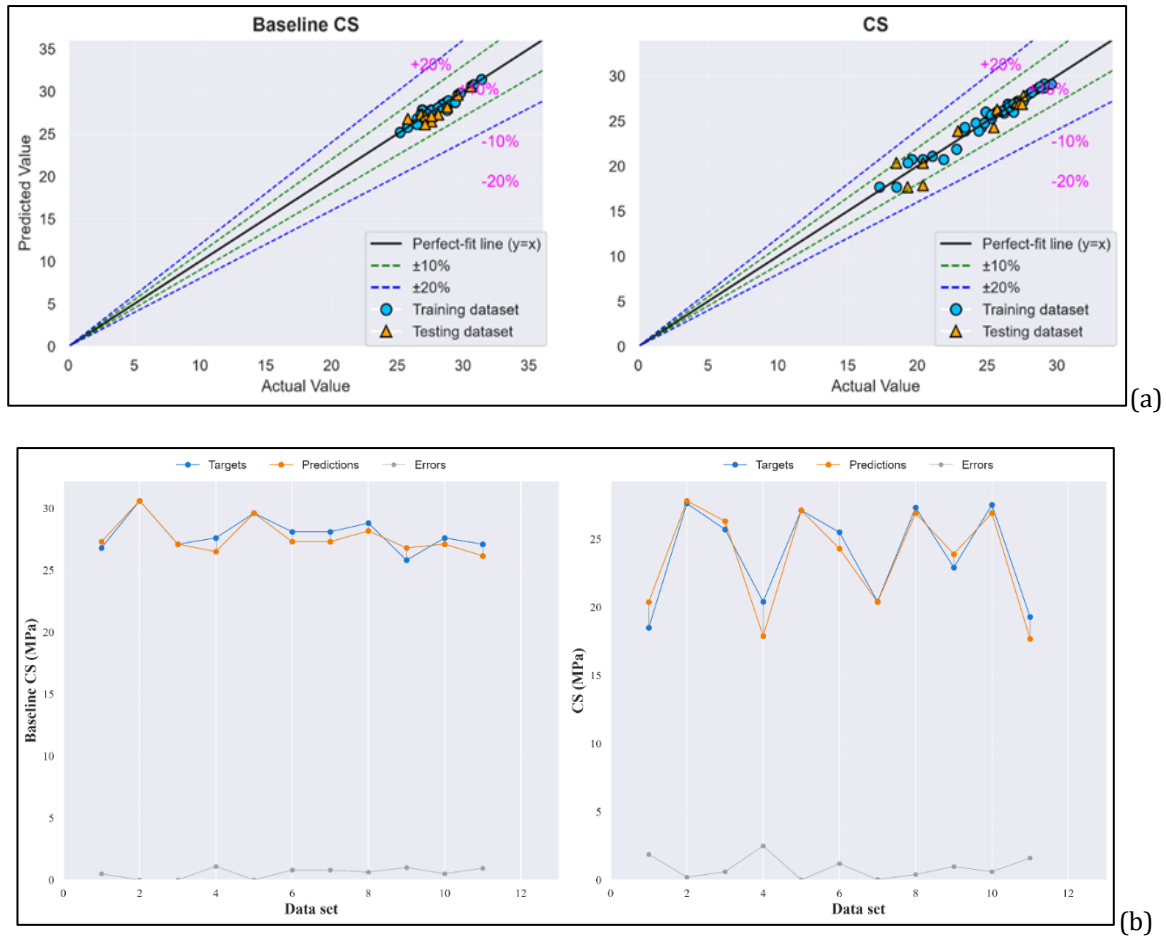


Figure 6: AdaBoost model (a) Relationship between the actual and predicted results (b) Distribution of actual and predicted values along with the errors.

3.2 K-Fold Cross-Validation

The model's legitimacy was assessed during execution using the K-fold cross-validation approach, which is typically used to assess a model's validity. Relevant data are divided into ten groups at random for K-fold cross-validation. One group is set aside for model validation, while the other nine are used for training. 80 % of the data was used to train the models, while the remaining 20% was used to test the models that were used. The correctness of the model is indicated by the high R^2 value and the lower values of the errors (RMSE, MAE, and MSE). To get an average result, the method must also be performed ten times. The k-fold cross-validation ($k = 2, 3, 4, 5, 6, 7, 8, 9, 10$) was evaluated using R^2 , MAE, MSE, and RMSE; their distributions for RF, GradBoost, XgBoost, and AdaBoost models are displayed in Table 3. When it comes to predicting outcomes, the model with the lowest error and the highest R^2 value performs better. The GradBoost model's best, lowest, and average R^2 values were 0.94, 0.86, and 0.87, respectively. whereas, the XgBoost model's best, lowest, and average R^2 values were 0.87, 0.81, and 0.83, respectively. Additionally, the GradBoost model's maximum RMSE, MAE, and MSE values were 0.94, 0.70, and 0.91 MPa, whereas the XgBoost model's corresponding values were 0.96, 0.76, and 0.94 MPa. The minimal RMSE, MAE, and MSE for the GradBoost model were 0.54, 0.41, and 0.36 MPa, respectively, while the corresponding values for the XgBoost model were 0.67, 0.53, and 0.48 MPa. The GradBoost model exhibited greater predictive accuracy than the XGBoost model, as evidenced by its lower RMSE, MAE, and MSE values, reflecting a minimized error profile.

Table 3: Results of k-fold cross-validation.

K-Fold	RF				GradBoost				XgBoost				AdaBoost			
	R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE
2	0.75	1.0	0.78	1.03	0.79	0.94	0.70	0.91	0.78	0.96	0.76	0.94	0.76	1.04	0.85	1.14
3	0.80	0.92	0.75	0.87	0.90	0.67	0.49	0.50	0.84	0.81	0.64	0.69	0.79	0.98	0.78	1.02
4	0.85	0.77	0.63	0.61	0.87	0.73	0.52	0.56	0.82	0.82	0.65	0.70	0.81	0.89	0.71	0.86
5	0.88	0.71	0.54	0.52	0.92	0.59	0.41	0.38	0.87	0.72	0.56	0.52	0.87	0.78	0.62	0.68
6	0.85	0.72	0.56	0.55	0.89	0.63	0.43	0.43	0.83	0.73	0.55	0.55	0.81	0.84	0.66	0.77
7	0.83	0.78	0.61	0.65	0.86	0.71	0.50	0.55	0.81	0.81	0.60	0.68	0.81	0.86	0.68	0.81
8	0.82	0.69	0.56	0.50	0.89	0.54	0.42	0.36	0.82	0.67	0.53	0.48	0.79	0.76	0.61	0.64
9	0.84	0.71	0.57	0.55	0.89	0.60	0.44	0.43	0.85	0.69	0.54	0.52	0.80	0.82	0.64	0.74
10	0.82	0.70	0.56	0.56	0.86	0.61	0.45	0.48	0.81	0.71	0.56	0.58	0.79	0.80	0.64	0.74

3.3 SHAP Analysis

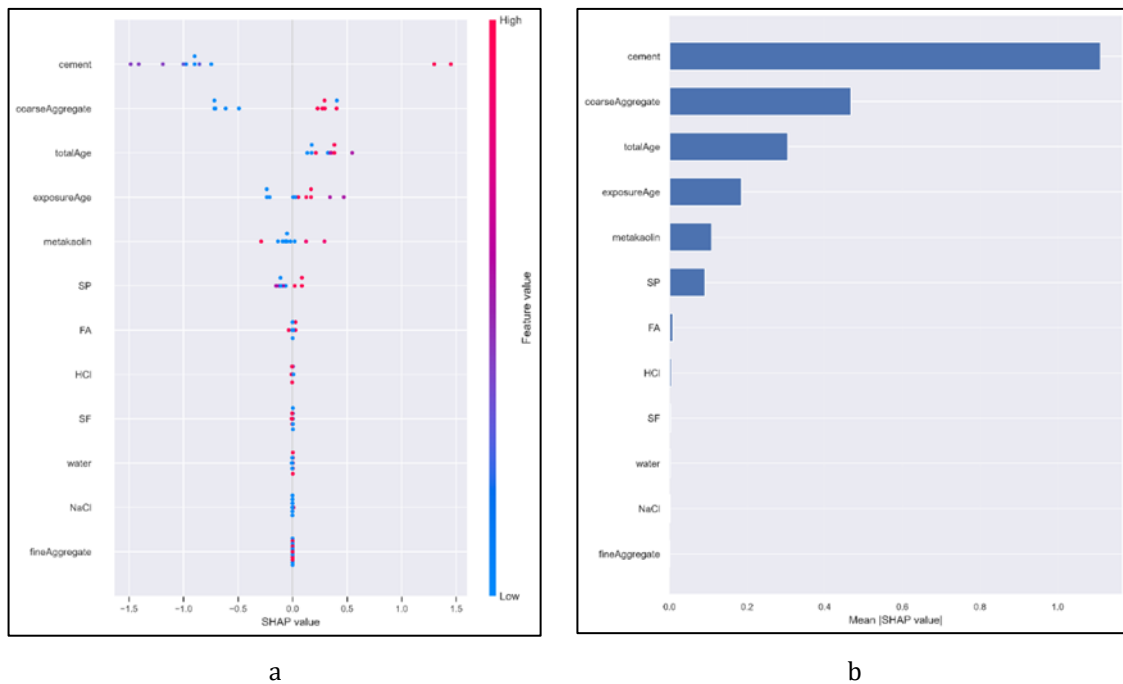


Figure 7: Influence of different parameters on baseline CS (a) SHAP impact values of the predictor, (b) Sensitivity analysis of mean SHAP values.

SHAP ensures mathematically rigorous and consistent interpretations by accounting for feature interactions and providing both per-prediction and overall model explanations. Its main goal is to improve machine learning models' transparency by accurately measuring feature influence. Figure 7(a) illustrates the SHAP analysis of the baseline CS. It shows the influence of various parameters on the baseline CS of the SCM concrete. The SHAP summary plot displays the distribution of SHAP values for each feature, highlighting both the magnitude and direction either positive or negative of their impacts on the baseline CS of the concrete. Cement appears to have the most significant impact on the baseline concrete, which means that, with higher values of cement strongly increases the predicted baseline CS of the SCM concrete. Similarly for coarse aggregate, total age of curing and exposure age to chloride ingress. While for the CS of the SCM concrete subjected to chloride ingress as shown in Figure 8(a) shows that exposure age had the most significant impact on the CS of the concrete subjected to chloride ingress, this means that the longer the exposure age the lower the CS of the SCM-concrete. This is also the case for the total age, but coarse aggregate, cement and metakaolin have a positive influence on the CS of the concrete when subjected to chloride ingress. Figure 7(b) and Figure 8(b) provides a complementary view by ranking the mean absolute SHAP values, giving a clearer picture of the

average contribution magnitude of each feature to the model output. However, parameters such as NaCl, fine aggregate, water, and SF show minimal mean SHAP values, indicating relatively negligible impact on the CS of concrete.

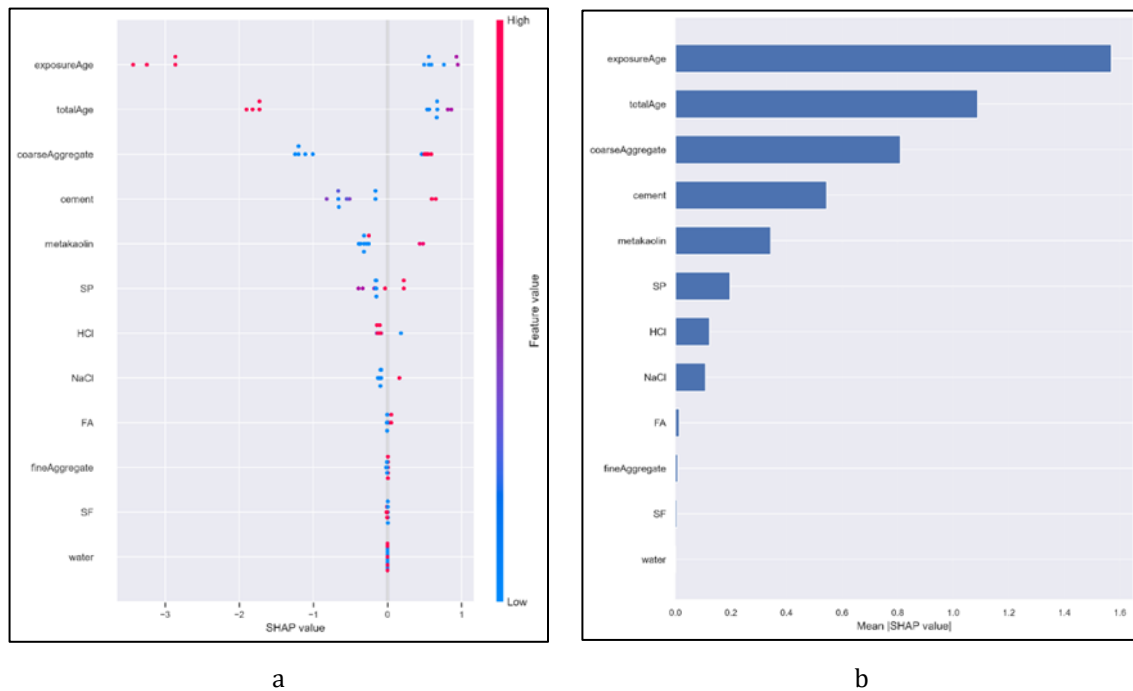


Figure 8: Influence of different parameters on CS of concrete subjected to chloride ingress (a) SHAP impact values of the predictor, (b) Sensitivity analysis of mean SHAP values.

4 CONCLUSION

The objective of this study was to evaluate and demonstrate how ML techniques can be used to forecast the compressive strength of SCM concrete subjected to chloride ingress. The baseline CS and CS of SCM concrete subjected to chloride ingress was forecasted using RF, GradBoost, XgBoost, and AdaBoost algorithms. The following conclusions can be drawn:

- The GradBoost model was more effective at prediction than the other models employed, this was due to a higher coefficient of determination (R^2) and lower error values (RMSE, MAE, MSE). The R^2 value for the GradBoost, XgBoost, RF and AdaBoost models was noted to be 0.8612, 0.8593, 0.8547, and 0.7938, respectively. Which are all within the acceptable range.
- The k-fold cross-validation approach and statistical analysis have further demonstrated the models' adequate performance. The GradBoost model's higher performance over the other models under analysis was also confirmed by these checks.
- SHAP analysis revealed that cement and coarse aggregate were the most critical features contributing majorly to the baseline CS of the SCM concrete, i.e. an increase in cement and coarse aggregate content will improve the baseline CS of the SCM concrete, while exposure age and total age contributed majorly to the CS of the SCM concrete subjected to chloride ingress, i.e. the longer the SCM concrete is subjected to chloride ingress, the lower the CS of the SCM concrete.
- Without requiring more time for sample casting and testing, ML approaches can more accurately predict the strength characteristics of SCM concrete subjected to chloride ingress.

This research bridges the gap between experimental data and practical engineering, demonstrating how ML-driven approaches can revolutionize the determination of the baseline CS and CS of SCM concrete subjected to chloride ingress. For future research, it is recommended to evaluate a wider range of machine learning algorithms for outcome prediction. Additionally, expanding the dataset size and integrating alternative machine learning architectures could further enhance model robustness and generalization. Also, other input parameters could be added to get better response from the models. In addition, more in-depth analyses such as hyper-parameter tuning should be carried out to improve the prediction accuracy of the model, and checks such as partial dependence plot (PDP) as an addition to SHAP analysis to help identify trends, thresholds, or nonlinear patterns, for a better evaluation and understanding of the outcomes from ML techniques.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

I am deeply grateful to my co-authors, for their dedication and insightful discussions and team work that made this work possible.

Disclosure of conflict of interest

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

Authors' Contributions

- Conceptualization, Methodology, Data Validation & Cleaning, Analysis, Visualization, Writing – Original Draft – Osasere D. Ikponmwosa
- Data Curation & Validation and Revised Draft – Adetola A. Falade
- Revised Draft – Saheed Adedokun

REFERENCES

- [1] Barbhuiya, S., Kanavaris, F., Das, B. B., & Idrees, M. (2024). Decarbonising cement and concrete production: Strategies, challenges and pathways for sustainable development. *Journal of Building Engineering*, 86, 108861. <https://doi.org/10.1016/j.jobe.2024.108861>
- [2] Snellings, R., Suraneni, P., & Skibsted, J. (2023). Future and emerging supplementary cementitious materials. *Cement and Concrete Research*, 171, 107199. <https://doi.org/10.1016/j.cemconres.2023.107199>
- [3] Liu, W., Mo, H., & Lee, C. K. (2026). Compressive strength prediction for SCM-based concretes using a multiple machine learning models and datasets approach. *Results in Engineering*, 30, 110262. <https://doi.org/10.1016/j.rineng.2026.110262>
- [4] Pillay, D. L., Olalusi, O. B., Kiliswa, M. W., Awoyera, P. O., Kolawole, J. T., & Babafemi, A. J. (2021). Engineering performance of metakaolin based concrete. *Cleaner Engineering and Technology*, 6, 100383. <https://doi.org/10.1016/j.clet.2021.100383>
- [5] Sunday, G. M., D, E. P., & A, S. I. (2020). Effects of Aggressive Environment on the Durability of Concrete Produced with Binary and Ternary Blend Cement. *International Journal of Engineering Research And*, V9(02). <https://doi.org/10.17577/ijertv9is020077>
- [6] Farnam, Y., Wiese, A., Bentz, D., Davis, J., & Weiss, J. (2015). Damage development in cementitious materials exposed to magnesium chloride deicing salt. *Construction and Building Materials*, 93, 384–392. <https://doi.org/10.1016/j.conbuildmat.2015.06.004>
- [7] Zhang, X., Zhang, Y., Liu, B., Liu, B., Wu, W., & Yang, C. (2021). Corrosion-induced spalling of concrete cover and its effects on shear strength of RC beams. *Engineering Failure Analysis*, 127, 105538. <https://doi.org/10.1016/j.engfailanal.2021.105538>
- [8] Pontes, J., Bogas, J. A., Real, S., & Silva, A. (2021). The rapid chloride Migration test in assessing the chloride penetration resistance of normal and lightweight concrete. *Applied Sciences*, 11(16), 7251. <https://doi.org/10.3390/app11167251>
- [9] Behera, A., Pratap, B., Kumar, P., Rout, M. K. D., & Sembeta, R. Y. (2025). Prediction of rapid chloride permeability using silica fume, fly ash, GGBS and micro fibers based geopolymer concrete. *Scientific Reports*, 16(1), 959. <https://doi.org/10.1038/s41598-025-30430-6>
- [10] Kumar, P., Gogineni, A., & Ammarullah, M. I. (2025). Sustainable bioengineering approach to industrial waste management: LD slag as a cementitious material. *Discover Sustainability*, 6(1). <https://doi.org/10.1007/s43621-025-00981-9>
- [11] Shetty M. S. and Jain A. K. (2019). Concrete Technology (Theory and Practice), 8th ed., New Delhi, India: S. Chand Publishing.
- [12] Ahmad, W., Ahmad, A., Ostrowski, K. A., Aslam, F., Joyklad, P., & Zajdel, P. (2021). Application of advanced machine learning approaches to predict the compressive strength of concrete containing supplementary cementitious materials. *Materials*, 14(19), 5762. <https://doi.org/10.3390/ma14195762>

- [13] Moradi, N., Tavana, M. H., Habibi, M. R., Amiri, M., Moradi, M. J., & Farhangi, V. (2022). Predicting the compressive strength of concrete containing binary supplementary cementitious material using machine learning approach. *Materials*, 15(15), 5336. <https://doi.org/10.3390/ma15155336>
- [14] Wang, R., Huo, Y., Wang, T., Hou, P., Gong, Z., Li, G., & Li, C. (2024). Machine Learning Method to Explore the Correlation between Fly Ash Content and Chloride Resistance. *Materials*, 17(5), 1192. <https://doi.org/10.3390/ma17051192>
- [15] 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*, 13(1), 15024. <https://doi.org/10.1038/s41598-023-42270-3>
- [16] 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. <https://doi.org/10.3389/fmats.2024.1445547>
- [17] Wang, W., Wang, S., & Lin, C. (2016). Effect of hot environment on strength and heat transfer Coefficient of Nano-Clay Concrete Paper title. *Journal of Materials Science and Chemical Engineering*, 04(07), 45–52. <https://doi.org/10.4236/msce.2016.47007>
- [18] Awan, M. S., Ali, A., S, P., & Awan, Y. S. (2019). Carbon Nano fibre reinforcements in concrete. *Indonesian Journal of Science and Technology*, 4(1), 1. <https://doi.org/10.17509/ijost.v4i1.4140>
- [19] Wu, X., Zheng, S., Feng, Z., Chen, B., Qin, Y., Xu, W., & Liu, Y. (2022). Prediction of the frost resistance of high-performance concrete based on RF-REF: A hybrid prediction approach. *Construction and Building Materials*, 333, 127132. <https://doi.org/10.1016/j.conbuildmat.2022.127132>
- [20] Ortenzi, M. A., Basilissi, L., Farina, H., Di Silvestro, G., Piergiovanni, L., & Mascheroni, E. (2015). Evaluation of crystallinity and gas barrier properties of films obtained from PLA nanocomposites synthesized via “in situ” polymerization of l-lactide with silane-modified nanosilica and montmorillonite. *European Polymer Journal*, 66, 478–491. <https://doi.org/10.1016/j.eurpolymj.2015.03.006>
- [21] Feng, D., Liu, Z., Wang, X., Chen, Y., Chang, J., Wei, D., & Jiang, Z. (2019b). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000. <https://doi.org/10.1016/j.conbuildmat.2019.117000>
- [22] Alamri, M., Ali, T., Ahmed, H., Qureshi, M. Z., Elmagarhe, A., Khan, M. A., Ajwad, A., & Mahmood, M. S. (2024). Enhancing the engineering characteristics of sustainable recycled aggregate concrete using fly ash, metakaolin and silica fume. *Heliyon*, 10(7), e29014. <https://doi.org/10.1016/j.heliyon.2024.e29014>
- [23] Falade, A. a. S. I. S. M. a. S. M. a. M. A. (2024). Effects of metakaolin as a pozzolanic material on chloride ingress in concrete. *NIPES Journal of Science and Technology Research*. <https://doi.org/10.5281/zenodo.14316852>
- [24] Ye, M., Li, L., Yoo, D., Li, H., Zhou, C., & Shao, X. (2023). Prediction of shear strength in UHPC beams using machine learning-based models and SHAP interpretation. *Construction and Building Materials*, 408, 133752. <https://doi.org/10.1016/j.conbuildmat.2023.133752>
- [25] Leo Breiman. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- [26] Schonlau, M.; Zou, R.Y. (2020). The random forest algorithm for statistical learning. *Stata J.* 20, 3–29.
- [27] Yin, L., Lin, S., Sun, Z., Li, R., He, Y., & Hao, Z. (2023). A game-theoretic approach for federated learning: A trade-off among privacy, accuracy and energy. *Digital Communications and Networks*, 10(2), 389–403. <https://doi.org/10.1016/j.dcan.2022.12.024>
- [28] Jerome H. Friedman. Greedy function approximation. (2001). a Gradient Boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232.