REVISTA MATÉRIA V.30
ISSN 1517-7076 articles e20240989, 2025

A supervised machine learning modeling with parametric optimization of the compressive strength of graphene-nano-engineered concrete

Md. Habibur Rahman Sobuz[1,5], Md. Kawsarul Islam Kabbo[1], Abdullah Alzlfawi[2], Mohammed Jameel[3], Noor Md. Sadiqul Hasan[4], Md. Munir Hayet Khan[5]

[1]Khulna University of Engineering & Technology, Department of Building Engineering and Construction Management. 9203, Khulna, Bangladesh.

[2]Majmaah University, College of Engineering, Department of Civil and Environmental Engineering. 11952, Al Majmaah, Saudi Arabia.

[3]King Khalid University, College of Engineering, Department of Civil Engineering. Asir, 61421, Abha, Saudi Arabia.

[4]International University of Business Agriculture and Technology, College of Engineering and Technology, Department of Civil Engineering. 1230, Dhaka, Bangladesh.

[5]INTI International University, Faculty of Engineering & Quantity Surveying. Persiaran Perdana BBN, Putra Nilai, 71800, Nilai, Negeri Sembilan, Malaysia.

ABSTRACT

This study explores the potential of nano-graphene particles for the sustainable manufacturing of concrete. The primary goal is to predict the compressive strength of graphene-incorporated concrete by evaluating the effects of materials such as cement, graphene, fly ash, water usage, aggregate levels, curing ages, and superplasticizer dosages. A total of 350 data entries were sourced from various literature. Multiple machine learning techniques, such as Adaptive Boosting, Decision Tree, Gradient Boosting, k-nearest Neighbors, Light Gradient Boosting, and XGBoost, were utilized to study how these variables influence compressive strength. The dataset was split into an 80%–20% train and test set for developing the prediction algorithms. Among the models, the XGBoost model delivered the highest precision, with a coefficient of determination of 0.921 in the training stage. To ascertain how the input parameters affected the outcome, SHAP and partial dependence plots were employed for each variable's contribution to the improvement of strength. Based on the parametric optimization, incorporating graphene at a range of 0–2 kg/m^3 in the concrete mix led to the maximum increase in compressive strength. This work enhances machine learning-based nano-engineered concrete production by removing experimental methods, reducing labor and resource demands, minimizing environmental footprint, and providing graphene-modified concrete design data.

Keywords: Graphene; Machine learning; Compressive strength; Nanomaterials; Parametric analysis.

1. INTRODUCTION

Concrete is the most prevalent construction material due to its exceptional versatility, availability of raw resources, and minimal maintenance expenses [1, 2]. However, there is a lot of environmental pressure since concrete manufacturing is a major contributor to CO_2 emissions on a global scale [3, 4]. Every year, the globe produces over 25 billion tons of concrete [5, 6]. Clinker production for Portland cement accounts for 7% of total CO_2 emissions annually [7]. Compliance with regulations pertaining to sustainable development and energy saving is improving within the concrete production industry [8, 9]. Furthermore, conventional concrete used as a building material exhibits brittleness and weakness in tension and flexural strength [10, 11]. It could fracture and deform under severe loads and environmental conditions, diminishing strength [12]. Enhancing cement-based materials at both the macro and nanoscale is crucial for mitigating fracture propagation and improving strength [13]. The incorporation of nano-engineered particles (NEPs) may augment the strength of cementitious materials by facilitating a bridging effect inside the cement matrix [14, 15]. The use of nanoscale concrete reinforcing components is essential to successfully inhibit fracture propagation since fractures often initiate at the nanoscale and then expand to a larger size [16]. Nanotechnology has achieved worldwide acclaim for its outstanding efficacy in several sectors. In the construction industry, NEPs have the capacity to transform the

Corresponding Author: Md. Habibur Rahman Sobuz Received on 31/12/2024 Accepted on 05/05/2025

DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0989

production of cement-based composites. Nanomaterials are often categorized as particles measuring between 1 and 100 nanometers in size [17, 18]. Additionally, several nanomaterials have been used to augment the strength of traditional concrete [19–22].

However, among nanomaterials, engineered-nano-graphene (EG) has garnered significant attention owing to its unique nanoscale architecture and superior mechanical and electrical characteristics [23]. Furthermore, EGs are incorporated into cement composites to enhance their mechanical attributes or to provide self-sensing capabilities [24]. LIU *et al.* [25] used EG concentrations between 0.5% and 2.5% of binder in concrete formulations. The findings indicated that the mechanical properties of the concrete were not significantly altered by the incorporation of EGs. Nonetheless, the incorporation of 1.5% EGs led to an 80% reduction in water penetration and a 12% augmentation in the chloride ion permeability. In a similar manner, CHEN *et al.* [26] examined specimens of EG-reinforced concrete with EG concentrations varying from 0.02% to 0.4%. The author established that the CS is enhanced by 22% at an EG dosage of 0.05%. SUN *et al.* [27] established the percolation threshold for the resistivity of EG-filled cementitious composites at roughly 2% by volume. Cementitious composites containing 5% by volume of EGs demonstrated a significant 15.6% fractional change in electrical resistivity when exposed to a Cs of 20 MPa. DIMOV *et al.* [28] similarly showed a significant gain of CS and rupture strength for EG-infused concrete. The EGs effectively bridge fissures and voids inside the matrix. Consequently, it enhances the density of the concrete and lowers the voids in the matrix. Concrete is a heterogeneous substance characterized by intrinsic variability in property evaluation [29]. The implementation of EG has significantly enhanced characteristics. Predicting these features is difficult because of the complex composition and nonlinear behavior of concrete [30]. Also, there are a lot of issues and limitations that might make it hard to do comprehensive lab work, such as finding an appropriate location to store and cure concrete mixtures, which can be removed by adopting modeling-based analysis and prediction [31, 32].

To eliminate experimental-based difficulties, finite element and numerical analysis have now emerged as an effective alternative to predict and model characteristics of concrete and structural elements [33–36]. In addition, machine learning (ML) based optimization and modeling techniques in civil engineering have recently attracted significant interest [37–40]. In addition to properties prediction, ML can also be used in other fields such as estimating bond strength of reinforced concrete [41], modeling surface roughness [42] and other properties of modified concrete [43, 44]. Several prominent ML methods, including artificial neural networks (ANN), random forests (RF), and decision trees (DT), have been successfully used to address complex regression issues in material characterization [45–47]. BARKHORDARI *et al.* [48] used hybrid deep neural networks to reduce the variance of the prediction model and achieve a precise estimation of the CS of fly ash-based concrete. DUAN *et al.* [49] investigated the compressive strength of recycled aggregate concrete through four AI techniques based on a meta-heuristic search of the sociopolitical algorithm and found ICA-XGBoost as the optimum model for prediction. Furthermore, JAF *et al.* [50] developed multi-scale models to predict the strength of fly ash-based concrete and concluded that the ANN model has the highest efficiency and accuracy in predicting the CS of concrete in different strength ranges. SALAMI *et al.* [51], along with KALOOP *et al.* [52] used genetic programming (GEP), adaptive boosting (ADB), and gradient boosting (GBR) to forecast concrete strength. MANGALATHU and JEON [53] developed shear strength prediction models for beam-column joints with RF, SOBUZ *et al.* [54] used boosting-based methods, including GBR, ADB, and Extreme Gradient Boost (XGB), to predict the CS of concrete. XGB exhibited the best performance, with an R^2 value over 0.90. XGB had the highest RMSE and MAE values, making it the most precise ML algorithm used by MARZEC *et al.* [55] for predicting the shear capacity of concrete beams. UDDIN *et al.* [56] demonstrated that Light Gradient Boosting (LGB) outperforms XGB, Support Vector Machine (SVM), and RF models in accurately evaluating the strength of 3D-printed concrete. Their research underscores the advantages of ML compared to regular empirical methods. Besides these, modeling and simulating are also familiar for structural elements and their behavior prediction [57–59].

Recent studies on ML methods for estimating the CS of nano-graphene-infused concrete identify two primary limitations: the need for more advanced and precise ML modeling and the absence of comprehensive parametric assessments, including SHAP and PDP. Moreover, prior research depended on datasets that lacked comprehensiveness and had restricted data ranges, so impacting the generalizability of the findings. This work used and contrasted six machine learning models—ADB, DT, GBR, kNN, LGB, and XGB—to address the deficiency in the current research. The study used eight essential input criteria and analyzed 350 data points from prior research, significantly enhancing the precision of the CS anticipation. Furthermore, SHAP and PDP were used in the research to assess the influence of factors on the prediction of CS and to elucidate the impact of input parameters on the CS of EGC. The structure of this manuscript is organized as follows: Section 2 details the research methodology and provides a comprehensive explanation of the machine learning algorithms and parametric modeling used in this study, including their theoretical backgrounds and working principles. After that, section 3 presents and discusses the results, including dataset characteristics, model performance evaluation,

SHAP and PDP analyses, and model comparison metrics in separate subsections. Finally, the last section (section 4) summarizes the key findings and concludes the study, highlighting the practical implications and future research directions.

2. RESEARCH METHODOLOGY

ML algorithms generally utilize statistical techniques to identify patterns within large datasets, which allows the model to improve its performance over time through experience [39]. This study utilizes six input features fed into the developed targeted model, where they undergo processing within a hidden layer composed of interconnected neurons. The hidden layer is generally accountable for learning sophisticated patterns and interactions between the input variables [60]. Each of the models used in this study has been described below:

2.1. Adaptive boosting (ADB)

ADB is a popular ensemble learning algorithm that uses adaptive weight distribution over training instances to select multiple classifier instances iteratively [61]. The technique is called adaptive boosting because a new set of weights is assigned to each instance. Greater weights are assigned to instances with incorrect labels. Minimizing classification error is interpreted as optimizing a non-smooth, non-differentiable cost function, for which an exponential loss may serve as the best approximation [62]. As a result, ADB performs exceptionally well in a wide range of classification tasks. The weight of the output and model for each sample is first determined after a sub-model has been learned for the first weak learner. The model is retrained after adjusting the sample for the next iteration based on the weights from the previous model. A new weak learner is then generated, and its model weight is determined, repeating this process for N iterations. If the preceding model carries a higher weight than the subsequent one, the prediction will be inaccurate [63]. The final model output is the weighted sum of these weak learners, which is created by combining the learners with low predictive power with different weights after N iterations. For a typical regression task, the training dataset Ô can be represented as follows:

$$\hat{O} = [(X_1,Y_1), (X_2,Y_2), \ldots \ldots, (X_m,Y_m)] \tag{1}$$

Here, the training data set consists of m samples, denoted as (X_i,Y_i) for $i = 1$ to m; X_i represents the input data vector, while Y_i represents the corresponding data value for output.

2.2. Decision tree (DT)

DT uses data from training to create a model that resembles a tree. The root node, decision node (sometimes it also called the internal node), and terminal node (also called the leaf node) are the three nodes that make up the DT model. Three crucial elements must be addressed during this process: Finding the best splits, defining a criterion for the splitting process to stop, and figuring out which output goes to which terminal node are the first three steps in the process. DTs organize cases into categories by arranging them in a hierarchical structure, with the root node and a number of leaf or end nodes representing each category. We assign specific values or classes to the non-partitioned subgroups of X that the terminal nodes represent. Here is the equation that represents the model:

$$R(T) = \frac{1}{N}\sum_{x_i}^{n}(y_i - \bar{y}_t)^2 \tag{2}$$

Here, R(T) is the error measurement, $\bar{y}_t$ is the fixed value of output for every final node, and (x_t, y_i) is the training dataset. N denotes occurrences in training.

2.3. Gradient boosting (GBR)

An ensemble machine learning technique for regression tasks is the Gradient Boosting (GB) model. For many regression situations, it is the preferred boosting strategy and is superior to other ensemble techniques like RF. GB is commonly employed for data analysis tasks such as classification and regression. In GB, regression trees are not chosen at random, and instead, strong dictionary pruning is utilized in its place. GB, like AdaBoost, adjusts for historical errors by progressively adding forecasters to the ensemble [52]. In GB, many functional trees are trained consecutively with the goal of decreasing the residuals in order to fix the mistakes of the earlier trees. Though simpler versions sometimes leave out the regularization component, GB minimizes an objective function by combining a loss term and a regularization term. This process is repeated until the necessary precision is achieved.

2.4. k-Nearest neighbors (kNN)

kNN is a non-parametric machine learning classification and regressor. The KNN algorithm categorizes a data point based on the classifications of its neighboring points. The distance between points is often calculated using metrics, primarily Euclidean distance, but additional metrics may be used as necessary [64]. kNN is recognized for its conceptual clarity and efficacy, particularly in scenarios where the selection boundary is nonlinear. Conversely, it requires substantial computer resources and is susceptible to extraneous characteristics, rendering it vulnerable to overfitting or underfitting without appropriate calibration.

2.5. Light gradient boosting (LGB)

The LGB technique is a well-established approach in ML-technique that utilizes GBR combined with decision tree methodologies [65]. This method incorporates a histogram-based strategy to transform continuous variables into discrete classes, facilitating the efficient analysis of extensive datasets while ensuring robust predictive capabilities, particularly for regression applications. Notable benefits of the LGB technique encompass swift model training, impressive accuracy in forecasting, consistency, minimal memory consumption, and the ability to perform training in parallel [65]. The LGB approach adopts a growth strategy that prioritizes leaf nodes, where the data is segmented at the leaf level, emphasizing the most significant data points during tree formation [66]. The ultimate result from the LGB method is achieved by averaging the forecasts from each DT, with information perceived as the expected lowering in entropy stemming from node splits based on particular attributes.

2.6. Extreme gradient boosting (XGB)

Established by Chen and Guestrin in 2016 [67], the XGB technique has become a reliable tool for data science researchers, known for its tree-based ensemble learning framework. XGB is based on the Gradient Boosting Regression (GBR) structure, utilizing multiple functions to predict outcomes [68], according to,

$$\bar{z}_t = z_i^0 + \eta \sum_{k=1}^{M} f_k(X_i) \tag{3}$$

The XGB algorithm predicts $\bar{z}_i$ for a sample based on its feature set as X_i. The model employs M estimators, each represented by a separate decision tree f_k. Initially, z_i^0 is defined as the mean of the training set values, shown in equation (3). XGB is used to address regression, classification, and ranking issues in supervised learning. Because of its excellent problem-solving capabilities and low feature engineering requirements, it has been widely used in industry. XGB is a learning technique that builds multiple decision trees, each of which attempts to improve predictions by fitting residuals to fix the mistakes of the ones that came before it. XGB integrates the predictions of numerous trees to produce a final result (CS) when modeling a continuous goal variable. For this, XGB uses an objective function F that accumulates a loss function and a regularization term explained by the equation (4):

$$F = \sum_{i=1}^{n} l(y_i, \bar{y}_t + \sum_{k=1}^{k} \bigcap (f_k) \tag{4}$$

Where, $(y_i, \bar{y}_i)$ is the loss function for the prediction $\bar{y}_i$ and true target y_i. $\cap$ (f_k) is the regularization erm for each tree f_k, which helps prevent overfitting.

2.7. SHapley Additive exPlanation (SHAP)

SHAP is a parametric technique that explains each prediction a model makes. It evaluates the weight of each input to the final prediction using concepts from game theory, particularly Shapley values [69]. SHAP calculates the relative contribution of the input values to the model's output in relation to the baseline prediction. The SHAP summary plot displays the input features in order of their significance in determining the output of the MLP model. Each data sample is represented by a dotted circle with red and blue hues to indicate the high and low values of a feature. The model's prediction accuracy is increased by adding each input feature value to the collection of all the features in the data model that are on the positive side of the SHAP summary plot. The SHAP can be expressed as:

$$h(\acute{z}) = \emptyset_o + \sum_{i=1}^{N} \emptyset_i \acute{z}_i \tag{5}$$

In equation (5), h stands for the model used to provide explanations, while z′ represents the foundational features. The symbol N indicates the maximum size of the coalition and ∅ refers to the attribution assigned to each feature.

2.8. Partial dependency plot (PDP)

ML models in business are designed to generate insights for decision-making. Interpretability is critical for helping developers understand, refine, and compare models, as well as allowing businesses to explain model predictions. ML models, which are frequently referred to as "black boxes," require interpretability tools such as partial dependence plots (PDPs) to illustrate the isolated effect of different characteristics on outcomes. PDPs can disclose whether a feature's connection to the target is linear, monotonic, or complex, providing a clear picture of feature influence and promoting model transparency [68].

$$\widehat{f_{xs}}(x_s) = E_{xc}\left[\widehat{f_{xs}}\left(x_s, x_c\right)\right] = \int \widehat{f_{xs}}\left(x_s, x_c\right) dP\left(x_c\right) \tag{6}$$

In partial dependency, equation (6), x_s denotes the feature(s) of interest, while x_c encompasses all other features utilized by model $\widehat{f_{xs}}$. The combined vectors x_s and x_c cover the entire feature space x. Partial dependency isolates how x_s influences predictions by averaging model outputs over x_c, taking into account both individual and combined effects.

2.9. Performance indices

To evaluate the accuracy and predictive capability of every model, numerous performance criteria have been tested, such as RMSE, MAE, and MAPE. The equations to quantify the performance of ML algorithms are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y - Y_i)^2} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y - Y_i\right|^2 \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \tag{9}$$

Where Y denotes the avg. of the experimental values, Y_i denotes the average of the expected values, and n denotes the total number of samples. Figure 1 displays the whole working procedure of ML-based analysis of CS of EGC.

3. RESULTS

3.1. Analysis of dataset

The dataset used for analyzing EGC properties comprises 350 unique mixture samples carefully collected from numerous published sources. Among the previously published literature, only the Scopus and Web of science indexed well-known journals are taken into consideration for ensuring accurate mix compositions. This compilation of data provides eight critical features for each sample, forming a comprehensive foundation for machine learning predictions. Among these, CS serves as the primary target variable that models aim to predict. The remaining input parameters include cement content, graphene, fly ash content, fine aggregates, coarse aggregates, water content, curing duration, and superplasticizers. The input parameters for this study were chosen based on their direct influence on the overall mechanical performance of concrete. These features are also the most commonly reported and experimentally validated variables affecting compressive strength in prior studies on conventional and nano-modified concrete. Graphene was specifically included due to its novel role as a nano-reinforcement, while the remaining constituents represent the essential mix components in concrete design. Curing age was considered to capture strength development over time. Through such an extensive

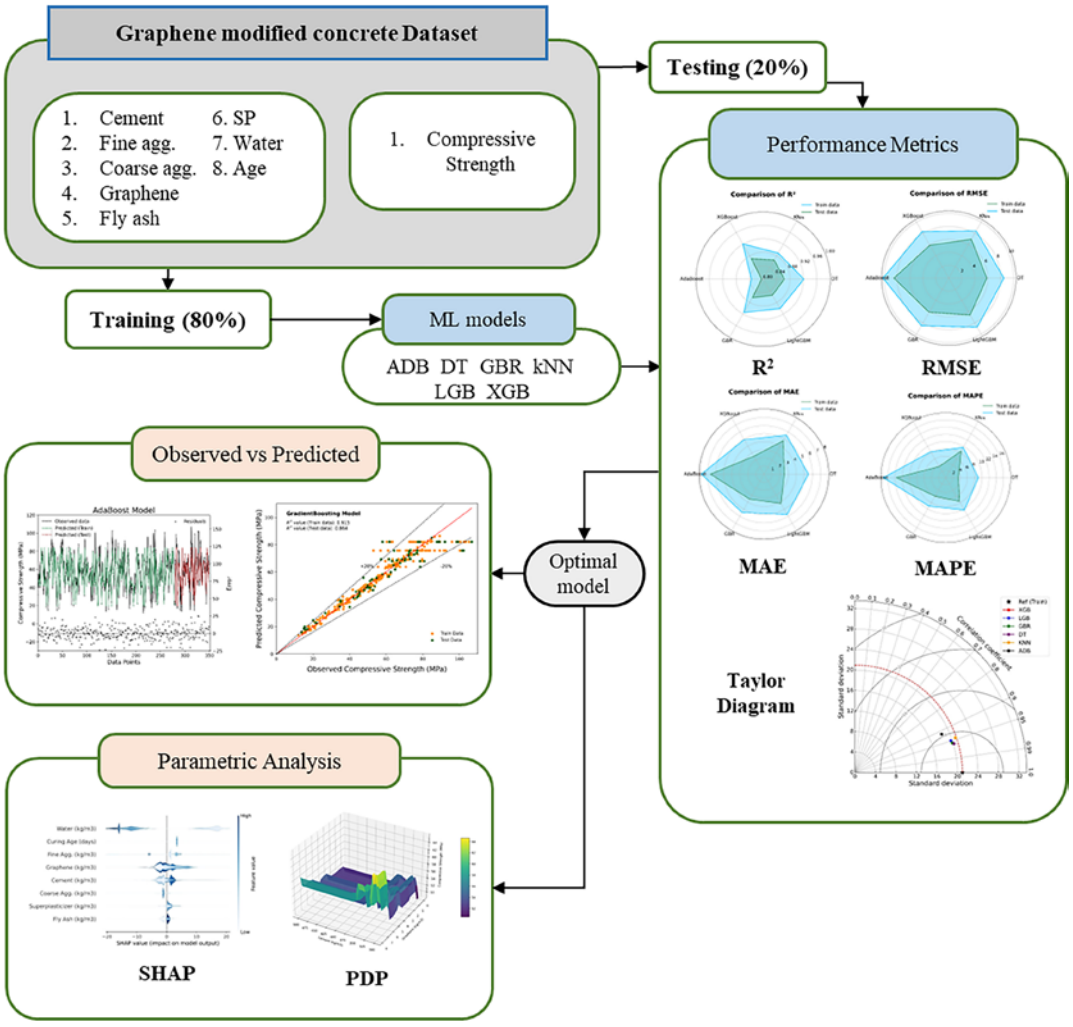

Figure 1: Methodological flow chart.

Table 1: Statistical distribution of database's features.

VARIABLE	MIN.	MAX.	AVG.	50%	ST. DEV.	COV.
Cement (kg/m^3)	300.00	500.00	393.16	418.00	63.03	15.06
Graphene (kg/m^3)	0.00	7.80	0.31	0.29	0.73	174.19
Fly Ash (kg/m^3)	0.00	210.00	45.07	100.00	91.42	138.92
Coarse Aggregate (kg/m^3)	520.00	1190.00	947.44	1104.00	184.10	15.63
Fine Aggregate (kg/m^3)	380.00	1000.00	738.15	715.00	151.54	18.01
Superplasticizer (kg/m^3)	0.00	10.50	3.41	4.02	3.89	81.82
Water (kg/m^3)	140.00	224.00	169.78	156.00	23.82	11.43
Curing Age (days)	1.00	120.00	28.48	28.00	25.38	107.55
Compressive Strength (MPa)	12.00	106.71	54.72	57.34	21.45	35.84

dataset, models can more accurately assess how variations in these components influence the material's structural properties, aiding in the development of optimized EGC mixes. Table 1 summarizes the input and output datasets, with key statistical measures for each parameter. The table contains minimum, average, and maximum values, 50% quartile values, as well as the standard deviation (St.Dev.) and coefficient of variation (COV), providing a concise assessment of the dataset's statistical properties.

Figure 2 depicts the correlations between CS and various components in EGC mixtures, emphasizing the ideal ranges for each ingredient. Figure 2(a) shows cement content ranging from 380 to 460 kg/m^3, with some combinations approaching 500 kg/m^3. CS values typically fall between the same ranges, indicating optimal cement content for strength. In Figure 2(b), graphene concentrations of 0 to 1 kg/m^3 lead to increased strength, with CS values ranging from 15 to 100 MPa. Figure 2(c) shows that CS peaks at low fly ash (FA) concentration (0–50 kg/m^3) or between 100 and 150 kg/m^3, indicating optimal ranges for FA use. A cluster of FA dosage has also been observed at a high concentration of nearly 210 kg/m^3. Figure 2(d) shows that coarse aggregates tend to concentrate at 800-1000 kg/m^3 or within the 1100-1200 kg/m^3 range, with highest CS found when the aggregate content exceeds 1100 kg/m^3. Figure 2(e) shows that fine aggregate levels range from 600 to 900 kg/m^3, with peak CS near 600 kg/m^3, suggesting an optimal balance for structural stability.

Figure 2(f) shows that superplasticizer usage is consistent across samples, with most values ranging from 1 to 5 kg/m^3, potentially improving workability consistency. Figure 2(g) shows water content changing between samples, with the best strength occurring at water levels between 140 and 200 kg/m^3. This highlights the need for perfect water balance to reach the desired concrete strength. Furthermore, Figure 2(h) visualizes that the dataset has a specific time of curing period that can be said to be 7, 28, 56, and 90 days.

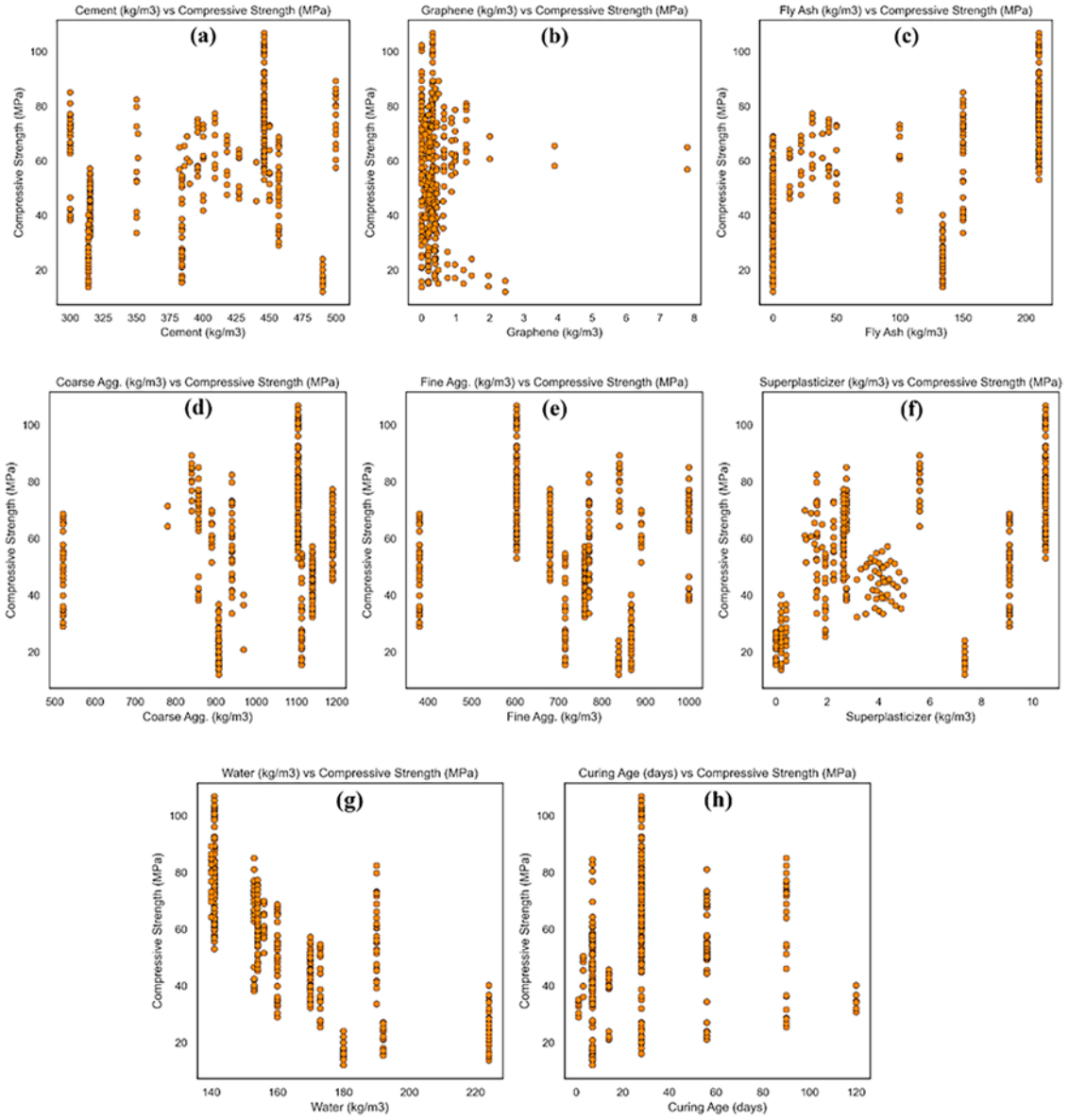

Figure 2: Distribution and association between all input variables with CS.

3.2. Correlation between features

The correlation coefficient varies between −1 and 1, with values exceeding 0 indicating a positive relationship and negative values reflecting the opposite trend. A higher absolute coefficient value suggests a stronger connection. Figure 3 illustrates the correlation coefficients between input and output variables, which range from −0.72 to +0.58. The heat map indicates that five input variables have a positive interaction with CS, with the exception of graphene, water content, and fine aggregate. The relationship between coarse aggregate and curing age with CS is relatively weak but positive (R = 0.14 and 0.12), suggesting a moderate level of consistency with strength. Fly ash, SP, and cement show high positive correlations with CS, with respective R values of 0.58, 0.53, and 0.37. Conversely, a very moderate negative correlation exists between graphene and CS, indicating that strength is negligibly reduced as the contents of graphene rise. Alongside this, a strong negative association is noted for water (R = −0.72).

3.3. Visual representation and errors of ML models

The graphical illustration presented in Figure 4 outlined the distribution of observed and predicted CS values across the six algorithms for both the training and testing phases. The residual scattering for all six ML models in the dataset is presented as well. Regarding the maximum error, kNN exhibited the highest values, recording 26.65 during training and 28.96 during testing, whereas DT and kNN demonstrated the lowest training maximum error, almost equal to zero. XGB, LGB, and GBR demonstrated comparable performance during testing, exhibiting maximum errors of 24.62, 24.68, and 24.55, respectively. To minimize error, DT, and KNN attained flawless zero values in both training and testing phases, whereas LGB exhibited the lowest balanced non-zero training and testing error at 0.002 and 0.058. Considering the average error rate, XGB and DT consistently demonstrated the lowest average error across both datasets, recording 2.54 and 2.52 in the training phase. In contrast, ADB showed the highest average errors, with 6.65 in training and 7.79 in testing. The analysis and illustrations demonstrate most of the model's effectiveness in achieving a balance of error during both the training and testing phases, although there are some difficulties related to overfitting and maintaining consistent error levels. In a research of SOBUZ *et al.* [54], they found the XGB to be an algorithm with superior performance and lower error rates compared to other models.

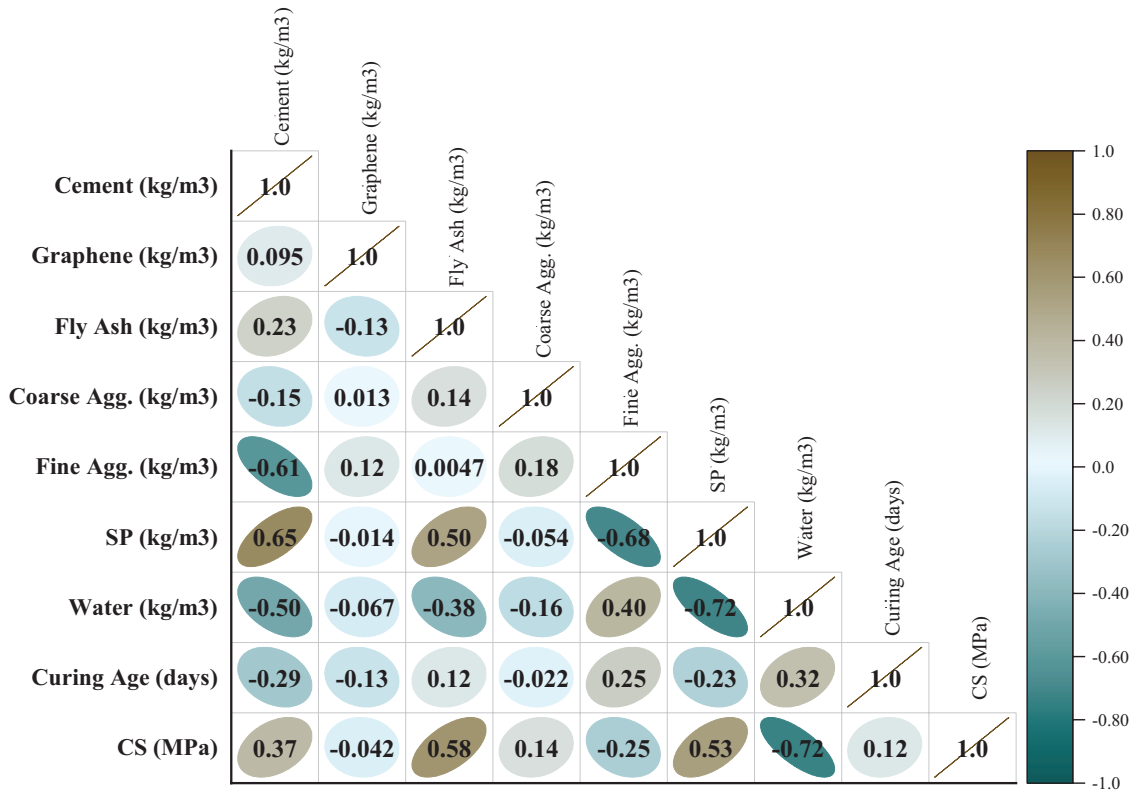

Figure 3: Correlation coefficient values between variables.

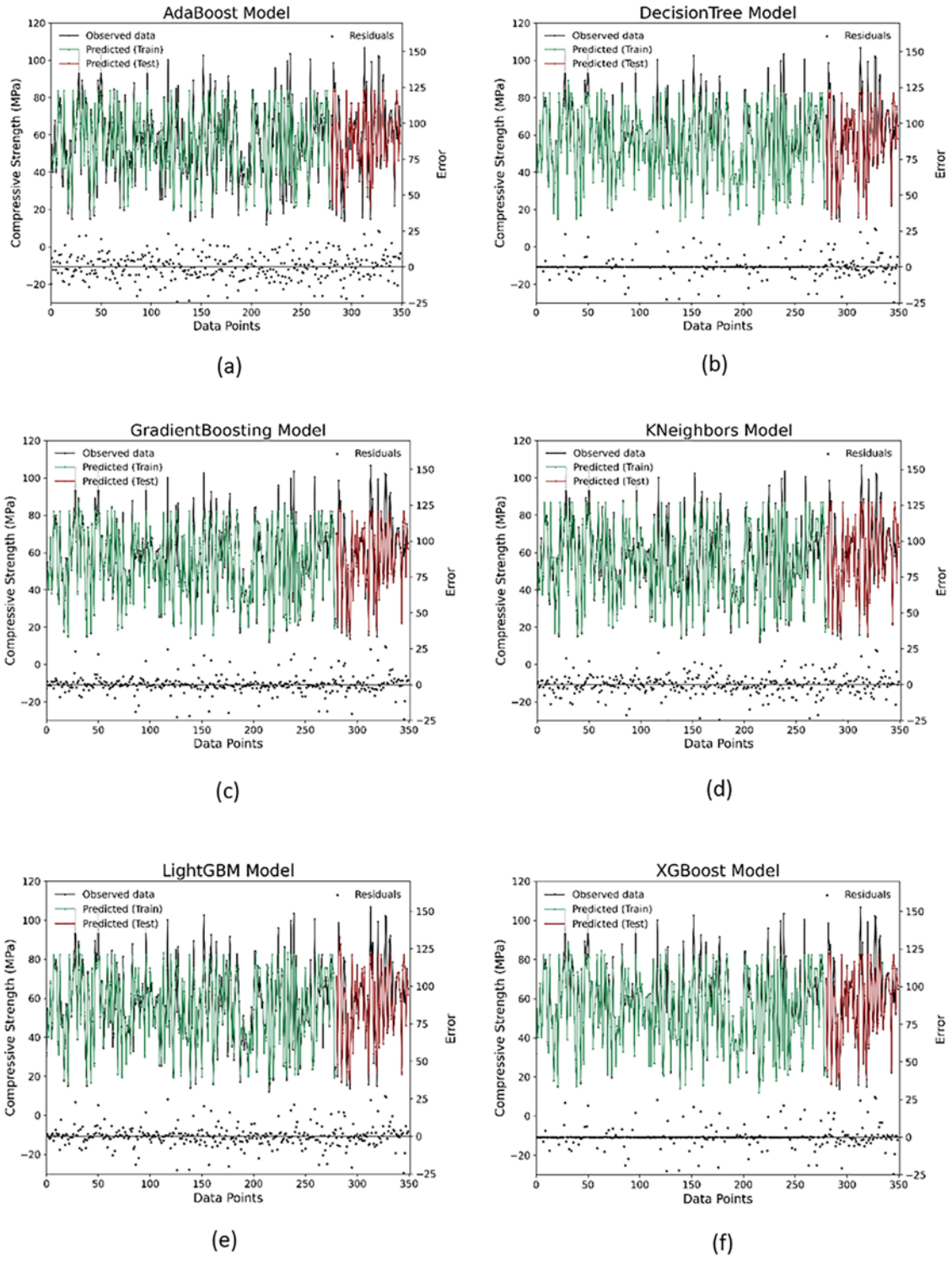

Figure 4: Error distribution and visualization of all model's outcomes.

3.4. Assessing the predictive accuracy of ML models

Figure 5 presents the linear regression of the ML models. The R^2 scores of the six ML algorithms exhibit varying performance patterns. Combinedly, XGB and DT demonstrate the maximum R^2 scores in training (0.921), suggesting the highest fit among all the models for the training data. However, XGB showed slightly better test R^2 scores of 0.87 comparing DT (R^2 of 0.863), revealing negligible potential issues with overfitting. GBR demonstrates a strong competitive edge over XGB and DT by achieving comparable R^2 scores in both training

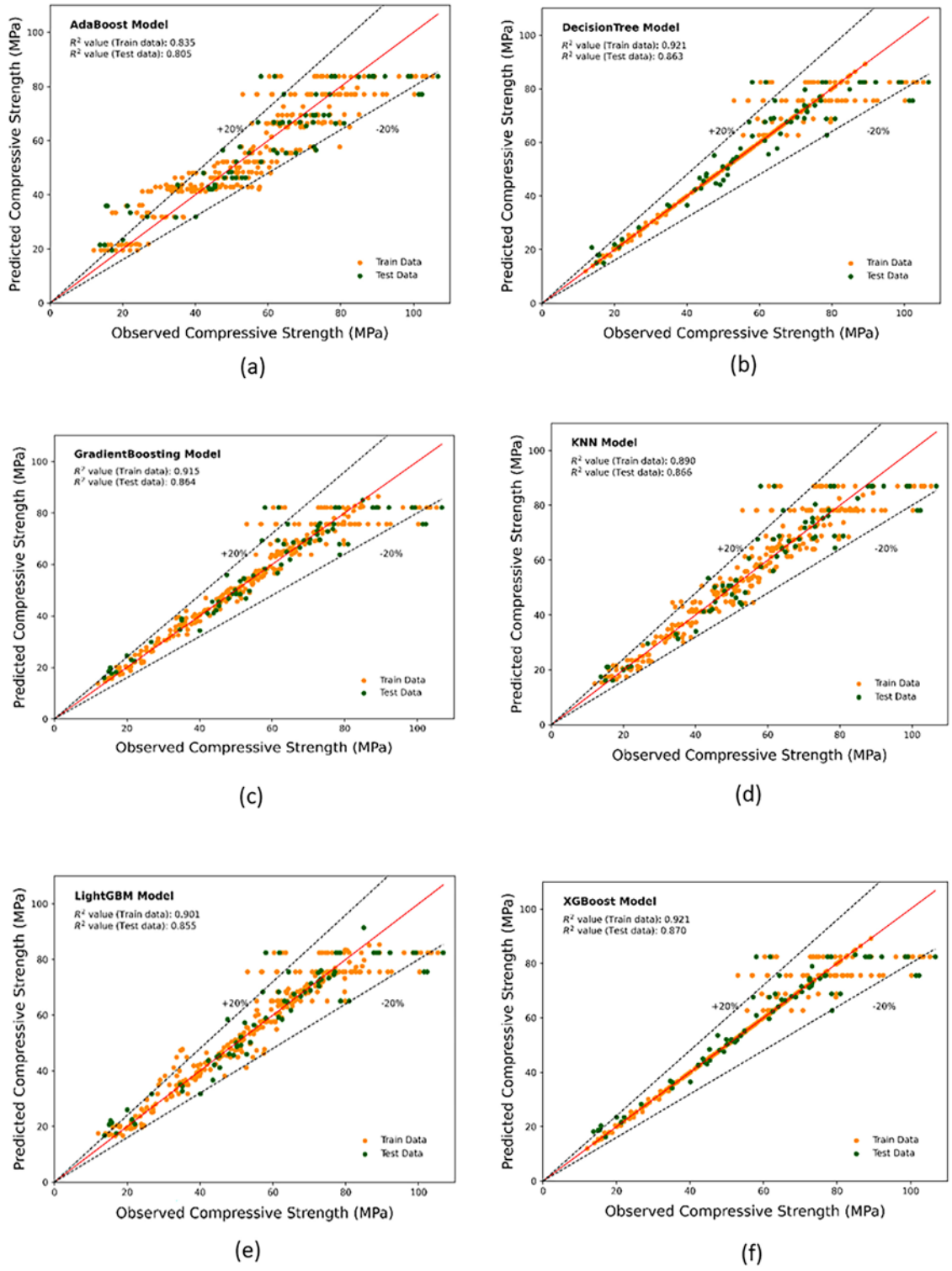

Figure 5: Linear regression plot of models.

and testing data, evidenced by a test R^2 of 0.863 and a training R^2 of 0.915. Following GBR, LGB also demonstrates strong performance, achieving an R^2 value of 0.90 during training and 0.855 during testing. In contrast, ADB exhibits suboptimal performance, reflected in its low R^2 scores of 0.835 in training and 0.804 in testing, which is also the lowest score among all six models used in this study. kNN demonstrates moderate R^2 scores, with 0.889 in training and 0.865 in testing, indicating a satisfactory level of predictive capability. Although the

kNN and ADB models have comparatively lower values of R^2, they can still be used to predict the CS of EGC. Finally, XGB emerged as the most effective model; the DT, GBR, and LGB models demonstrated commendable performance in prediction as well.

3.5. Performance evaluation criteria of ML models

Figure 6 compares the performance evaluation of the six models using R^2, MAE, MAPE and RMSE. With the lowest test RMSE of 8.30 MPa among the models tested, the XGB model stands out among the others, suggesting that its predictions are the most accurate. Its test MAPE of 8.30% indicates that the average percentage error is less than 10%, suggesting exceptional accuracy, and its test MAE of 4.93 MPa, which is the lowest, shows a small average absolute error in predictions. This phenomenon may arise from XGB's ability to capture complex patterns during training, which, without proper regularization, could lead to overfitting of the training data and reduced generalization during testing. Regularization techniques, such as adjusting the learning rate or increasing tree constraints, could help mitigate this issue.

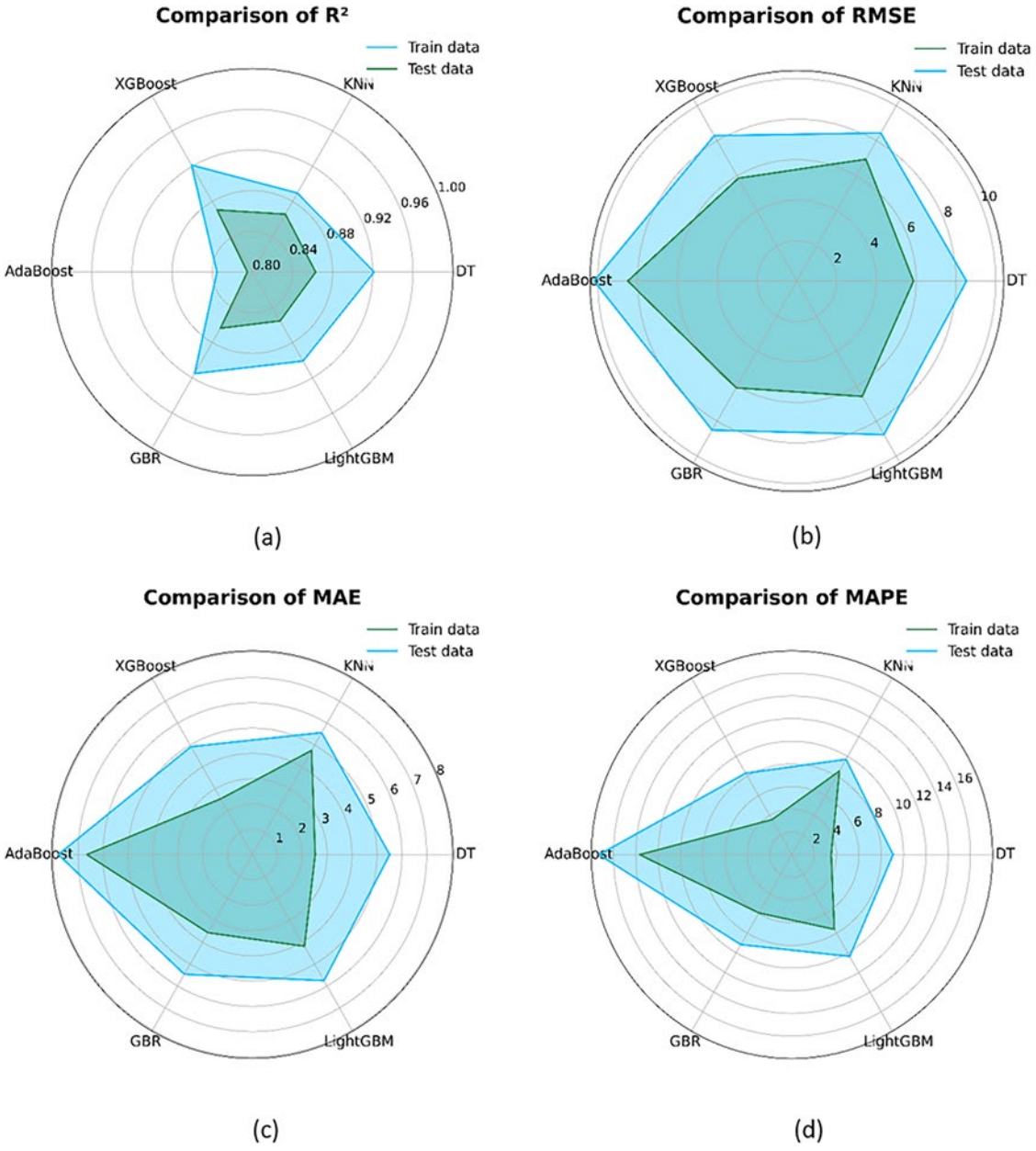

Figure 6: Comparison of ML model's performance based on several performance metrics.

With a test RMSE of 8.53 MPa, test MAE of 5.51 MPa, and test MAPE of 9.06%, DT comes in second. Despite being marginally less accurate than DT, GBR has a comparable error margin of XGB and DT. With test RMSE of 8.51 MPa, test MAE of 5.46 MPa, and test MAPE of 9.16%, the GBR model, on the other hand, has marginal error values and less significant scatter from actual values. This suggests that GBR effectively captures the data's underlying structure without overly complex patterns that compromise generalization. Its stability stems from iterative optimization and inherent regularization mechanisms. With test RMSEs of 8.77 MPa and 10.18 MPa, respectively, and MAEs of 5.75 MPa and 7.79 MPa, the LGB and ADB models perform less accurately, increasing error when compared to XGB, DT, or GBR alone.

These results imply that XGB and DT models are better at reducing prediction errors and, thus, more suited for precision. Furthermore, with a test RMSE of 8.45 MPa, test MAE of 5.57 MPa, and a high R^2 of 0.0.87, the kNN model performs moderately in comparison to XGB and DT, demonstrating dependable predictions with comparatively low errors. Overall, XGB has the least MAE, MAPE, and RMSE in both train and test stages, making it the most accurate.

In the Taylor diagram (Figure 7), six models are compared based on their standard deviation (SD) and correlation coefficient relative to a reference. The model with the lowest SD and highest correlation is closest to the reference, indicating better performance. All the models (except ADB) show similar performance, with standard deviations around 20 and correlation coefficients near 0.93, suggesting they capture the reference data pattern quite well. According to the tailor diagram, the kNN model is the closest to the reference point than the standalone other models. Furthermore, XGB, DT, GBR, and LGB closely followed one another and were positioned in almost similar places in the Taylor diagram (both in Figure 7(a) and 7(b). As can be seen, XGB showed maximum performance when compared to all the models used in this study. Except for ADB, the other four models also showed satisfactory performance following the XGB model, and they can be effectively used to predict the CS of EGC.

3.6. SHAP analysis results

The importance of input characteristics in the model's result projection is evaluated by the average SHAP values shown in Figure 8. The mean SHAP value of a feature is calculated by averaging its SHAP values throughout the whole dataset. The most important parameters that affect the CS of EGC are those with high mean SHAP values. Based on Figure 8, water content is the most impactful on CS, with a very high mean SHAP value of +14. It is followed by curing age (+5.13), fine aggregate (+4.37), graphene (+2.9), and cement (+2.01). Other parameters show a mean SHAP value lower than +2.0. It can be inferred that water, age of curing, and fine aggregate have a notable impact on the strength of EGC, while other ingredients have a comparatively lesser impact on the model prediction.

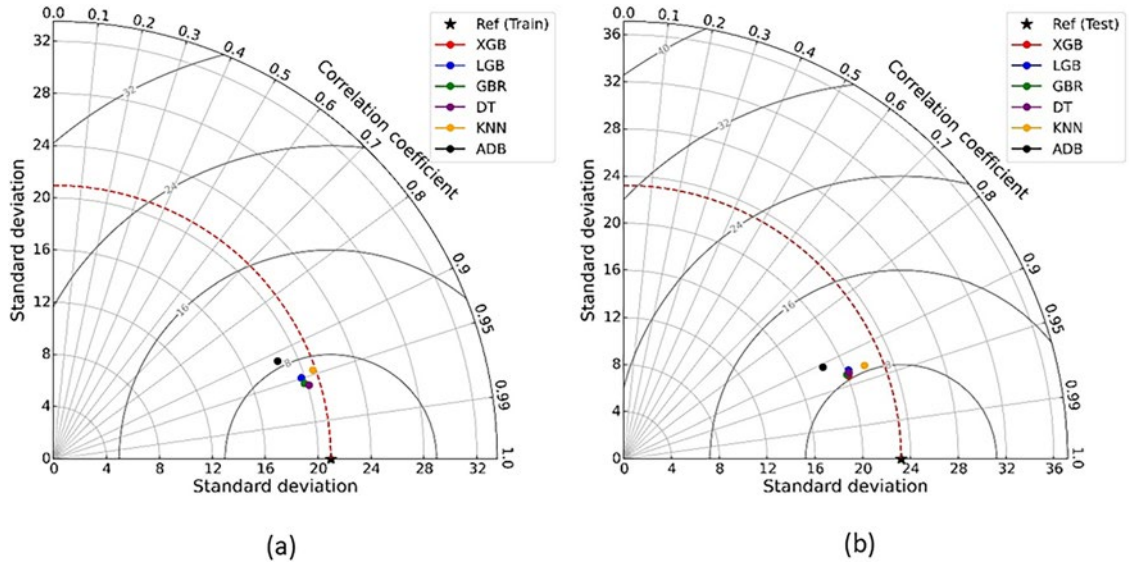

Figure 7: Taylor plot for (a) train and (b) test phase.

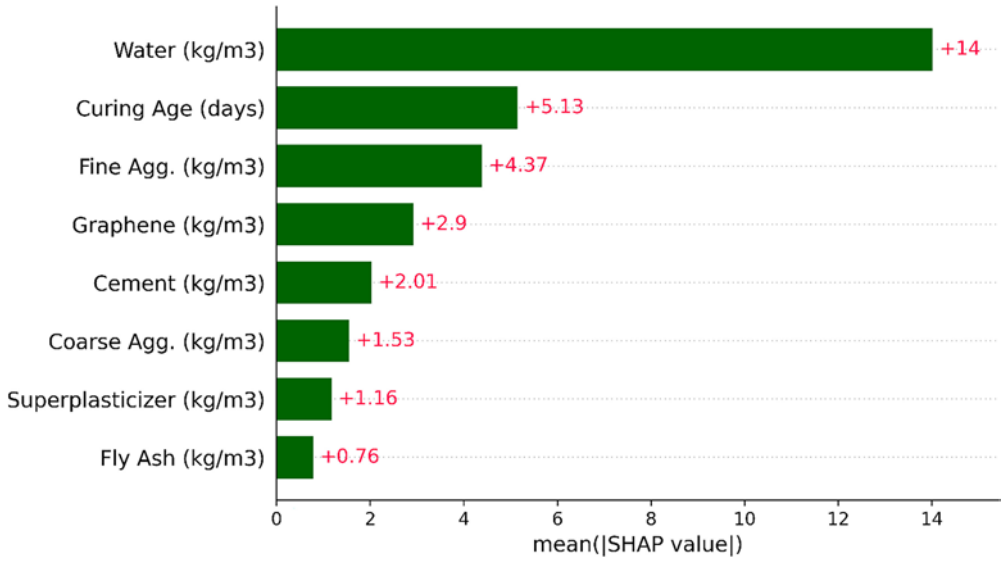

Figure 8: Illustration of SHAP values for all features.

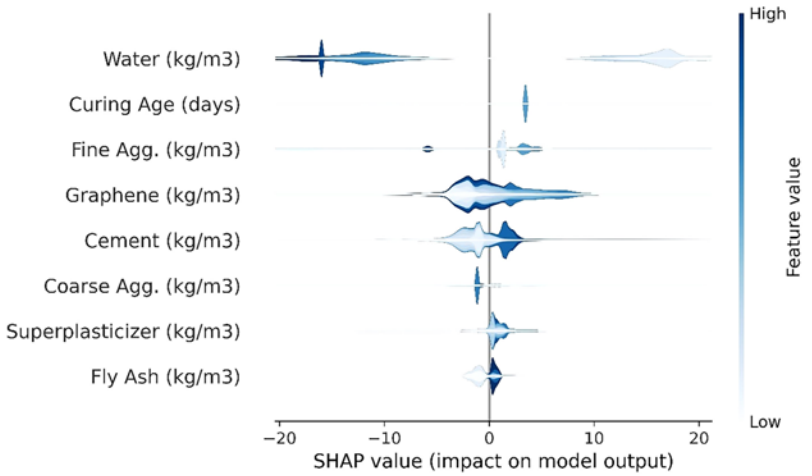

Figure 9: SHAP summary plots for EGC.

Figure 9 presents a SHAP feature plot, a powerful visualization tool for analyzing the influence of different parameters on the anticipation generated by the model. In the plot, the X-axis indicates a feature, while the perpendicular axis defines the SHAP values, reflecting the feature's influence on the model's output. The width of the violin plot at any point reflects the density of the SHAP values, with wider sections signifying a higher frequency of those values for the specific feature. Higher feature values, depending on the specific feature, have a more significant effect, either positive or negative. SHAP values imply the influence of individual parameters on the model's outcome, with SHAP values raising the prediction while the opposite values diminish it. For example, graphene, cement, SP, and fly ash show high feature values in positive SHAP value points, suggesting that higher amounts of these contents are associated with increased CS. This finding also aligns with experimental studies. For instance, a higher SP implies that there is less water concerning the amount of cement. As a result, this lower water leads to a denser cement paste when the concrete hardens, leading to higher CS in the concrete. A similar type of analysis was conducted by KARIM *et al.* [70], where it was seen that concrete age, aggregate content, and cement concrete are the most important parameters that influenced the CS of concrete incorporating RHA.

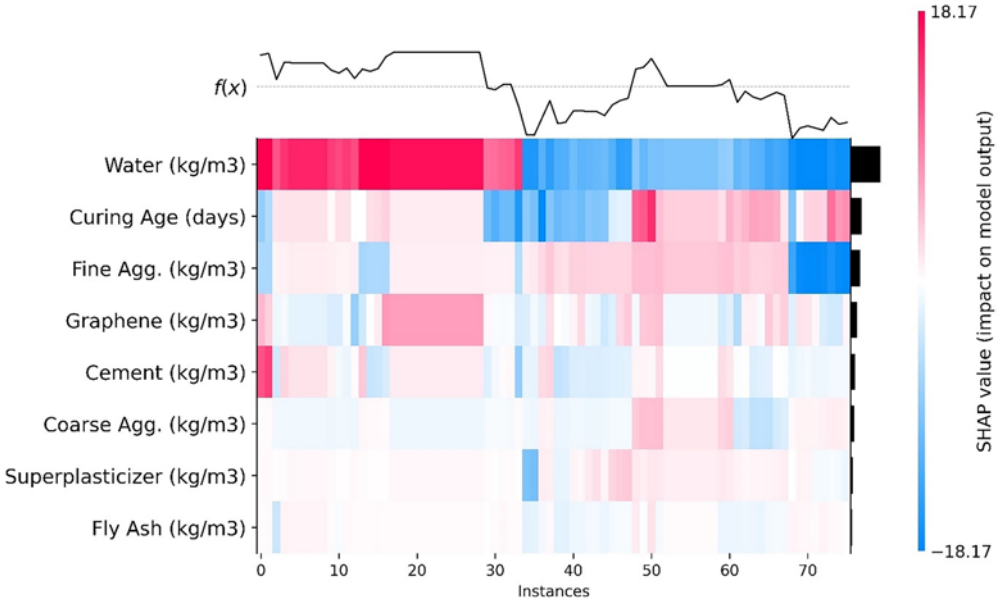

Figure 10: Distribution and association between all input variables with CS.

The SHAP heatmap in Figure 10 visualizes the impact of various features on each prediction instance, showing the detailed SHAP values across the dataset. Each row corresponds to a feature, and each column represents an instance. The color scale from blue to red indicates the SHAP value's magnitude and direction. According to the SHAP heatmap, Water and age of curing have a significant positive influence on SHAP magnitude, especially in the first 35 cases for water content. However, their effects are more mixed in the intermediate range. For curing duration, exceeding 45 instances showed notable positive SHAP magnitude comparing earlier instances. In the cases of fine aggregate, higher cases (>70) show a beneficial effect on model output. Graphene has positive SHAP values in most instances, where the negative feature values are so marginal that they can be neglected. Fly ash demonstrates neutral SHAP values in the whole bar of instances. This could be due to which could be the low amount of data that represents fly ash content in the dataset used in this analysis. However, for a clear visualization of these individual features' impact on the model's output, Figure 11 has been elaborately added to illustrate the individual SHAP value plot for all features.

3.7. PDP analysis

The impact of different input parameter content on the CS of EGC is depicted in Figure 12. The data values on both the X and Y axes in the figures demonstrate how input parameters affect CS. As can be seen in Figure 12(a), the most significant impact of graphene is observed at concentrations between 2 and 4 kg/m^3. Within this range, significant rises in CS are visible, along with around cement contents of 400–450 kg/m^3. These peaks are for the synergistic effect between graphene and cement, where the EG content enhances the hydration process and inner matrix of the concrete, which aggressively increases the CS of EGC. The highest CS values are observed within this region, exceeding 60 MPa in some combinations. From Figure 12(b), the replacement of cement with fly ash, up to a concentration of 200 kg/m^3, improves CS considerably. In terms of optimum concentration levels for fly ash, it appears to be around the range of 200 kg/m^3, as beyond this amount, the CS tends to fall. Additionally, coarse aggregate beyond 950 kg/m^3 visualizes a sudden fall of CS, which suggests that the design of EGC concrete mix should have a CA concentration lower than 950 kg/m^3.

Similarly, CS increases slowly and linearly with fine aggregate concentration up to 800 kg/m^3, then has a sudden decrease, declines linearly till 850 kg/m^3, and then stabilizes in a horizontal line for 49 MPa (Figure 12(c)). In the same figure, CS rises sharply between 52 and 58 MPa when the SP content varies between 0 to 10 kg/m^3. Furthermore, Figure 12(d) shows that higher water reduces the CS when incorporating more than 160 kg/m^3 although a longer curing age improves CS because it promotes better bonding through prolonged hydration. The findings derived from the 2D-PDP analysis provide critical insights for optimizing input feature concentrations to achieve maximum CS of EGC.

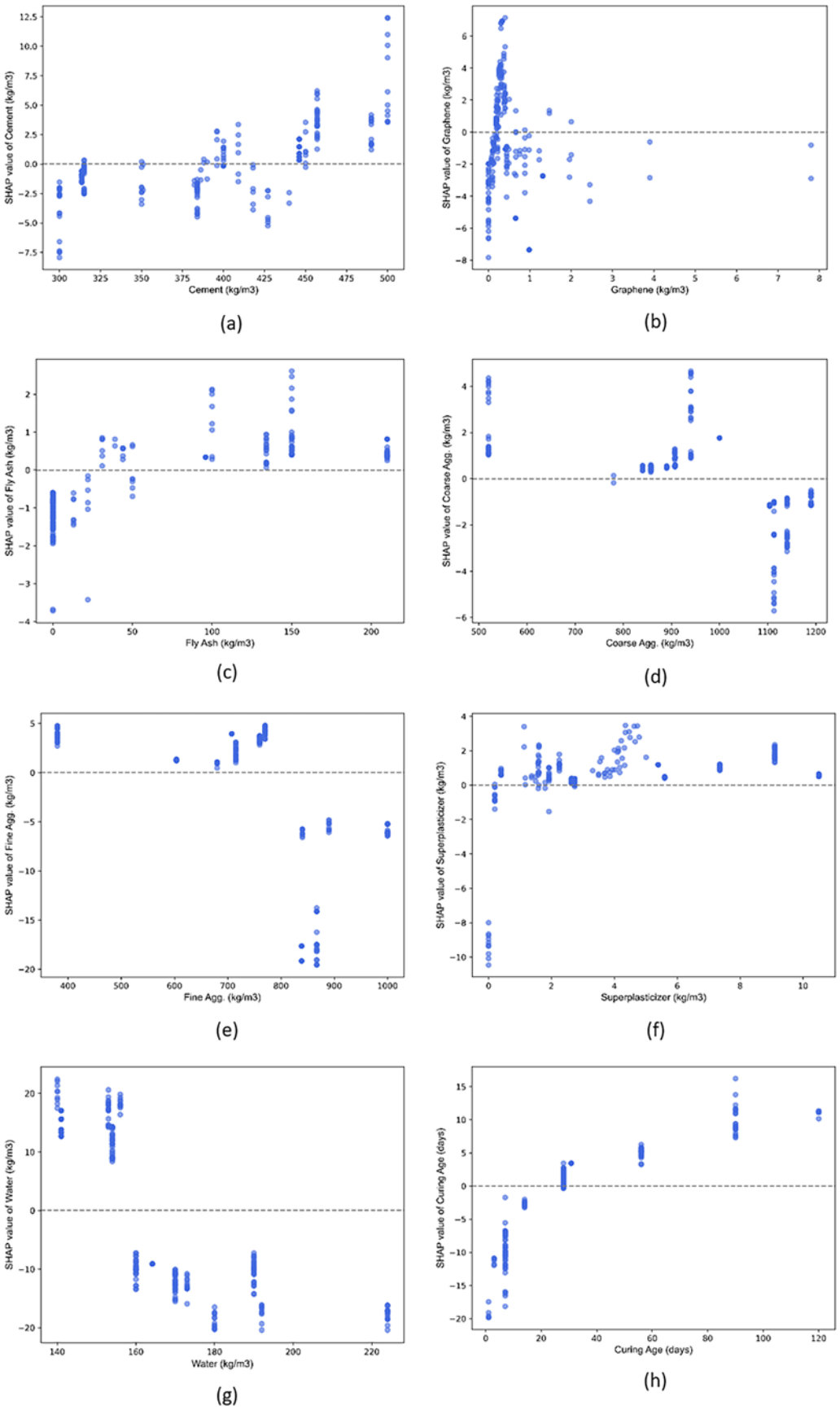

Figure 11: SHAP individual feature plots for the input parameters.

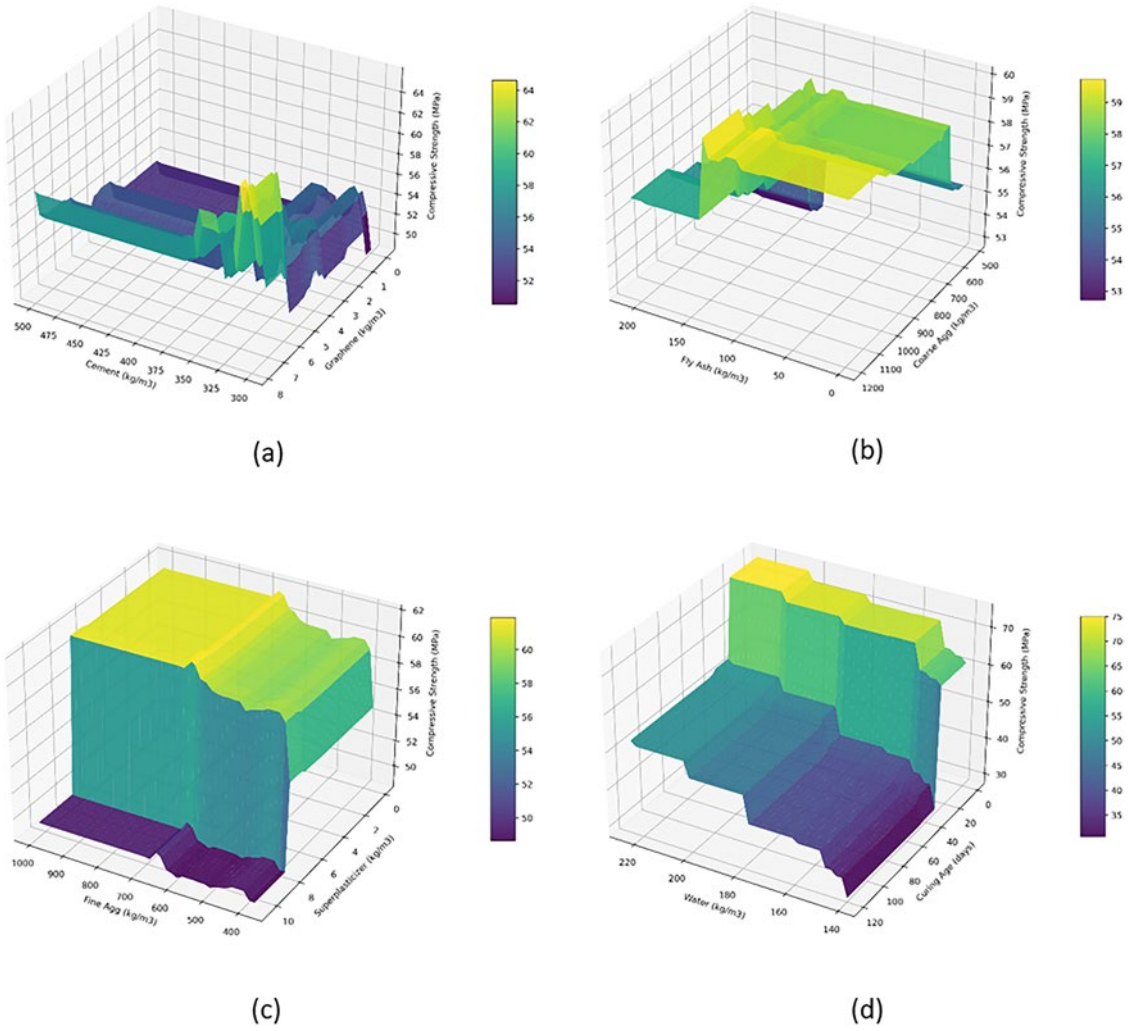

Figure 12: 2D PDP output of input variables.

4. CONCLUSIONS

This study successfully applied and compared six machine learning models—Adaptive Boosting, Decision Tree, Gradient Boosting, k-Nearest Neighbors, Light Gradient Boosting, and Extreme Gradient Boosting—to predict the compressive strength of engineered-graphene concrete using a comprehensive dataset. The most important findings are summarized as follows:

- All six machine learning models (ADB, DT, GBR, kNN, LGB, and XGB) demonstrated strong performance and high generalization capabilities. With the exception of ADB, the models achieved R^2 scores exceeding 0.85, signifying their reliability in accurately predicting concrete strength.
- The XGB model outperformed (R^2 of 0.863) other boosting and tree-based models, showcasing superior performance. Following XGB, DT, and GBR also exhibited balanced high predictive capability (R^2 of 0.863 and 0.864 for testing).
- The XGB model stands out among the other models since it has the lowest RMSE of 8.30 MPa, indicating its predictions are the most accurate. Second place goes to DT, which has a test RMSE of 8.53 MPa, a test MAE of 5.51, and a test MAPE of 9.06% among its metrics.
- SHAP analysis highlighted that water, curing age, fine aggregate, and graphene were the most influential factors affecting EGC. Additionally, the interaction between SP and fly ash had a moderate impact on the strength predictions. PDP analysis indicated that incorporating engineered graphene at a range of 0–2 kg/m^3 in the concrete mix led to the maximum increase in CS. The rise is from almost 53 MPa to 60 MPa, up to the EG content of 0.50 kg/m^3.

The findings indicate that the boosting-based algorithms are very efficient in forecasting the CS of graphene-modified concrete, resulting in decreased costs for materials and substantial time savings. Despite these promising results, some limitations should be acknowledged. The dataset's representation of extreme graphene dosages and certain material combinations was limited, potentially restricting the models' generalizability in those ranges. Moreover, the study did not include advanced regularization techniques or hybrid model structures to maintain computational simplicity and focus on boosting-based algorithms. Future research should incorporate more diverse and experimentally controlled datasets, explore rheological and durability properties, and investigate hybrid deep learning frameworks to further enhance prediction accuracy and generalizability.

5. ACKNOWLEDGMENTS

The analysis and modeling took place at KUET's Building Information Modeling (BIM) laboratory at the Department of BECM, Khulna, Bangladesh. All of the authors of this work would like to extend their heartfelt thanks to the lab employees who assisted them.

6. BIBLIOGRAPHY

[1] FAYED, S., MANSOUR, W., "Structural performance of sea sand recycled aggregate concrete filled Solid/Hollow aluminum tubular Columns: an experimental work", *Structures*, v. 47, pp. 1323–1340, 2023.

[2] YANG, L., GAO, Y., CHEN, H., *et al.*, "Three-dimensional concrete printing technology from a rheology perspective: a review", *Advances in Cement Research*, v. 36, n. 12, pp. 567–586, 2024. doi: http://doi.org/10.1680/jadcr.23.00205.

[3] SOBUZ, M.H.R., KHAN, M.H., KABBO, M.K.I., *et al.*, "Assessment of mechanical properties with machine learning modeling and durability, and microstructural characteristics of a biochar-cement mortar composite", *Construction & Building Materials*, v. 411, pp. 134281, 2024. doi: http://doi.org/10.1016/j.conbuildmat.2023.134281.

[4] FAYED, S., MADENCI, E., BAHRAMI, A., *et al.*, "Experimental study on using recycled polyethylene terephthalate and steel fibers for improving behavior of RC columns", *Case Studies in Construction Materials*, v. 19, pp. e02344, 2023. doi: http://doi.org/10.1016/j.cscm.2023.e02344.

[5] WANG, X., LI, L., WEI, M., *et al.*, "Experimental study on the mechanical properties of short-cut basalt fiber reinforced concrete under large eccentric compression", *Scientific Reports*, v. 15, n. 1, pp. 10845, 2025. doi: http://doi.org/10.1038/s41598-025-94964-5. PubMed PMID: 40155690.

[6] HUANG, H., HUANG, M., ZHANG, W., *et al.*, "Progressive collapse of multistory 3D reinforced concrete frame structures after the loss of an edge column", *Structure and Infrastructure Engineering*, v. 18, n. 2, pp. 249–265, 2022. doi: http://doi.org/10.1080/15732479.2020.1841245.

[7] CHEN, D., CHEN, Y., MA, L., *et al.*, "A state of review on manufacturing and effectiveness of ultra-high-performance fiber reinforced concrete for long-term integrity of concrete structures", *Advances in Concrete Construction*, v. 17, n. 5, pp. 293–310, 2024. doi: http://doi.org/10.12989/acc.2024.17.5.293.

[8] LI, D., NIE, J.-H., WANG, H., *et al.*, "Path planning and topology-aided acoustic emission damage localization in high-strength bolt connections of bridges", *Engineering Structures*, v. 332, pp. 120103, 2025. doi: http://doi.org/10.1016/j.engstruct.2025.120103.

[9] GAO, D., LI, Z., DING, C., *et al.*, "Uniaxial tensile stress-strain constitutive relationship of 3D/4D/5D steel fiber-reinforced concrete", *Construction & Building Materials*, v. 470, pp. 140539, 2025. doi: http://doi.org/10.1016/j.conbuildmat.2025.140539.

[10] HUANG, H., GUO, M., ZHANG, W., *et al.*, "Seismic behavior of strengthened RC columns under combined loadings", *Journal of Bridge Engineering*, v. 27, n. 6, pp. 05022005, 2022. doi: http://doi.org/10.1061/(ASCE)BE.1943-5592.0001871.

[11] RONG, C., PENG, Y., SHI, Q., *et al.*, "Eccentric compression performance of concrete filled steel tube slotted columns: experiment and simulation analysis", *Structures*, v. 74, pp. 108580, 2025.

[12] LEE, H.-P., CHAK, W.-F., TEOW, K.-L., *et al.*, "Investigation of catalyzed biomass thermoelectric concrete with palm oil fuel ash", In: *International Conference on Sustainable Civil Engineering Structures and Construction Materials*, Sarawak, Malaysia, 8-9 December 2020. Cham: Springer, pp. 451–463, 2020.

[13] WANG, J., ZHANG, Y., WANG, K., *et al.*, "Development of similar materials with different tension-compression ratios and evaluation of TBM excavation", *Bulletin of Engineering Geology and the Environment*, v. 83, n. 5, pp. 190, 2024. doi: http://doi.org/10.1007/s10064-024-03674-1.

[14] SOBUZ, M.H.R., AAYAZ, R., RAHMAN, S.A., *et al.*, "Assessment of hybrid fiber reinforced graphene nano-engineered concrete composites: From experimental testing to explainable machine learning modeling", *Journal of Materials Research and Technology*, v. 36, pp. 1409–1430, 2025. doi: http://doi.org/10.1016/j.jmrt.2025.03.003.

[15] LEE, H.-P., JUSLI, E., LING, J.H., *et al.*, "Concrete Paving Blocks incorporating palm oil boiler ash and palm oil clinker as substitute concrete materials", *Journal of Engineering Science and Technology*, v. 18, n. 6, pp. 75–81, 2023.

[16] AL-AJMAI, F.F., AL-OTAIBI, F.A., ALDAIHANI, H.M., *et al.*, "Effect of type of ground cover on the ground cooling potential for buildings in extreme desert climate", *Jordan Journal of Civil Engineering,* v. 12, n. 3, pp. 534–546, 2018.

[17] SHAHAB, Z., ANWAR, W., ALYAMI, M., *et al.*, "Experimental investigation and predictive modeling of compressive strength and electrical resistivity of graphene nanoplatelets modified concrete", *Materials Today. Communications*, v. 38, pp. 107639, 2024. doi: http://doi.org/10.1016/j.mtcomm.2023.107639.

[18] ZHANG, C., KHORSHIDI, H., NAJAFI, E., *et al.*, "Fresh, mechanical and microstructural properties of alkali-activated composites incorporating nanomaterials: a comprehensive review", *Journal of Cleaner Production*, v. 384, pp. 135390, 2023. doi: http://doi.org/10.1016/j.jclepro.2022.135390.

[19] MURAD, Y., "Compressive strength prediction for concrete modified with nanomaterials", *Case Studies in Construction Materials*, v. 15, pp. e00660, 2021. doi: http://doi.org/10.1016/j.cscm.2021.e00660.

[20] JIAO, H., WANG, Y., LI, L., *et al.*, "A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials", *Materials Today. Communications*, v. 35, pp. 106335, 2023. doi: http://doi.org/10.1016/j.mtcomm.2023.106335.

[21] ZEYAD, A.M., MAHMOUD, A.A., EL-SAYED, A.A., *et al.*, "Compressive strength of nano concrete materials under elevated temperatures using machine learning", *Scientific Reports*, v. 14, n. 1, pp. 24246, 2024. doi: http://doi.org/10.1038/s41598-024-73713-0. PubMed PMID: 39414873.

[22] MUGHEES, A., SHARMA, A., ANSARI, S.S., *et al.*, "Prediction of the compressive strength of nano-titanium based concrete composites using machine learning", *Materials Today: Proceedings*, 2023. In Press. doi: http://doi.org/10.1016/j.matpr.2023.03.540.

[23] REN, Z., ZENG, H., ZENG, X., *et al.*, "Effect of nanographite conductive concrete mixed with magnetite sand excited by different alkali activators and their combinations on the properties of conductive concrete", *Buildings*, v. 13, n. 7, pp. 1630, 2023. doi: http://doi.org/10.3390/buildings13071630.

[24] MOHSEN, M.O., AL-DISEET, M.M., ABURUMMAN, M.O., *et al.*, "Hybrid effect of GNPs, GOs, and CNTs on the flexural and compressive strengths of cement paste", *Journal of Building Engineering*, v. 73, pp. 106679, 2023. doi: http://doi.org/10.1016/j.jobe.2023.106679.

[25] LIU, Q., WU, W., XIAO, J., *et al.*, "Correlation between damage evolution and resistivity reaction of concrete in-filled with graphene nanoplatelets", *Construction & Building Materials*, v. 208, pp. 482–491, 2019. doi: http://doi.org/10.1016/j.conbuildmat.2019.03.036.

[26] CHEN, G., YANG, M., XU, L., *et al.*, "Graphene Nanoplatelets Impact on Concrete in improving freeze-thaw resistance", *Applied Sciences*, v. 9, n. 17, pp. 3582, 2019. doi: http://doi.org/10.3390/app9173582.

[27] SUN, S., HAN, B., JIANG, S., *et al.*, "Nano graphite platelets-enabled piezoresistive cementitious composites for structural health monitoring", *Construction & Building Materials*, v. 136, pp. 314–328, 2017. doi: http://doi.org/10.1016/j.conbuildmat.2017.01.006.

[28] DIMOV, D., AMIT, I., GORRIE, O., *et al.*, "Ultrahigh performance nanoengineered graphene–concrete composites for multifunctional applications", *Advanced Functional Materials*, v. 28, n. 23, pp. 1705183, 2018. doi: http://doi.org/10.1002/adfm.201705183.

[29] FAROOQ, F., AHMED, W., AKBAR, A., *et al.*, "Predictive modeling for sustainable high-performance concrete from industrial wastes: a comparison and optimization of models using ensemble learners", *Journal of Cleaner Production*, v. 292, pp. 126032, 2021. doi: http://doi.org/10.1016/j.jclepro.2021.126032.

[30] ZHENG, W., ZAMAN, A., FAROOQ, F., *et al.*, "Sustainable predictive model of concrete utilizing waste ingredient: individual alogrithms with optimized ensemble approaches", *Materials Today. Communications*, v. 35, pp. 105901, 2023. doi: http://doi.org/10.1016/j.mtcomm.2023.105901.

[31] LONG, X., MAO, M.-H., SU, T.-X., *et al.*, "Machine learning method to predict dynamic compressive response of concrete-like material at high strain rates", *Defence Technology*, v. 23, pp. 100–111, 2023. doi: http://doi.org/10.1016/j.dt.2022.02.003.

[32] NIU, Y., WANG, W., SU, Y., *et al.*, "Plastic damage prediction of concrete under compression based on deep learning", *Acta Mechanica*, v. 235, n. 1, pp. 255–266, 2024. doi: http://doi.org/10.1007/s00707-023-03743-8.

[33] MAGLAD, A.M., MANSOUR, W., TAYEH, B.A., *et al.*, "Experimental and analytical investigation of fracture characteristics of steel fiber-reinforced recycled aggregate concrete", *International Journal of Concrete Structures and Materials*, v. 17, n. 1, pp. 74, 2023. doi: http://doi.org/10.1186/s40069-023-00637-w.

[34] MANSOUR, W., FAYED, S., BASHA, A., "Experimental and numerical analysis of the punching behavior of RC isolated footings", *Steel and Composite Structures, An International Journal*, v. 45, n. 5, pp. 665–682, 2022.

[35] MANSOUR, W., LI, W., WANG, P., *et al.*, "Experimental and numerical evaluations of the shear performance of recycled aggregate RC beams strengthened using CFRP sheets", *Engineering Structures*, v. 301, pp. 117368, 2024. doi: http://doi.org/10.1016/j.engstruct.2023.117368.

[36] BASHA, A., TAYEH, B.A., MAGLAD, A.M., *et al.*, "Feasibility of improving shear performance of RC pile caps using various internal reinforcement configurations: tests and finite element modelling", *Engineering Structures*, v. 289, pp. 116340, 2023. doi: http://doi.org/10.1016/j.engstruct.2023.116340.

[37] SANKARAPANDIAN, K., ALSHAHRANI, H.M., ALOTAIBI, F.A., *et al.*, "Estimating punching performance in fiber-reinforced polymer concrete slabs utilizing machine learning and gradient-boosted regression techniques", *Matéria*, v. 30, pp. e20240668, 2025. doi: http://doi.org/10.1590/1517-7076-rmat-2024-0668.

[38] ALAHMARI, T.S., KABBO, M.K.I., SOBUZ, M.H.R., *et al.*, "Experimental assessment and hybrid machine learning-based feature importance analysis with the optimization of compressive strength of waste glass powder-modified concrete", *Materials Today. Communications*, v. 44, pp. 112081, 2025. doi: http://doi.org/10.1016/j.mtcomm.2025.112081.

[39] KABBO, M.K.I., SOBUZ, M.H.R., ADITTO, F.S., *et al.*, "Experimental assessment and machine learning quantification of structural eco-cellular lightweight concrete incorporating waste marble powder and silica fume", *Journal of Building Engineering*, v. 105, pp. 112557, 2025. doi: http://doi.org/10.1016/j.jobe.2025.112557.

[40] SOBUZ, M.H.R., KABBO, M.K.I., KHATUN, M., *et al.*, "Microstructural assessment and supervised machine learning-aided modeling to explore the potential of quartz powder as an alternate binding material in concrete", *Case Studies in Construction Materials*, v. 22, pp. e04568, 2025. doi: http://doi.org/10.1016/j.cscm.2025.e04568.

[41] CAVALERI, L., BARKHORDARI, M.S., REPAPIS, C.C., *et al.*, "Convolution-based ensemble learning algorithms to estimate the bond strength of the corroded reinforced concrete", *Construction & Building Materials*, v. 359, pp. 129504, 2022. doi: http://doi.org/10.1016/j.conbuildmat.2022.129504.

[42] CAVALERI, L., CHATZARAKIS, G.E., DI TRAPANI, F., *et al.*, "Modeling of surface roughness in electro-discharge machining using artificial neural networks", *Advanced Materials Research*, v. 6, n. 2, pp. 169, 2017.

[43] ALKAYEM, N.F., SHEN, L., MAYYA, A., *et al.*, "Prediction of concrete and FRC properties at high temperature using machine and deep learning: a review of recent advances and future perspectives", *Journal of Building Engineering*, v. 83, pp. 108369, 2024. doi: http://doi.org/10.1016/j.jobe.2023.108369.

[44] NGUYEN, N.-H., VO, T.P., LEE, S., *et al.*, "Heuristic algorithm-based semi-empirical formulas for estimating the compressive strength of the normal and high performance concrete", *Construction & Building Materials*, v. 304, pp. 124467, 2021. doi: http://doi.org/10.1016/j.conbuildmat.2021.124467.

[45] HASAN, N.M.S., SOBUZ, M.H.R., SHAURDHO, N.M.N., *et al.*, "Eco-friendly concrete incorporating palm oil fuel ash: fresh and mechanical properties with machine learning prediction, and sustainability assessment", *Heliyon*, v. 9, n. 11, pp. e22296, 2023. doi: http://doi.org/10.1016/j.heliyon.2023.e22296. PubMed PMID: 38045200.

[46] ASTERIS, P.G., MOKOS, V.G., "Concrete compressive strength using artificial neural networks", *Neural Computing & Applications*, v. 32, n. 15, pp. 11807–11826, 2020. doi: http://doi.org/10.1007/s00521-019-04663-2.

[47] ASHRAFIAN, A., PANAHI, E., SALEHI, S., *et al.*, "Mapping the strength of agro-ecological lightweight concrete containing oil palm by-product using artificial intelligence techniques", *Structures*, v. 48, pp. 1209–1229, 2023.

[48] BARKHORDARI, M.S., ARMAGHANI, D.J., MOHAMMED, A.S., *et al.*, "Data-driven compressive strength prediction of fly ash concrete using ensemble learner algorithms", *Buildings*, v. 12, n. 2, pp. 132, 2022. doi: http://doi.org/10.3390/buildings12020132.

[49] DUAN, J., ASTERIS, P.G., NGUYEN, H., *et al.*, "A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model", *Engineering with Computers*, v. 37, n. 4, pp. 3329–3346, 2021. doi: http://doi.org/10.1007/s00366-020-01003-0.

[50] JAF, D.K.I., ABDULRAHMAN, P.I., MOHAMMED, A.S., *et al.*, "Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO2) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete", *Construction & Building Materials*, v. 400, pp. 132604, 2023. doi: http://doi.org/10.1016/j.conbuildmat.2023.132604.

[51] SALAMI, B.A., OLAYIWOLA, T., OYEHAN, T.A., *et al.*, "Data-driven model for ternary-blend concrete compressive strength prediction using machine learning approach", *Construction & Building Materials*, v. 301, pp. 124152, 2021. doi: http://doi.org/10.1016/j.conbuildmat.2021.124152.

[52] KALOOP, M.R., KUMAR, D., SAMUI, P., *et al.*, "Compressive strength prediction of high-performance concrete using gradient tree boosting machine", *Construction & Building Materials*, v. 264, pp. 120198, 2020. doi: http://doi.org/10.1016/j.conbuildmat.2020.120198.

[53] MANGALATHU, S., JEON, J.-S., "Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques", *Engineering Structures*, v. 160, pp. 85–94, 2018. doi: http://doi.org/10.1016/j.engstruct.2018.01.008.

[54] SOBUZ, M.H.R., KHATUN, M., KABBO, M.K.I., *et al.*, "An explainable machine learning model for encompassing the mechanical strength of polymer-modified concrete", *Asian Journal of Civil Engineering,* v. 26, pp. 931–954, 2024. doi: https://doi.org/10.1007/s42107-024-01230-6.

[55] MARZEC, I., BOBIŃSKI, J., TEJCHMAN, J., *et al.*, "Finite element analysis on failure of reinforced concrete corner in sewage tank under opening bending moment", *Engineering Structures*, v. 228, pp. 111506, 2021. doi: http://doi.org/10.1016/j.engstruct.2020.111506.

[56] UDDIN, M.N., YE, J., DENG, B., *et al.*, "Interpretable machine learning for predicting the strength of 3D printed fiber-reinforced concrete (3DP-FRC)", *Journal of Building Engineering*, v. 72, pp. 106648, 2023. doi: http://doi.org/10.1016/j.jobe.2023.106648.

[57] GONG, B., LI, H., "A couple Voronoi-RBSM modeling strategy for RC structures", *Structural Engineering and Mechanics, An Int'l Journal*, v. 91, n. 3, pp. 239–250, 2024.

[58] WANG, K., CAO, J., YE, J., *et al.*, "Discrete element analysis of geosynthetic-reinforced pile-supported embankments", *Construction & Building Materials*, v. 449, pp. 138448, 2024. doi: http://doi.org/10.1016/j.conbuildmat.2024.138448.

[59] LI, D., CHEN, Q., WANG, H., *et al.*, "Deep learning-based acoustic emission data clustering for crack evaluation of welded joints in field bridges", *Automation in Construction*, v. 165, pp. 105540, 2024. doi: http://doi.org/10.1016/j.autcon.2024.105540.

[60] SOBUZ, M.H.R., KABBO, M.K.I., ALAHMARI, T.S., *et al.*, "Microstructural behavior and explainable machine learning aided mechanical strength prediction and optimization of recycled glass-based solid waste concrete", *Case Studies in Construction Materials*, v. 22, pp. e04305, 2025. doi: http://doi.org/10.1016/j.cscm.2025.e04305.

[61] FREUND, Y., SCHAPIRE, R.E., "Experiments with a new boosting algorithm", In: *Machine Learning: Proceedings of the Thirteenth InternationalConference*, Bari, Italy, 3-6 July 1996.

[62] GUO, L., GE, P.-S., ZHANG, M.-H., *et al.*, "Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine", *Expert Systems with Applications*, v. 39, n. 4, pp. 4274–4286, 2012. doi: http://doi.org/10.1016/j.eswa.2011.09.106.

[63] KANG, M.-C., YOO, D.-Y., GUPTA, R., "Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete", *Construction & Building Materials*, v. 266, pp. 121117, 2021. doi: http://doi.org/10.1016/j.conbuildmat.2020.121117.

[64] COVER, T., "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, v. 4, n. 5, pp. 515–516, 1968.

[65] XI, B., LI, E., FISSHA, Y., *et al.*, "LGBM-based modeling scenarios to compressive strength of recycled aggregate concrete with SHAP analysis", *Mechanics of Advanced Materials and Structures*, v. 31, n. 23, pp. 5999–6014, 2024. doi: http://doi.org/10.1080/15376494.2023.2224782.

[66] ALABDULLAH, A.A., IQBAL, M., ZAHID, M., *et al.*, "Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis", *Construction & Building Materials*, v. 345, pp. 128296, 2022. doi: http://doi.org/10.1016/j.conbuildmat.2022.128296.

[67] CHEN, T., GUESTRIN, C., "Xgboost: a scalable tree boosting system", In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, 13-17 August 2016.

[68] FRIEDMAN, J.H., "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, v. 29, n. 5, pp. 1189-1232, 2001. doi: http://doi.org/10.1214/aos/1013203451.

[69] MERRICK, L., TALY, A., "The explanation game: explaining machine learning models using shapley values", In: *Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference*, Dublin, 25-28, August 2020. Cham: Springer, pp. 17–38, 2020. doi: http://doi.org/10.1007/978-3-030-57321-8_2.

[70] KARIM, R., ISLAM, M.H., DATTA, S.D., *et al.*, "Synergistic effects of supplementary cementitious materials and compressive strength prediction of concrete using machine learning algorithms with SHAP and PDP analyses", *Case Studies in Construction Materials*, v. 20, pp. e02828, 2024. doi: http://doi.org/10.1016/j.cscm.2023.e02828.