



OPEN Predicting the mechanical performance of industrial waste incorporated sustainable concrete using hybrid machine learning modeling and parametric analyses

Md. Alhaz Uddin¹, Md. Habibur Rahman Sobuz²✉, Md. Kawsarul Islam Kabbo²,
Md. Kanan Chowdhury Tilak², Ratan Lal³, Md. Selim Reza⁴, Fahad Alsharari¹,
Mohamed AbdelMongy¹ & Masuk Abdullah⁵✉

The construction sector is proactively working to minimize the environmental impact of cement manufacturing by adopting alternative cementitious substances and cutting carbon emissions tied to concrete. This study investigates the viability of using waste industrial materials as a replacement of cement in concrete mixes. The primary goal is to predict the compressive strength of waste-incorporated concrete by evaluating the effects of materials such as cement, fly ash (FA), silica fume (SF), ground granulated blast furnace slag (GGBFS), metakaolin (MK), water usage, aggregate levels, and superplasticizer dosages. A total of 441 data entries were sourced from various publications. Multiple machine learning techniques, such as light gradient boosting (LGB), extreme gradient boosting (XGB), and decision trees (DT), along with hybrid approaches like XGB-LGB and XGB-DT, were utilized to study how these variables influence compressive strength. The dataset was partitioned into training and testing, and statistical tools were employed to assess the correlation between input variables and strength. Model accuracy was gauged using metrics such as mean absolute percentage error (MAPE), root mean square error (RMSE), and the coefficient of determination (R^2). Among the models, the XGB and DT approach delivered the highest precision, with an R^2 of 0.928 in the training stage. Among hybrid models, XGB-DT exhibited a balanced performance having R^2 value of 0.907 and 0.785 for training and testing phase. Additionally, SHAP (SHapley Additive exPlanations) and partial dependence plots (PDP) were employed to pinpoint the optimal ranges for each variable's contribution to the improvement of compressive strength. SHAP and PDP analyses identified coarse aggregate, superplasticizers, water and cement content have high influence on model's output. Additionally, 150–200 kg/m³ of GGBFS as key factors for optimizing compressive strength. The study concludes that the hybrid models along with the single models, can effectively forecast the compressive strength of concrete incorporating industrial byproducts, assisting the construction industry in efficiently evaluating material properties and understanding the influence of various input factors.

Keywords Sustainable concrete, Hybrid machine learning, Compressive strength, Industrial waste, Parametric analysis.

Concrete stands as one of the most extensively used construction materials in the construction sector. Accelerated economic advancement and urban growth have caused a notable surge in new construction endeavors, along with the deconstruction of older infrastructure^{1,2}. Cement serves as a fundamental component in the creation of

¹Department of Civil Engineering, College of Engineering, Jouf University, Sakaka 72388, Saudi Arabia. ²Department of Building Engineering and Construction Management, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh. ³HJ Russell & Company, 171 17th St NW #1600, Atlanta, GA 30363, USA. ⁴Department of Software Engineering, Faculty of Science and Information Technology, Daffodil Smart City (DSC) Birulia, Daffodil International University, Savar, Dhaka 1216, Bangladesh. ⁵Department of Vehicles Engineering, University of Debrecen, Óttemető str. 2-4, Debrecen 4028, Hungary. ✉email: habib@becm.kuet.ac.bd; masuk@eng.unideb.hu

concrete. Nevertheless, the manufacturing industries of binders, more specifically cement, emit vast quantities of carbon dioxide, contributing to environmental degradation^{3–6}. The cement sector accounts for approximately 8% of global yearly CO₂ emissions linked to fossil fuel consumption^{7–9}. Consequently, decreasing cement usage in concrete production is an urgent priority for the worldwide construction field in this century¹⁰. Implementing renewable and eco-friendly energy alternatives is vital for reducing carbon emissions in the cement production process. Furthermore, the depletion of natural resources and the growing deposits of industrial waste have prompted recent studies into using by-products in concrete mixtures¹¹. Incorporating these discarded materials in concrete design is a contemporary approach to advancing sustainable resource practices and mitigating environmental harm, including CO₂ emissions¹². Combatting carbon release and reducing dependency on natural resources present major global issues that must be overcome to enable eco-production and ensure a sustainable path forward^{13,14}.

In contrast, the escalating amount and complexity of industrial refuse present major difficulties for proper disposal and recycling efforts, further complicated by a lack of available disposal locations and stricter environmental mandates. Global awareness of sustainable waste handling necessitates swift action in developing and applying efficient waste treatment and disposal strategies^{15,16}. Failing to manage natural resources effectively today could lead to harmful effects for future populations, stressing the urgency of improving recycling processes, finding material alternatives, and making better use of resources to minimize environmental harm¹⁷. The constant industrial and economic expansion worldwide has substantially boosted waste production, which correlates with the increasing demand for construction projects. This demand has led to the need for vast amounts of building materials, with concrete being the most extensively employed construction material across the globe¹⁸. Consequently, cement production has surged to nearly 4.1 billion tons each year, with predictions forecasting an increase to 5.8 billion tons by 2050. Addressing the ecological effects of this growth calls for both sustainable waste management practices and innovative approaches to conserving resources while ensuring responsible material production^{19–23}.

Numerous methods have been developed to mitigate the detrimental effects of cement production and the environmental impacts of concrete manufacturing²⁴. Increasing the incorporation of industrial waste as supplementary cementitious materials (SCMs) and other alternatives to cement clinker presents a valuable chance to lower global CO₂ emissions and conserve natural resources^{25–28}. To cut carbon emissions by one billion tonnes annually, it is necessary to substitute roughly 50% of the clinker used in cement production with alternative materials. This means integrating approximately 1.6 billion tonnes of industrial byproducts into cement production each year^{29,30}. Reaching these emission reduction goals demands a multi-pronged strategy, including reducing cement content in concrete mixtures, decreasing clinker consumption in cement manufacturing, and minimizing the use of concrete in new construction projects^{31,32}. Embracing these approaches will allow the cement industry to significantly lessen its environmental footprint while advancing sustainability initiatives. This includes technological innovations such as partially replacing cement with SCMs such as fly ash (FA) or ground granulated blast furnace slag (GGBFS), among others^{33–35}. Additionally, the adoption of alternative additives with a smaller environmental impact is being explored to lessen the ecological damages caused by cement production^{36,37}. These actions aim to enhance sustainability without compromising the performance quality of concrete^{38–44}.

Evaluating the characteristics of concrete incorporating industrial byproducts typically requires extensive laboratory experimentation, which is not only expensive but also time-consuming⁴⁵. Moreover, logistical challenges, such as limited space for curing and storing concrete specimens, often complicate thorough experimentation⁴⁶. Concrete's compressive strength is determined by numerous parameters, like the duration of curing, the materials incorporated in the mixture, and the type of aggregates employed, all of which necessitate detailed analysis that may prolong the testing process^{47–49}. Although several empirical equations have been proposed to estimate compressive strength and improve testing efficiency, these methods are sometimes unable to determine the complex, nonlinear interactions among concrete properties and their corresponding strength^{50–52}. As a result, more advanced models are required to accurately associate input variables with the performance of concrete made using industrial waste. In this context, Artificial Intelligence (AI) is demonstrated as an effective tool because of its ability to quickly learn from data, model intricate relationships, and predict results based on various input parameters^{53,54}. Several studies have highlighted the success of advanced machine learning (ML) approaches in forecasting the mechanical attributes of different categories of concrete, underscoring the potential of AI for improving the predictive capability^{55–57}.

To predict the fresh and hardened properties of self-compacting concrete, Support Vector Regression (SVR) with the Exponential Radial Basis Function (ERBF) kernel produced a high level of accuracy ($R > 0.95$)^{58–62}. The RF model performed exceptionally well when it came to predicting the strength of recycled aggregate concrete ($R^2 = 0.97$), with cement being identified as the input that had the greatest impact (32.5%)⁶³. When it came to the prediction of the flexural strength of steel fiber-reinforced concrete (SFRC), GBR achieved the maximum precision ($R^2 = 0.96$), outperforming both Random Forest ($R^2 = 0.94$) and Extreme Gradient Boosting ($R^2 = 0.86$)^{64,65}. With R^2 values of 0.92 (slump) and 0.93 (compressive strength), Multi-Expression Programming (MEP) was shown to be superior to Gene Expression Programming (GEP) when it came to alkali-activated concrete (AAC). The results of this research demonstrate the potential of machine learning models to perform accurate, efficient, and cost-effective analysis of concrete properties. Additionally, Golafshani et al.⁶⁶ used the XGBoost approach in conjunction with the multi-objective grey wolf optimizer to estimate the elastic modulus of geopolymer concrete, and achieve high accuracy in predicting concrete parameters. This would reduce the need for experimental testing and help with the optimization of materials⁶⁷. In addition, gene expression programming (GEP) was applied alongside artificial neural networks (ANN) to evaluate the compressive strength of geopolymer concrete containing GGBFS⁶⁸. In a separate study, Gupta et al.⁶⁹ implemented a multilayer perceptron (MLP) and ANN to assess the compressive strength of polymer-based composites containing FA and GGBFS. The MLP model

delivered a highly accurate prediction, with an R^2 of 0.968, effectively capturing the impact of key factors like the silica modulus, water-to-binder ratios, and curing conditions^{70,71}. Khan and Abbas⁷² constructed a multilayer stacked ML model to assess the compressive strength of reactive powder concrete (RPC), establishing strong links between compressive strength and variables such as SF content, fiber proportions, and concrete age. The stacking algorithm proved the most accurate, securing an R^2 of 0.96, followed by XGBoost at 0.954, Random Forest and ETR at 0.95, and KNN at 0.77. Similarly, Alabdullah et al.⁷³ contrasted LightGBM and XGBoost models for forecasting Rapid Chloride Penetration Test (RCPT) outcomes. Their findings indicated that LightGBM outperformed XGBoost with an R^2 of 0.9738, and SHAP analysis highlighted that compressive strength, aging, and water-binder ratio were key contributors to improving chloride resistance, with optimum performance occurring at a w/b ratio of 0.30–0.35 and 15% metakaolin replacement. Additionally, Jagadesh et al.⁷⁴ examined the 28-day compressive strength of self-compacting concrete (SCC) using recycled aggregates, utilizing a dataset of 515 mix designs with various machine learning methods. Their evaluation revealed that the Category Boosting (CB), K Nearest Neighbors (KNN), and Extra Trees (ERT) models produced the highest R^2 values and the lowest mean squared error (MSE), underscoring their effectiveness in accurately mapping input-output relationships^{75–79}.

There are two significant gaps in existing research regarding ML techniques for forecasting the compressive strength of concrete made with industrial waste: the necessity for more advanced and accurate hybrid ML models and the scarcity of detailed parametric evaluations, such as SHAP and partial dependence plot (PDP) analyses⁸⁰. Moreover, previous studies have generally relied on relatively small, limited datasets. Farooq et al.⁸¹ found that ensemble methods like boosting and AdaBoost were more effective than individual models. Shen et al.⁸² highlighted XGB as highly precise in forecasting the compressive strength of ultra-high-strength concrete. Alhakeem et al.⁸³ demonstrated that gradient-boosting regression trees (GBRT) were highly reliable for evaluating the compressive strength of sustainable concrete. Uddin et al.⁸⁴ observed that LightGBM offered greater accuracy compared to ML models such as SVR, DT and RF algorithms for evaluating the compressive strength of 3D-printed concrete. Amin et al.⁸⁵ showed that XGBoost regressor have higher precision than SVM in assessing the compressive strength of nano-materials and fiber incorporated concrete. Khan et al.⁸⁶ found that both RF and XGB achieved R^2 values exceeding 0.90 when predicting the compressive strength of steel fiber-reinforced composites. Song et al.⁸⁷ concluded that boosting methods were more accurate than neural networks and DT in estimating the compressive strength of concrete containing FA. Rahman et al.⁸⁸ applied eleven ML algorithms to forecast the shear capacity in concrete beams, with XGB delivering the best results, evidenced by an RMSE of 1.346 and an MAE of 0.704.

This research aimed to address gaps in the current literature by applying both traditional and hybrid ML techniques, including XGB, LGB, DT, XGB-LGB, and XGB-DT models, to assess the compressive strength of concrete incorporating industrial waste. A dataset of 430 data points from prior research was compiled alongside nine key input variables, which notably improved the accuracy of the compressive strength predictions. The models' performance was cross-validated by comparing anticipated results with actual experimental values, utilizing several statistical metrics during the training, testing, and validation stages. Furthermore, SHAP and PDP analyses were performed to explore the effect of various factors on compressive strength prediction. These analytical techniques offered valuable insights, aiding in the creation of optimized concrete mix designs with fewer trials, ultimately minimizing the need for extensive testing and reducing time, material use, labor, and associated costs.

Methodology

Development of ML models

Extreme gradient boosting (XGB) technique

XGB is an improved variation of the gradient boosting method originally developed by Friedman⁸⁹, utilizing multiple additive functions to improve predictive accuracy and efficiency⁹⁰. This advanced tree-boosting technique refines gradient boosting to better handle complex datasets using-

$$\bar{p}_i = p_i^0 + \eta \sum_{r=1}^N f_r(Z_i) \quad (1)$$

In the XGB model, the predicted outcome ($\bar{p}_i$) for the i -th data point depends on its corresponding feature set

Z_i . The model utilizes N estimators, with each estimator f_r (where r ranges from 1 to N) representing a separate decision tree structure. Initially, p_i^0 is set to the average of the observed values in the training data, acting as the baseline prediction. The learning rate g , also known as the shrinkage factor, progressively adjusts the model as new trees are introduced, helping to minimize the chances of overfitting, a challenge faced by most machine learning models.

The training process is iterative; at each r -th step, a new estimator is incorporated to refine the prediction from the prior step, $\bar{p}_i^{(r-1)}$, with the updated prediction $\bar{p}_i$ derived from the previous result and the additional estimator f_r . Overfitting is a crucial concern, and minimizing it is key to the model's performance as

$$\bar{p}_i^{(k)} = \bar{p}_i^{(r-1)} + \eta f_r \quad (2)$$

In the XGB model, each estimator f_r is defined by optimizing the leaf weights in the r -th tree through minimization of the objective function defined by

$$obj = \gamma E + \sum_{h=1}^E \left[G_h \omega_h + \frac{1}{2} (H_h + \lambda) \omega_h^2 \right] \quad (3)$$

In the XGB model, E denotes the number of leaves in the r -th tree, ω_j with (for $h=1$ to E) representing the weights of these leaves. The regularization parameters r and λ help maintain a simpler tree structure to prevent overfitting. The parameters G_h and H_h are the sums of the initial and second-order derivatives of the loss function for the data points linked to the h -th leaf, aiding in balancing model complexity and predictive accuracy.

The r -th tree is generated by dividing a single leaf into multiple leaves. This process focuses on optimizing the gain parameter, which is described as

$$gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (4)$$

In the model, G_L and H_L correspond to the left leaf, while G_R and H_R relate to the right leaf after a split occurs. A split is considered valid if the gain parameter is more significant than 0. Elevating the regularization parameters λ and γ , lowers the gain parameter, which contributes to keeping the tree structure simpler by limiting the complexity of leaf splits. However, this adjustment may also lower the model's predictive capacity to accurately fit the training dataset.

Light gradient boosting machines (LGB) technique

The LGB technique is a well-established approach in machine learning that utilizes gradient boosting combined with decision tree methodologies^{91,92}. 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^{91,92}. 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^{73,93}. The ultimate result from the LGB method is achieved by averaging the forecasts from each decision tree, with information gain defined as the expected reduction in entropy stemming from node splits based on particular attributes, which can be represented as follows:

$$IG(M, U) = F_n(M) - \sum_{W \in Values(W)} \frac{|M_w|}{M} F_n(M_w) \quad (5)$$

$$F_n(M) = \sum_{c=1}^B -pc \log_2 pc \quad (6)$$

In this context, (M) denotes the overall entropy information for set M , whereas pc represents the fraction of M that is categorized as class c . The variable B signifies the different classes, M_w indicates the subset of M where the feature is equal to w , and w refers to the specific value of the attribute W .

Decision tree (DT) algorithm

The first model introduced is the DT technique, which is well-regarded in the literature for its adaptability in capturing intricate nonlinear relationships and its strong interpretability^{94,95}. The DT method segments data into smaller groups, creating a tree-like format. This structure comprises three primary categories of nodes: (i) the root node (RN), representing the topmost point, (ii) decision nodes (DN), where conditional evaluations take place to determine how data is divided into additional subtrees based on the outcomes, and (iii) leaf nodes (LN), which function as terminal nodes that signify the final output. To assess the uniformity of the dataset, DT algorithms calculate two types of entropy measures⁹⁶.

$$E(D) = \sum_{k=1}^c -P_k \log_2 P_k \quad (7)$$

Where P_k represents the fraction of dataset D that is categorized as class k .

$$E(N, Z) = \sum_{c \in X} P(c) E(c) \quad (8)$$

In this context, N indicates the target characteristic, Z corresponds to the decision characteristic, c signifies the values associated with characteristic Z , $P(c)$ represents the chance of c occurring, and $E(c)$ denotes the entropy linked to c .

$$Gain(N, Z) = E(N) - E(N, Z) \quad (9)$$

where N stands for the target attribute, Z signifies the decision attribute, and, c pertains to the values associated with the tuples for that attribute $E(N)$ represents the calculated entropy of the target attribute, while $E(N, Z)$ refers to the entropy of the tuples of Z within the N attribute⁹⁷.

The hybrid ML technique

This research introduced a combined ML framework that integrates various conventional ML techniques and synthesizes the outputs from XGB, DT, and LGB algorithms. This led to the creation of two ensemble frameworks: XGB-LGB and XGB-DT, which were developed using stacking methodologies. The decision to employ hybrid techniques, specifically XGB-LGB and XGB-DT hybrid models, was driven by their demonstrated ability to enhance predictive accuracy and robustness by combining the capability of individual algorithms. Ensemble algorithms reduce the risk of overfitting and improve generalization by leveraging the complementary learning capabilities of base models. For instance, XGBoost is effective in handling complex non-linear relationships, while Decision Trees offer interpretability and low computational cost, and LightGBM is known for its high efficiency and scalability. The integration of these models thus results in a synergistic framework capable of capturing a wider range of data patterns.

Hyperparameter tuning

The performance of the algorithms is significantly affected by hyperparameter tuning, and the selection of optimal hyperparameters is essential for the development of effective ML algorithms. The GridSearchCV class from the scikit-learn library was utilized in this research to automate the adjustment of hyperparameters in these machine learning models. Hyperparameters play a crucial role in the development of machine learning algorithms with high accuracy. GridSearchCV is an efficient and reliable technique for fine-tuning hyperparameters, although it can be time-consuming. This study examined the hyperparameters for machine learning algorithms used in this study, as presented in Table 1.

Parametric analysis

Shapley additive explanations (SHAP)

SHAP interprets a model’s predictions by expressing them as a linear combination of input features, facilitating a clear insight into how each variable influences the final outcome^{98–100}. This methodology is grounded in game theory, specifically the Shapley values, which assist in identifying the contribution of each participant in a collaborative setting⁹¹. In the realm of machine learning, the input features act as these participants, while the model’s outputs represent the overall “game”. SHAP determines the significance of each feature by observing variations in predictions when specific features are included or excluded from the input set, thereby providing a comprehensive scenario of how each feature affects the model’s results.

z(t') = \varphi_o + \sum_{j=1}^M (\varphi_j X'_j) \tag{10}

Here, the Shapley values \varphi_j function as the coefficients in this linear setup, where t' represents a generalized version of the original input vector t, and z stands as the interpretative model. These coefficients provide insights into the contributions of individual features to the overall prediction.

\varphi_j = \sum_{G \subseteq H \setminus \{j\}} \frac{|G|! (|H| - |G| - 1)!}{|H|!} [f_{SU\{j\}}(X_{SU\{j\}}) - f_s(X_s)] \tag{11}

Algorithm	Parameter	Grid search range	Optimum
XGBoost	Number of estimators	[100, 200, 400]	100
	Learning rate	[0.3, 0.1, 0.01]	0.1
	Maximum depth	[3, 4, 6]	4
	Lambda	[0.1, 1, 10]	1
	Gamma	[0.001, 0.1, 1]	1
	Sample leaf	[2, 5, 10]	5
Light GBM	Number of estimators	[100, 200, 500]	200
	Learning rate	[0.2, 0.1, 0.01]	0.1
	Maximum depth	[4, 5, 6]	4
	Number of leaves	[5, 25, 30]	30
	Boosting types	[gbdt, rf, dart]	gbdt
Decision tree	Max depth	[5, 10, 20]	10
	Min samples split	[2, 20, 30]	20
	Min samples leaf	[3, 5, 10]	5
	Maximum features	[sqrt, log2]	sqrt
	Criterion	[squared_error, friedman_mse]	squared_error
	Splitter	[best, random]	best
	Max leaf nodes	[10, 20, 50]	50

Table 1. Hyperparameters for ML models used in this study.

In this case, H represents the full collection of input features, and G refers to a subset derived from H , excluding the feature associated with index j .

Partial dependence plots (PDP) method

Single-dimensional Partial Dependence Plots (PDPs-1D) are highly effective for depicting how one particular input variable correlates with the predicted result. In general, PDPs shed light on the significance of each input factor in influencing the model's predictions. As a holistic tool, they analyze all instances to expose the broader link between a feature and the projected outcome¹⁰¹. These visualizations illustrate how individual input factors shape the predictions by showing the influence of varying their values on the machine-learning model¹⁰². Additionally, PDPs uncover the relationship between specific input features and the forecasted outcomes, enhancing the understanding of each feature's role in shaping the final prediction.

Evaluation metrics for prediction performance

The prediction accuracy of the algorithms was assessed using three distinct statistical methods employed at three stages: training, validation, and testing. These statistical performance determinants are root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Additionally, the coefficient of determination (R^2) is used to evaluate how well the prediction coincides with the experimental outcome. RMSE helped to calculate the typical deviation between actual and predicted values, while MAPE assessed the percentage discrepancy between them.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Q_i - R_i)^2} \quad (12)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{Q_i - R_i}{R_i} \right| \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (Q_i - R_i)^2}{\sum_{i=1}^m (Q_i - \mu_i)^2} \quad (14)$$

In this context, Q_i indicates the observed values, R_i signifies the predicted values, μ_i represents the average, and m corresponds to the total data points. Figure 1 illustrates the overall methodological flowchart used in this study.

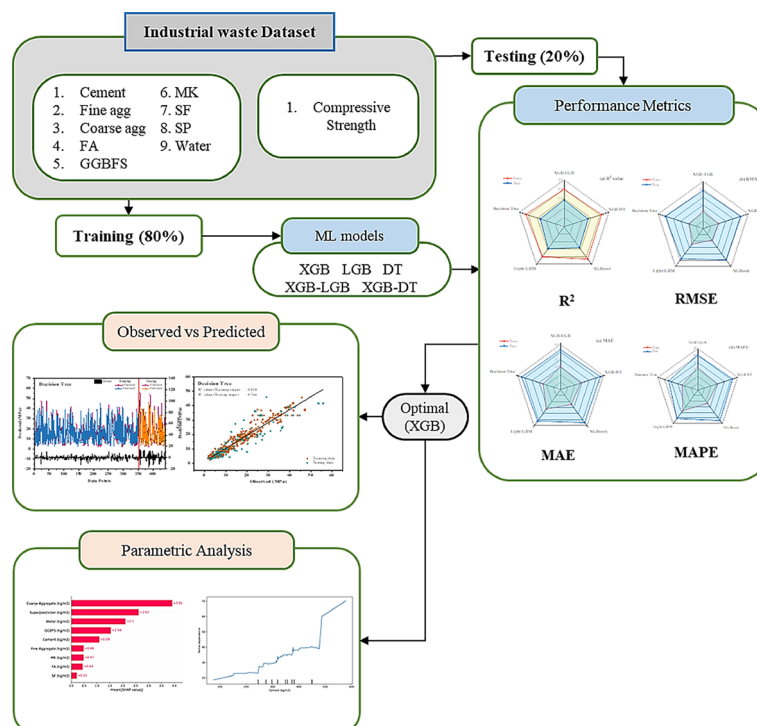


Fig. 1. Methodological flow chart.

Results and discussion
Analysis of dataset

In this research, a detailed dataset consisting of 441 records was compiled from various sources in the literature. The concrete mixtures studied involved the use of multiple industrial by-products, such as FA, SF, GGBFS, and MK, in combination with ordinary Portland cement, aggregates, SP, and water. The critical input variables encompassed the quantities of cement, FA, SF, GGBFS, MK, curing time, water amount, aggregate volume, and superplasticizer concentration. Table 2; Fig. 2 combined offer a clear overview of the input and output data, as well as their statistical characteristics. The table outlines various statistical metrics for each parameter, such as minimum and maximum values, standard deviation, and coefficient of variation (COV).

Figure 2 illustrates the association between compressive strength values and various concrete mix components. Figure 2(a), cement content shows a high concentration between 300 and 400 kg/m³. For coarse aggregate, as shown in Fig. 2(b), the data is predominantly clustered between 1200 and 2000 kg/m³, while fine aggregate data in Fig. 2(c) is primarily between 0 and 300 kg/m³. The relationship between compressive strength and industrial waste content, as shown in Fig. 2(d-g), is more complex, reflecting variability in the proportions of materials like FA, SF, and GGBFS from different studies. In Fig. 2(h), superplasticizer (SP) shows minimal variation, with around two-thirds of the data reporting quantities below 6 kg/m³. Lastly, Fig. 2(i) indicates that the water content associated with compressive strength is generally between 100 and 150 kg/m³.

Pearson correlation matrix

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 -0.57 to +0.59. The heat map indicates that six input variables have a positive interaction with compressive strength, with the exception of water content and aggregate. The relationship between cement and compressive strength is relatively weak but positive ($R=0.15$), suggesting a moderate level of consistency with strength. Metakaolin (MK), GGBFS, fine aggregate, water, and superplasticizer (SP) also show positive correlations with compressive strength, with respective R values of 0.17, 0.33, 0.16, 0.16, and 0.44. Conversely, a moderate negative correlation exists between SF and FA concerning compressive strength, indicating that strength diminishes as the contents of FA and SF rise. A similar negative association is noted for coarse aggregate.

XGB model performance assessment

Figure 4 showcases the outcomes of applying the XGB algorithm to forecast compressive strength. In Fig. 4(a), a combination of a line graph and scatter plot demonstrates the correlation between predicted and actual values for both the training and testing datasets. The prediction error for the training dataset remains below 6 MPa, while for the testing data, the error margin is approximately 20 MPa. Figure 4(b) highlights the relationship between actual and predicted values, showing R^2 values of 0.928 for the training data and 0.777 for the testing data. This indicates a high prediction accuracy for the training dataset and moderate accuracy for the testing set. The XGB model demonstrated high accuracy in predicting the compressive strength of waste marble powder concrete ($R^2>0.97$), indicating its reliability for optimizing WMP concentration in concrete mixtures¹⁰³.

Evaluation of the performance of the LGB model

Figure 5(a) displays the findings from the training and testing stages of the LGB model. The highest error observed is around 15 MPa in the training stage and about 20 MPa in the testing stage. The error range during the training phase is broader for the LGB model when compared to the testing data of the XGB model, although the testing errors for both models are similar. In Fig. 5(b), the R^2 values for the LGB algorithm are 0.895 for training and 0.79 for testing. The LGB model demonstrated consistent performance in predicting the CS of Recycled Aggregate Concrete, attaining an impressive Coefficient of Determination (R^2) value of 0.88, underscores its potential as a powerful tool in sustainable construction analysis¹⁰⁴. Overall, the XGB model shows a better alignment between predicted and actual data than the LGB model.

Input materials	Unit	Minimum	Maximum	Mean	Deviation	COV
Cement	kg/m ³	90	1370	300.55	194.53	64.73
Coarse Aggregate	kg/m ³	506	2649	1487.70	312.01	20.97
Fine Aggregate	kg/m ³	0	467	110.45	141.54	128.15
FA	kg/m ³	0	247	39.7	51.39	129.45
GGBFS	kg/m ³	0	247	11.42	30.38	266.03
MK	kg/m ³	0	420.1	17.74	51.87	292.39
SF	kg/m ³	0	120	17.05	36.72	215.35
Superplasticizers	kg/m ³	0	6.3	1.21	1.74	143.92
Water	kg/m ³	78	230.05	137.13	30.43	22.19
Compressive strength	MPa	1.5	55.73	16.43	11.09	67.51

Table 2. Descriptive analysis of the input concrete ingredients.

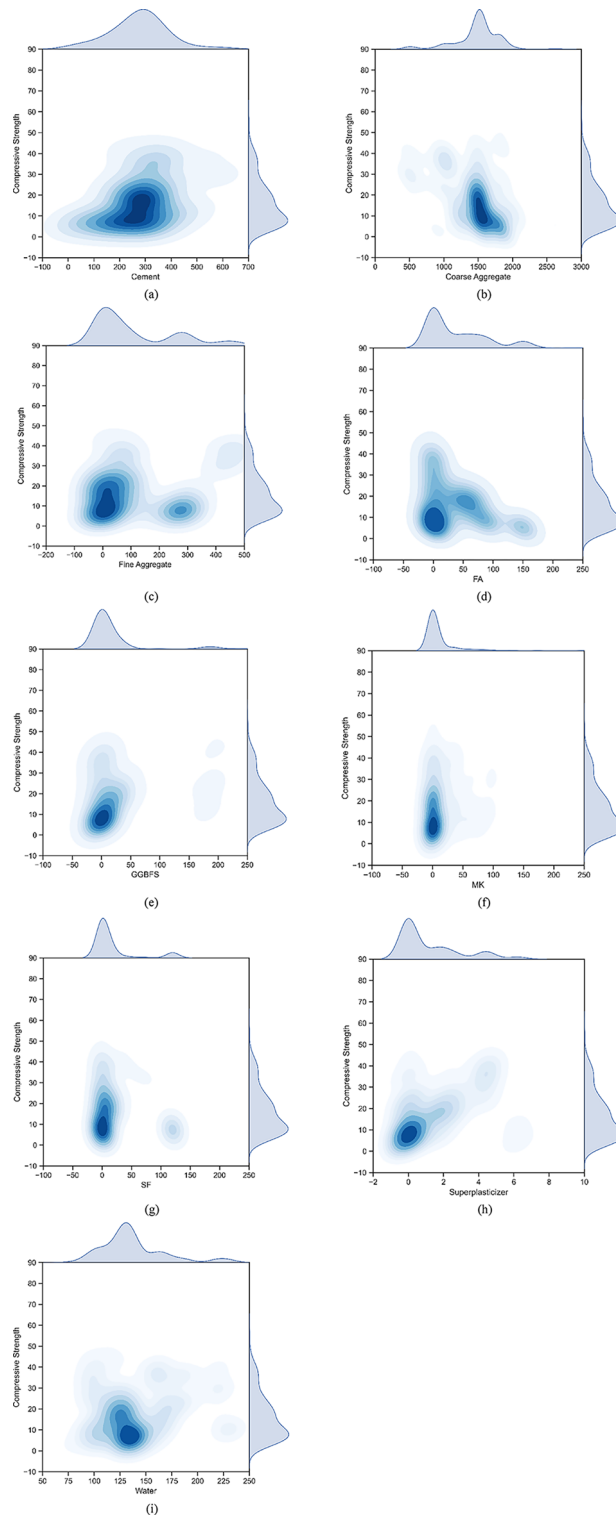


Fig. 2. Contour map of input and output variables.

Evaluation of the performance of the DT model

Figure 6(a) displays the outcomes from the training and testing phases for the Decision Tree (DT) model, indicating a maximum error of approximately 20 MPa for both datasets. The training phase for the DT model exhibits a higher error compared to the XGB and LGB models, while the testing errors remain consistent across all models. Figure 6(b) presents the R^2 values for the DT model, which are 0.928 for the training set and 0.766 for the testing set. The DT model, which achieved an impressive average accuracy of 88.32%, demonstrated its effectiveness as a dependable and potent instrument for predicting construction waste, paving the way for more

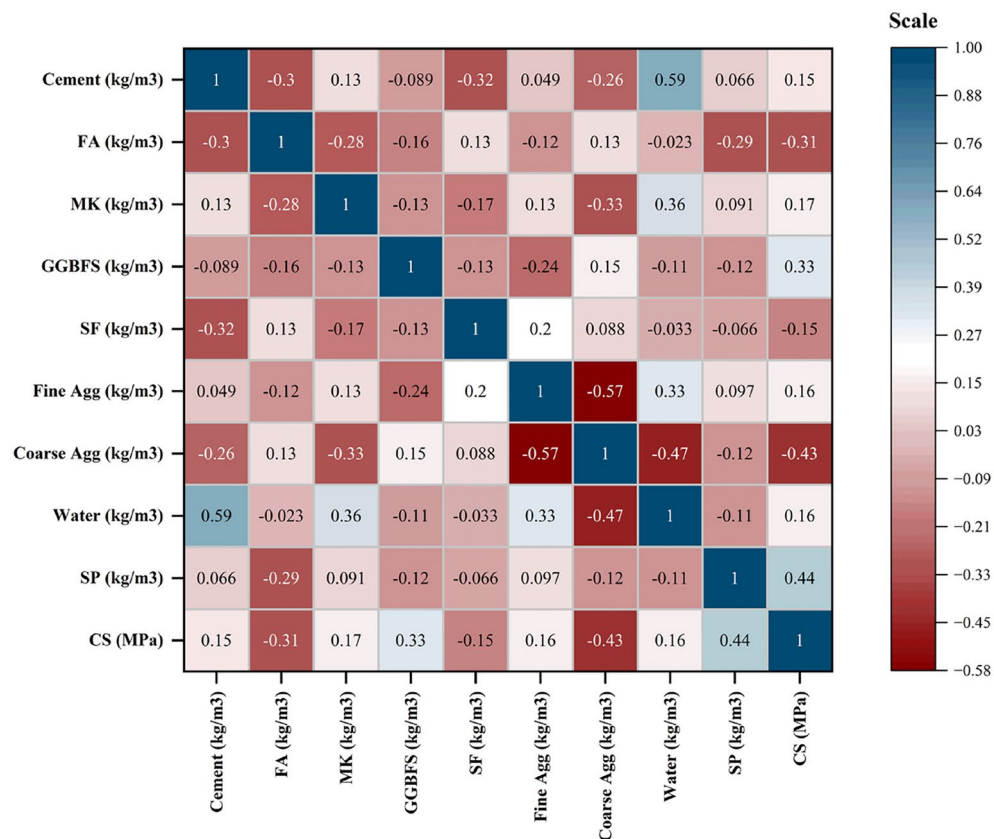


Fig. 3. Matrix depicting the correlation between input and target parameters.

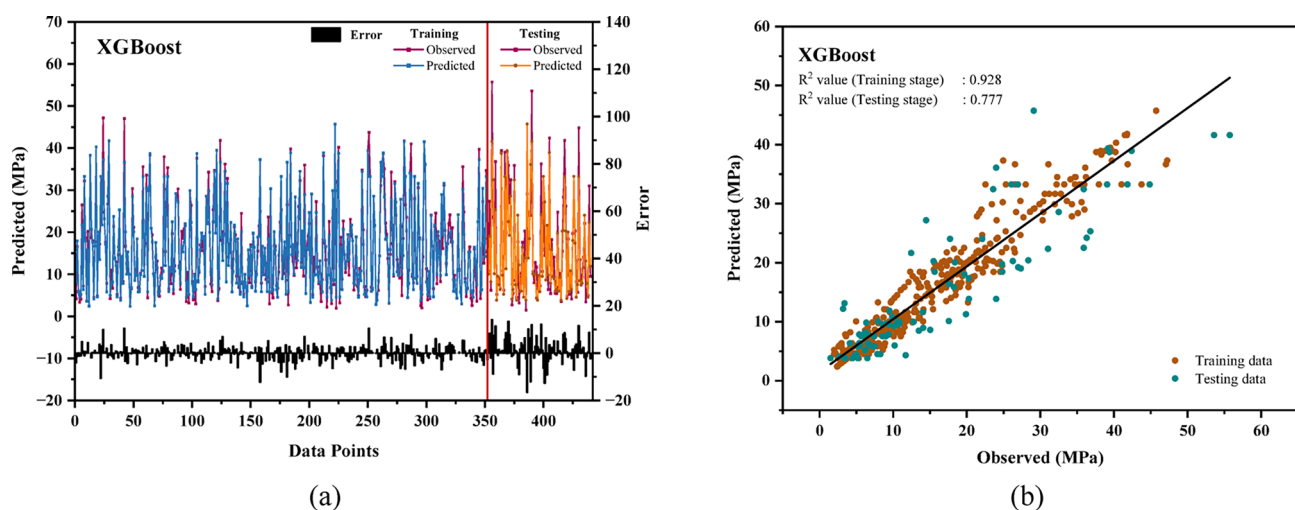


Fig. 4. (a) The relative error in compressive strength predictions using the XGB model for train-test datasets. (b) The linear relationship between predicted and actual compressive strength of the XGB model.

sustainable construction practices and intelligent waste management¹⁰⁵. Among the individual models (XGB, LGB, and DT), the DT model demonstrates similar prediction performance on both new and existing datasets, although the XGB model achieves superior accuracy overall.

Evaluation of the performance of the hybrid model

The hybrid models, XGB-LGB and XGB-DT, exhibited superior performance compared to their individual versions. In Fig. 7(a), the XGB-LGB model records a reduced error, with testing errors remaining below 8 MPa and training errors under 20 MPa. Similarly, Fig. 8(a) illustrates a comparable range of errors for the XGB-DT

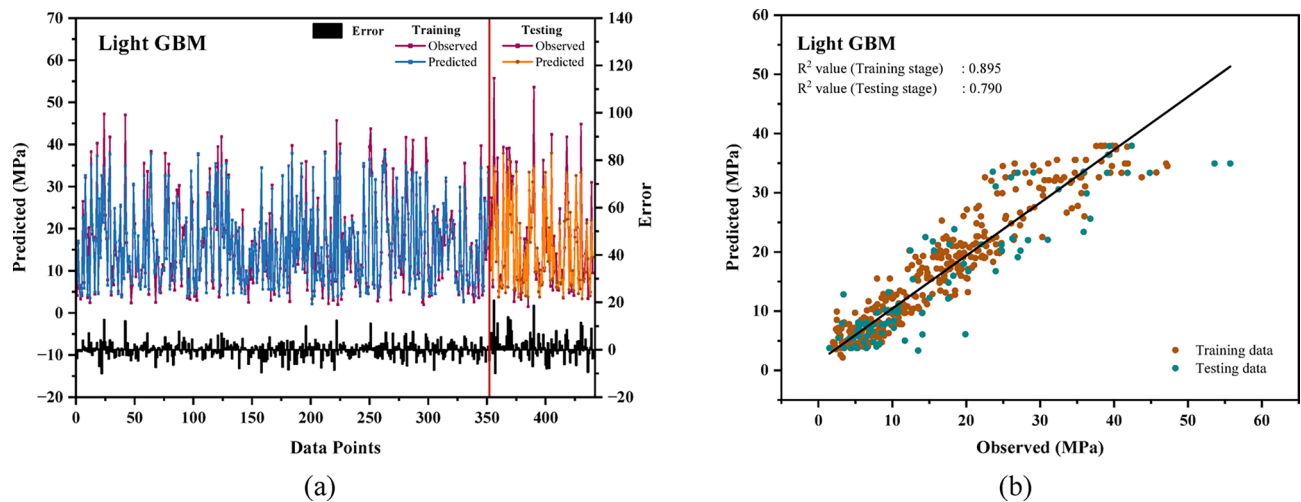


Fig. 5. (a) Relative error of compressive strength in the LGB model for train-test datasets. (b) The linear relationship between predicted and actual compressive strength of the LGB model.

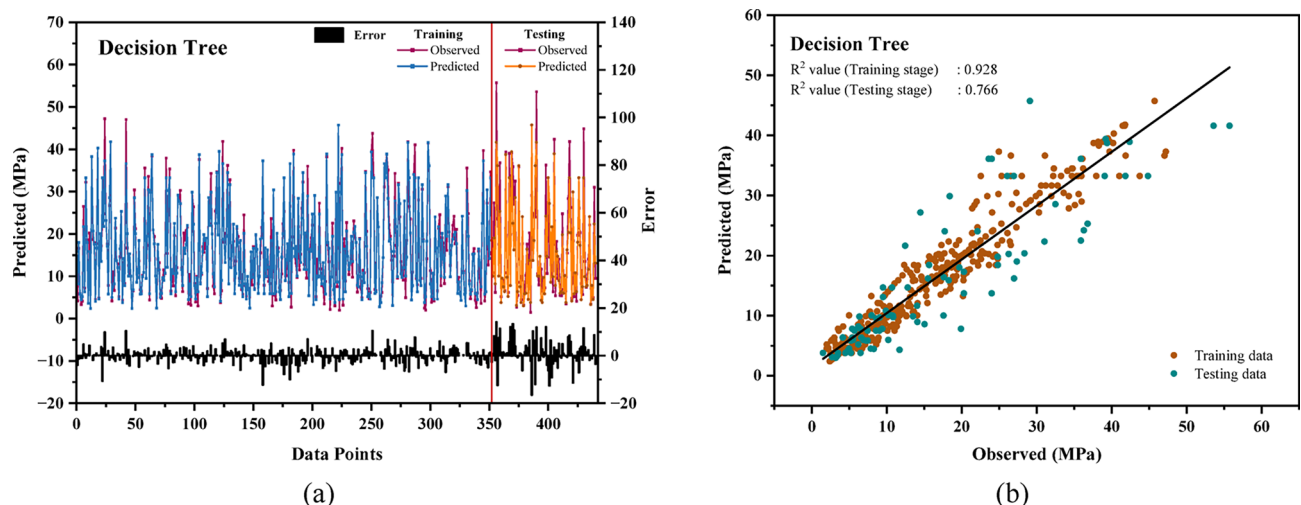


Fig. 6. (a) Relative error of compressive strength for the Decision Tree (DT) model for train-test datasets. (b) The linear relationship between predicted and actual compressive strength of the DT model.

model. The regression coefficients for both models, shown in Figs. 7(b) and 8(b), indicate greater accuracy than the standalone models, highlighting their enhanced predictive capability. The hybrid XGB-LGB model exhibited exceptional accuracy ($R^2=0.95$, $RMSE=5.255$ MPa) in predicting the CS of rice husk ash-based concrete, underscoring its potential as a precise and efficient instrument for sustainable construction applications¹⁰⁶.

Evaluation of the performance of developed models

The Taylor diagram provides an effective visual aid for comparing the performance of different models with experimental data. It estimates how well the predicted values align with the experimental values by determining the correlation coefficient (R^2) and standard deviation. An R^2 -value of 1 indicates that the anticipated values match the observed values perfectly in terms of relationship and variability. Figure 9 illustrates a Taylor diagram for five training models that demonstrate a correlation coefficient higher than 0.7, indicating relatively strong alignment with the train data, although slight differences can be noted in their variability. The standard deviation ranges from approximately 10 to 12.5, with the XGB and DT models positioned close to the reference point, suggesting they have higher precision.

The XGB-DT model, while showing a lower correlation coefficient, also exhibits a slightly elevated standard deviation compared to the XGB and DT models, reflecting its relatively lower consistency in prediction. Conversely, the XGB-LGB combination demonstrates strong performance with a slightly better correlation than some other models. Overall, the XGB, DT, and XGB-LGB models appear to balance correlation and standard deviation, indicating their close alignment with the reference data and solid predictive performance.

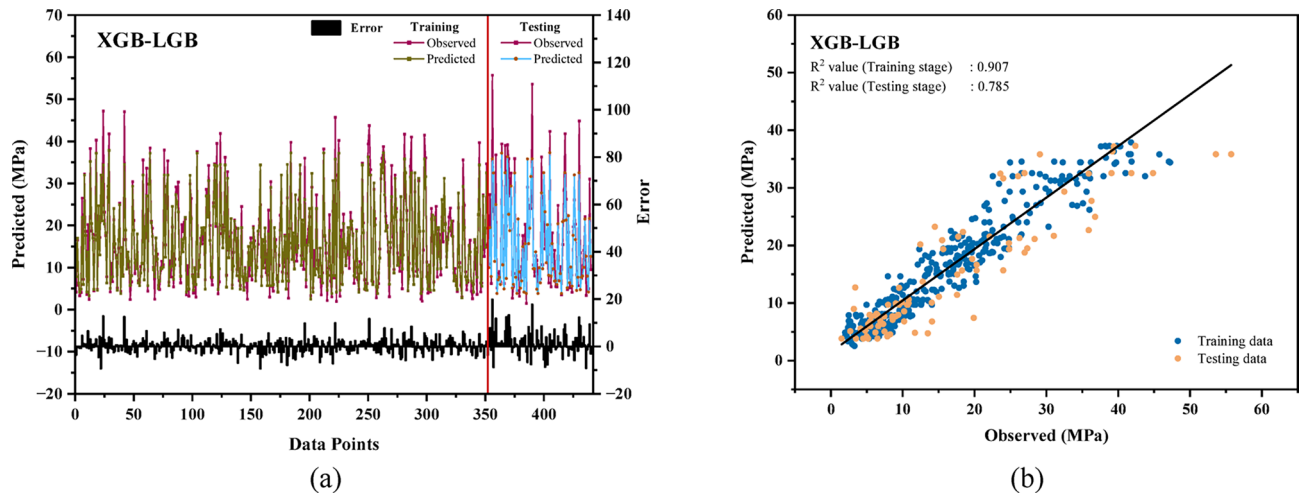


Fig. 7. (a) Relative error in compressive strength for the XGB-LGB model for train-test datasets. (b) The linear relationship between predicted and actual compressive strength of the XGB-LGB model.

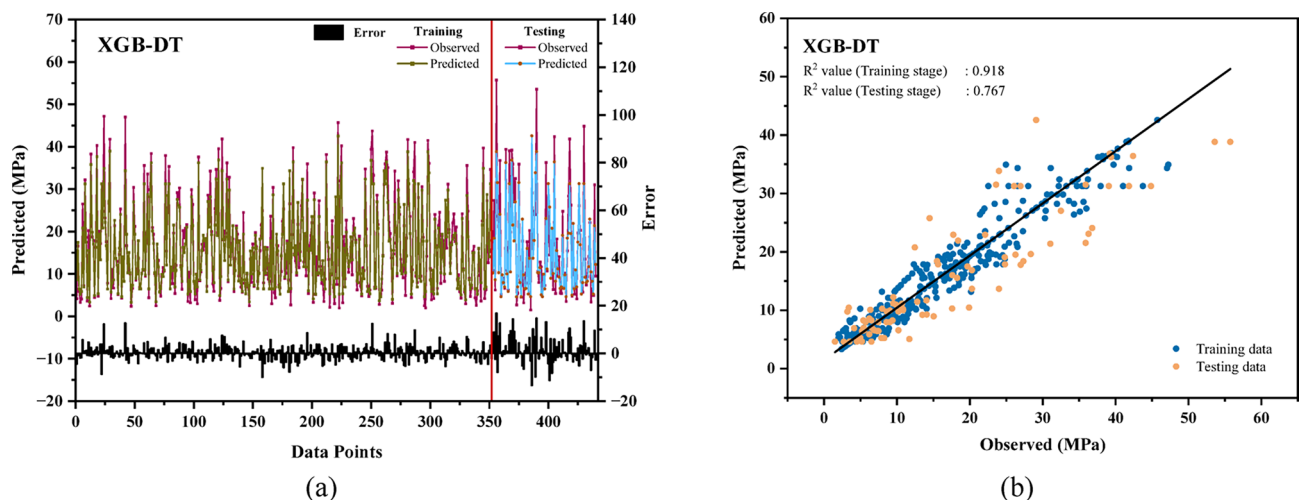


Fig. 8. (a) Relative error in compressive strength for the XGB-DT model across training and testing datasets (b) Linear relationship between predicted and actual observations in the XGB-DT model.

Figure 10 displays a radar chart that evaluates various machine learning models according to R^2 , RMSE, MAE, and MAPE for both training and testing datasets. In Fig. 10(a), the XGB and DT models exhibit the strongest performance, achieving R^2 values of 0.928 for training both models and 0.777 and 0.766 for testing, respectively. In contrast, the LGB model recorded the lowest R^2 values, with 0.895 for training and 0.79 for testing. Among the two hybrid models, XGB-DT demonstrates the strongest performance during training with an R^2 value of 0.917, while XGB-LGB achieves the best performance during testing with an R^2 value of 0.785, which is closest to the highest testing value. Figure 10(b) indicates that the DT model achieved the smallest RMSE for both datasets, at 2.8634 for training and 6.0766 for testing, while the LGB model had the highest RMSE, XGB-DT also exhibits lower RMSE values compared to XGB-LGB. In Fig. 10(c), the DT model once again displayed the lowest MAE for testing at 1.9416. MAE value of 1.945 and 2.25 for XGB and XGB-DT, respectively, whereas the LGB model reported the highest at 2.5967. Lastly, Fig. 10(d) illustrates the DT model with the least MAPE at 16.6413 for testing, while the LGB model had the highest at 23.4427. Overall, the XGB and DT models surpassed the others, demonstrating superior R^2 values and reduced error rates, whereas, among the hybrid models, XGB-DT achieved the best performance.

Furthermore, the time efficiency of each analytical approach plays a critical role in practical implementation. In this study, training durations for all models were monitored, and it was observed that the DT model offered the shortest computational time due to its simpler structure, while the XGB-LGB hybrid model required the longest training period, reflecting the computational complexity of ensemble techniques. Despite the slightly extended training time, the hybrid models (XGB-LGB and XGB-DT) demonstrated improved accuracy, justifying the

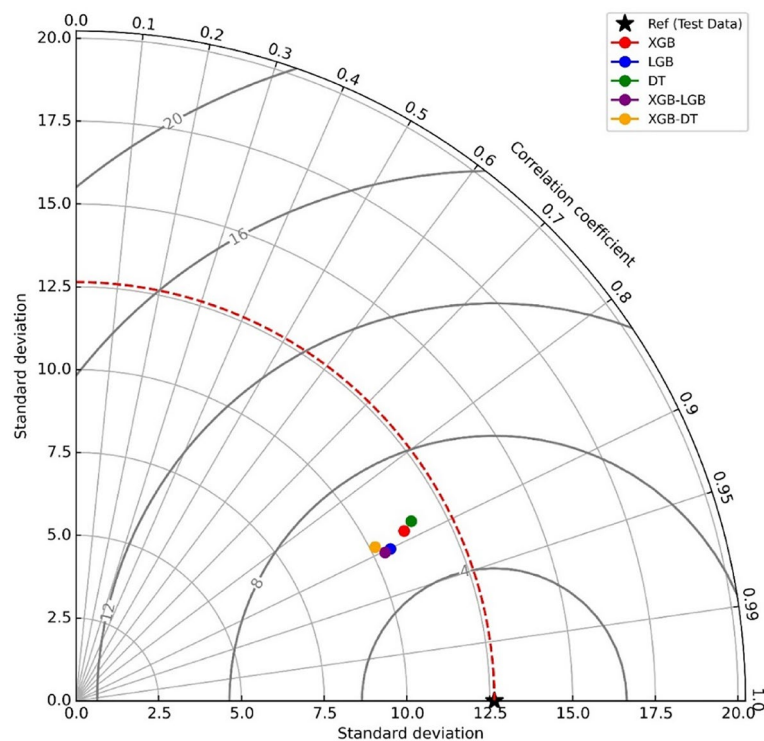


Fig. 9. Taylor diagram for evaluating the effectiveness of ML models used in this study.

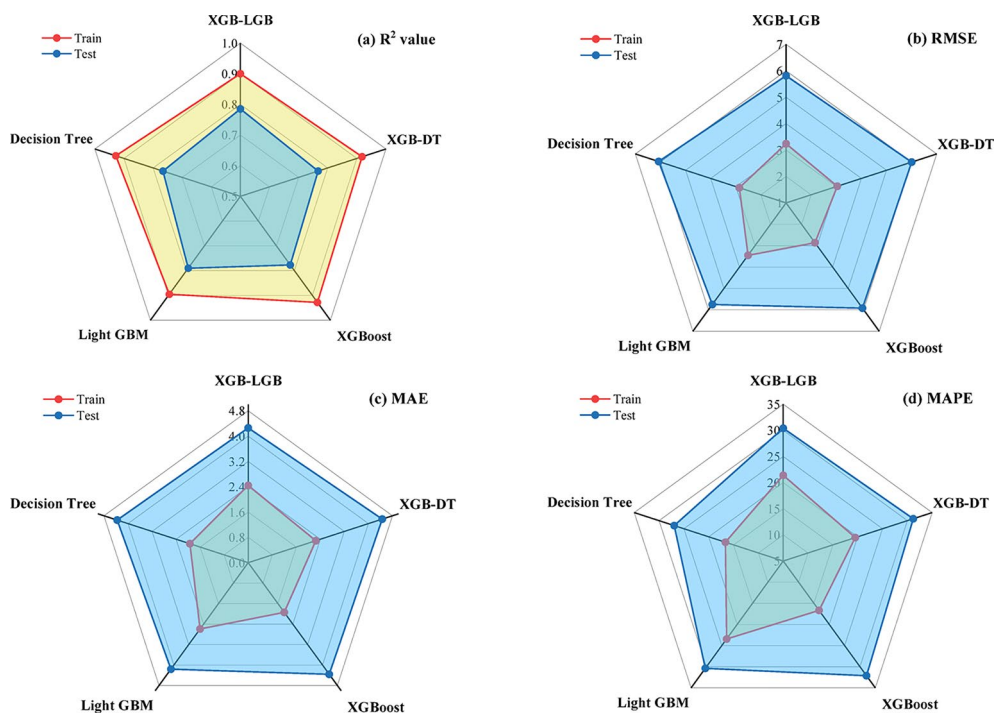


Fig. 10. Performance of ML models.

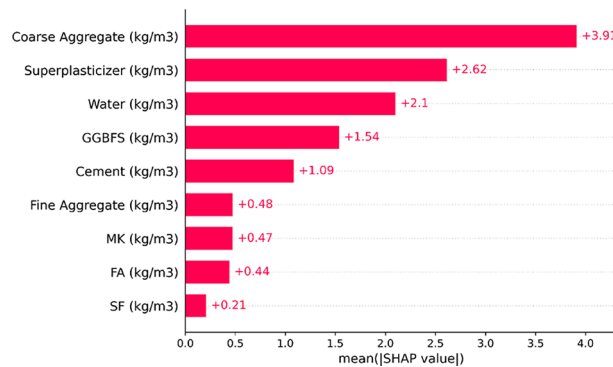


Fig. 11. Mean values of SHAP.

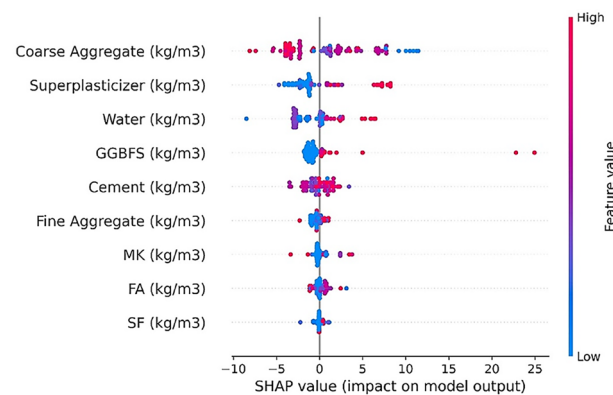


Fig. 12. SHAP impact values.

trade-off between time and performance. In addition, the boosting-based algorithms (XGB and LGB) required prolonged time comparing DT for processing and predicting the CS of modified concrete.

SHAP analysis

Identifying the essential elements for precise model predictions is essential. This study utilized mean SHAP values to assess the importance of each input variable. Figure 11 depicts the average SHAP values for the input features of the XGB model. Coarse aggregate was found to have the most significant influence on the model's predictions, with a SHAP value of 3.91. Superplasticizer ranked as the second most important factor influencing concrete compressive strength. Among the binders, GGBFS contributed the most, followed by MK, FA, and SF, all of which improved strength by enhancing the concrete's microstructural stiffness. Water showed a higher SHAP value than fine aggregate, though both were among the less significant factors. The SHAP analysis demonstrated that chloride resistance is substantially influenced by compressive strength, aging, and the water-binder ratio. The optimal performance was achieved with a water-binder ratio of 0.30–0.35 and a 15% metakaolin replacement, which ensured enhanced durability against chloride penetration⁷³.

Overview diagram of SHAP effect metrics

The application of SHAP game theory facilitated the ranking of the significance and predictive influence of all input variables. SHAP quantifies the average marginal effect of each variable over all possible combinations. The relative importance of each variable was computed and ranked in descending order to emphasize the overall feature relevance. Figure 12 illustrates the significance of each variable through colored points, with the Y-axis indicating feature importance and the X-axis representing SHAP values. This graph demonstrates the relationship between each input and its corresponding SHAP value. For example, cement exhibited a notable SHAP value, signifying a positive contribution to concrete compressive strength. Conversely, coarse aggregate displayed a negative correlation between its SHAP value and feature values. SP and GGBFS exhibited moderate relevance, whereas SF and FA were ranked lower. MK and fine aggregate, despite their moderate SHAP values, had a beneficial effect on concrete strength. An expanded dataset would offer clearer insights into the patterns and effects of each input.

SHAP heat map

Figure 13 displays a heat map depicting SHAP values for different features across individual data instances. The $f(x)$ graph showcases the fluctuations in SHAP values among data points with comparable $f(x)$ scores. In the

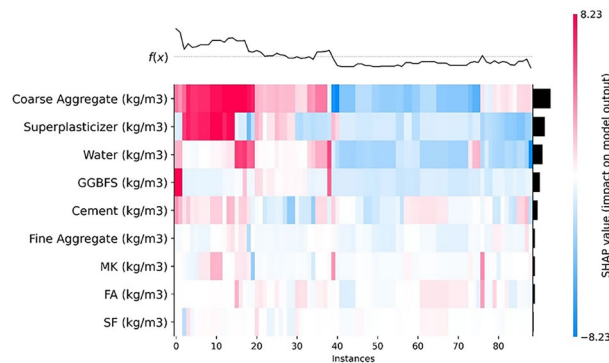


Fig. 13. SHAP heat map.

initial 40 instances, coarse aggregate and SP exhibit significant SHAP magnitudes; however, from data points 40 to 90, the SHAP values turn negative for coarse aggregate. Coarse aggregate displays a higher SHAP value in the initial 40 cases, which declines for the remaining points. Other features consistently exhibit lower SHAP values, likely due to data limitations in past studies. This research aims to fill that gap, offering insights for developing more accurate predictive models.

Investigation of SHAP interaction graphs

Figure 14 illustrates the relationships among input variables and their impact on compressive strength. In Fig. 14(a), the relationship between cement and FA strengthens as cement content increases. Figure 14(b) highlights that coarse aggregate and water have a greater impact, up to 1000 kg/m³, after which the influence decreases. In Fig. 14(c), the interaction between fine aggregate and water shows a moderate correlation up to 100 kg/m³, with a higher impact above 400 kg/m³. According to Fig. 14(d), FA negatively affects compressive strength, suggesting that keeping FA between 10 and 100 kg/m³ yields better strength. In Fig. 14(e), using around 200 kg/m³ of GGBFS with coarse aggregate positively impacts compressive strength. Figure 14(f) shows that an optimal MK content between 10 and 50 kg/m³ improves compressive strength with coarse aggregate. Figure 14(g) shows SF having a stronger effect at around 40 kg/m³ when combined with SP, and Fig. 14(h) highlights that SP performs best with coarse aggregate in the 3 to 5 kg/m³ range. Finally, Fig. 14(i) demonstrates that water significantly influences compressive strength as coarse aggregate increases. The outcomes of this SHAP analysis are consistent with the results reported by Li et al.¹⁰⁷, which details the beneficial and detrimental influences on the strength of concrete.

PDP analysis

Figure 15 illustrates the impact of various input factors on the compressive strength of concrete. In Fig. 15(a), an increase in cement content from 300 to 375 kg/m³ correlates with a rise in compressive strength. Figure 15(b) shows that coarse aggregate contributes positively to compressive strength, peaking at 1500 kg/m³ before declining from 20 MPa to 10 MPa. The partial dependence plot (PDP) in Fig. 15(c) indicates that fly ash (FA) has a negligible effect on compressive strength. Conversely, Fig. 15(d) reveals that fine aggregate substantially increases compressive strength when its quantity exceeds 275 kg/m³, with values climbing from 15.5 MPa to 17.25 MPa. Figure 15(e) highlights the favorable impact of ground granulated blast furnace slag (GGBFS), where compressive strength rises from 18 MPa to 28 MPa as GGBFS content increases from 150 to 200 kg/m³, stabilizing beyond 200 kg/m³. The influence of metakaolin (MK) is depicted in Fig. 15(f), where its effect becomes significant at around 25 kg/m³. Figure 15(g) illustrates that silica fume (SF) enhances compressive strength up to 30 kg/m³, after which its impact stabilizes. Finally, Figs. 15(h) and (i) show that superplasticizer (SP) gradually improves compressive strength up to 5 kg/m³ before declining, while water plays a crucial role in compressive strength within the range of 140 to 190 kg/m³. The performance of RHA-FA concrete is considerably enhanced by the optimal range for enhancing compressive strength, which is achieved with a fly ash content of 20–30% and rice husk ash proportions below 10%, as revealed by PDP analysis¹⁰⁸.

Conclusions

This study aimed to assess the reliability of AI models in predicting the compressive strength of concrete and to examine the effects of various raw materials on compressive strength. Nine critical input factors were evaluated for their predictive power, and their interrelations were thoroughly investigated. The outcomes of this research led to the following conclusions:

- All five machine learning algorithms (XGB-LGB, XGB-DT, XGB, LGB, and DT) exhibited robust predictive capabilities and a strong ability to generalize. With R^2 values surpassing 0.89, these models proved highly efficient in forecasting concrete strength, demonstrating their dependability for precise predictions.
- The ensemble ML models, XGB-LGB and XGB-DT, provided enhanced performance compared to the single models, showcasing improved accuracy and effectiveness. Furthermore, the hybrid models reduce overfitting risk by minimizing the difference between training and testing R^2 value.

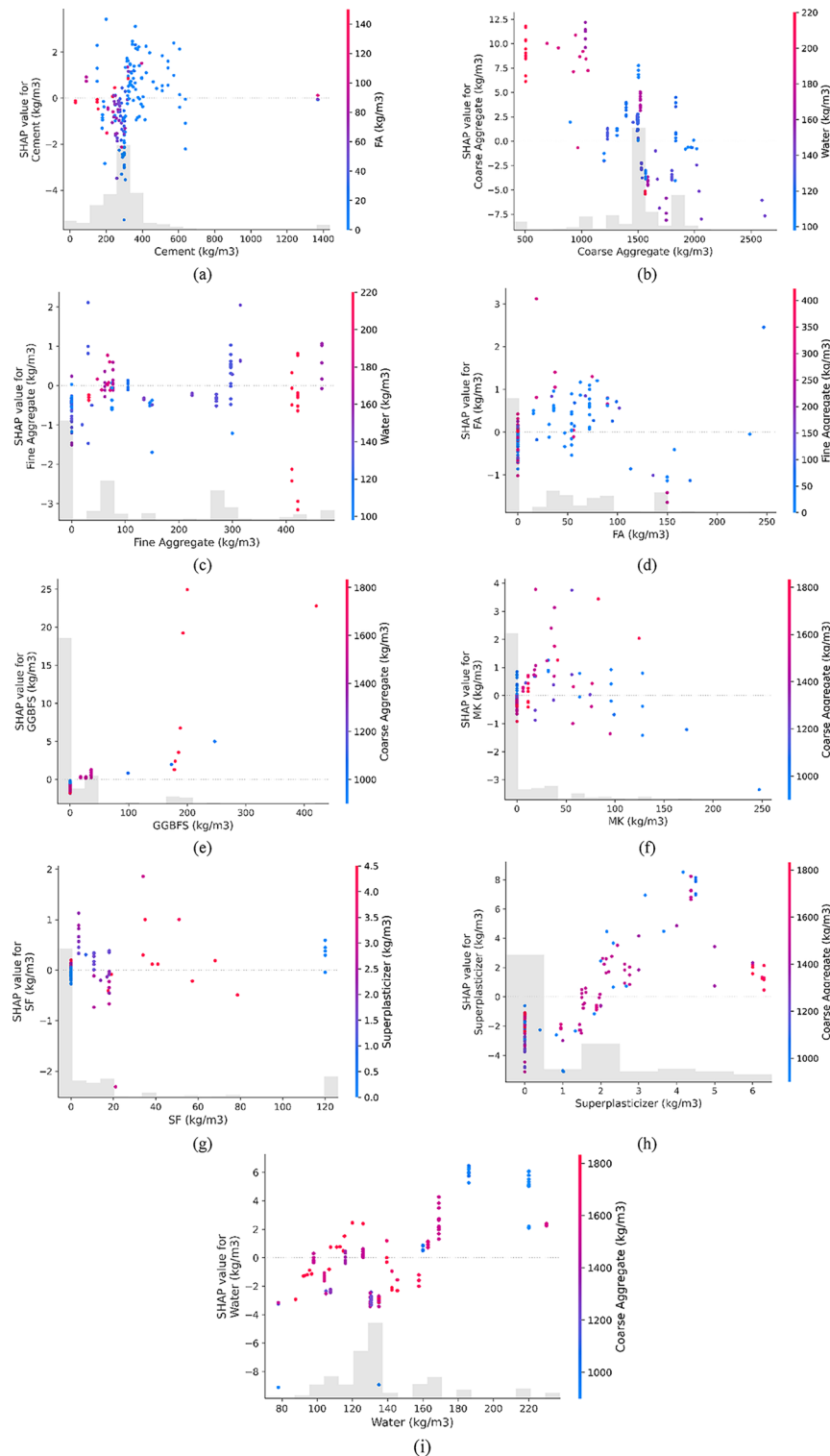


Fig. 14. Interaction plots for all input parameters.

- Among all the single ML models, DT exhibited the top performance, with the lowest RMSE of 2.863, MAPE of 16.641, and MAE of 1.941. Furthermore, the hybrid XGB-LGB model demonstrates strong balanced performance with a slightly better correlation than other models.
- The SHAP analysis revealed that coarse aggregate, superplasticizers, water, GGBFS, and cement were the most significant elements impacting the predicted compressive strength. Additionally, the PDP evaluation indicated that the best mixture for optimizing concrete strength comprised 150–200 kg/m³ of GGBFS.

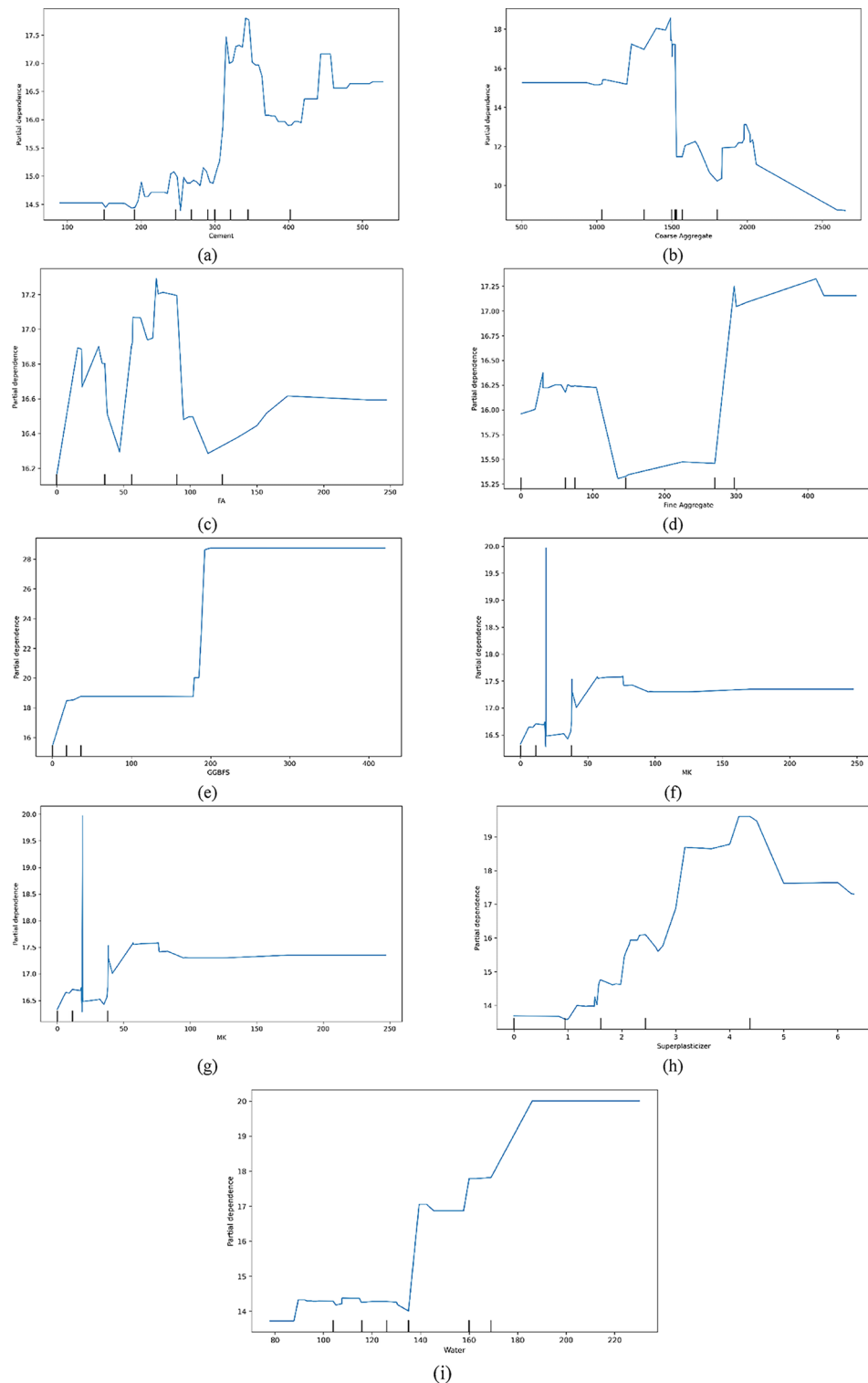


Fig. 15. Examination of the stacking model through PDP for the input variables.

The established methods could assess the compressive strength of industrial waste-based concrete utilizing a variety of input variable values, saving time and effort for further experiments. Implementing these algorithms on a construction site speeds up work and reduces the need for time-consuming laboratory testing.

Limitations and further studies

Although it is vital to note that other elements, such as the chemical and physical qualities of materials, fluctuations in temperature, and resistance to corrosion, chloride, and acid attacks, could potentially influence

the compressive strength of industrial waste-based concrete, it is also important to make sure that these other factors are taken into consideration. That being said, this study was unable to conduct a complete analysis of the impact of these characteristics because there was a limited amount of data available. Consequently, the objective of future research should be to broaden the scope of the database so that it incorporates these additional parameters. This will make it possible to construct more robust predictive models that are based on deep learning. In addition, additional research could investigate the possibility of applying the machine learning frameworks that have been developed for the purpose of forecasting other essential qualities of concrete. These properties include flexural strength, tensile strength, durability, porosity, flowability, shrinkage, cost, and carbon footprint.

Data availability

Our objective is to maintain control over unsupervised usage that may lead to unintentional duplication of research efforts or reduced novelty in future studies. However, the dataset will be provided upon request. Please contact Dr. Md. Habibur Rahman Sobuz (email: habib@becm.kuet.ac.bd) if anyone needs the data for this study.

Received: 8 October 2024; Accepted: 11 July 2025

Published online: 20 July 2025

References

- Hossain, M. M. et al. *BIM-based smart safety monitoring system using a mobile app: a case study in an ongoing construction site*. Construction Innovation, ahead-of (2023).
- Sobuz, M. H. R. et al. Combined influence of modified recycled concrete aggregate and Metakaolin on high-strength concrete production: experimental assessment and machine learning quantifications with advanced SHAP and PDP analyses. *Constr. Build. Mater.* **461**, 139897 (2025).
- Kanthé, V. Effect of superplasticizer on strength and durability of rice husk Ash concrete. *Iranica J. Energy Environ.* **12** (3), 204–208 (2021).
- Khan, M. M. H. et al. Effect of various powder content on the properties of sustainable self-compacting concrete. *Case Stud. Constr. Mater.* **19**, e02274 (2023).
- Sobuz, M. H. R. et al. Evaluating the Properties of Demolished Aggregate Concrete with Non-destructive Assessment, in *Advances in Civil Engineering, Lecture Notes in Civil Engineering*, and Editors. Springer Singapore: Singapore. pp. 223–233. (2022).
- Kabbo, M. K. I. et al. *Experimental assessment and machine learning quantification of structural Eco-Cellular lightweight concrete incorporating waste marble powder and silica fume*. *J. Building Eng.*, : p. 112557. (2025).
- Celik, K. et al. Mechanical properties, durability, and life-cycle assessment of self-consolidating concrete mixtures made with blended Portland cements containing fly Ash and limestone powder. *Cem. Concr. Compos.* **56**, 59–72 (2015).
- Andrew, R. M. Global CO₂ emissions from cement production. *Earth Syst. Sci. Data.* **10** (1), 195–217 (2018).
- Amani, A., Ramezani-pour, A. M. & Palassi, M. Investigation on the sustainable use of electric Arc furnace slag aggregates in eco-friendly alkali-activated low fineness slag concrete as a green construction composite. *J. Clean. Prod.* **307**, 127257 (2021).
- Sobuz, M. H. R. et al. Assessment of hybrid fiber reinforced graphene nano-engineered concrete composites: from experimental testing to explainable machine learning modeling. *J. Mater. Res. Technol.* **36**, 1409–1430 (2025).
- Momshad, A. M. et al. *Assessing the engineering properties and environmental impact with explainable machine learning analysis of sustainable concrete utilizing waste banana leaf ash as a partial cement replacement*. Cleaner Engineering and Technology, 24: p. 100886. (2025).
- Sobuz, M. H. R. et al. Assessment of mechanical properties with machine learning modeling and durability, and microstructural characteristics of a biochar-cement mortar composite. *Constr. Build. Mater.* **411**, 134281 (2024).
- Datta, S. D. et al. *Critical project management success factors analysis for the construction industry of Bangladesh*. International Journal of Building Pathology and Adaptation, ahead-of (2023).
- Rahman, M. & Sobuz, H. R. *Comparative study of IPS & PPVC precast system—A case study of public housing buildings project in Singapore*. in *Proceedings of the 4th International Conference on Civil Engineering for Sustainable Development (ICCESD 2018)*, KUET, Khulna, Bangladesh. (2018).
- Aditto, F. S. et al. Fresh, mechanical and microstructural behaviour of high-strength self-compacting concrete using supplementary cementitious materials. *Case Stud. Constr. Mater.* **19**, e02395 (2023).
- Hasan, N. M. S. et al. Investigation of lightweight and green concrete characteristics using coconut shell aggregate as a replacement for conventional aggregates. *Int. J. Civil Eng.* **22** (1), 37–53 (2024).
- Saha, A., Sobuz, M. H. R., Hoque, M. I. & Mujahid, R. Influence of waste glass aggregates on the rheological properties of self-consolidated concrete. *Australian J. Civil Eng.* **18** (2), 272–285 (2020).
- Sobuz, M. H. R. et al. Mechanical properties and flexural response of palm shell aggregate lightweight reinforced concrete beam. *Sustainability* **15** (22), 15783 (2023).
- Md. Sadiqul Hasan, N., Rahman Sobuz, H., Sharif Auwalu, A. & Tamanna, N. Investigation into the suitability of Kenaf Fiber to produce structural concrete. *Adv. Mater. Lett.* **6** (8), 731–737 (2015).
- Ahmed, E. & Sobuz, H. R. Flexural and time-dependent performance of palm shell aggregate concrete beam. *KSCE J. Civ. Eng.* **15** (5), 859–865 (2011).
- Uddin, M. A. et al. *The Effect of Curing time on Compressive Strength of Composite Cement Concrete* 204–208 (Applied Mechanics and Materials, 2012).
- Sobuz, M. H. R., Datta, S. D. & Akid, A. S. M. Investigating the combined effect of aggregate size and sulphate attack on producing sustainable recycled aggregate concrete. *Australian J. Civil Eng.* **21** (2), 224–239 (2023).
- Hasan, R. et al. Eco-friendly self-consolidating concrete production with reinforcing jute fiber. *J. Building Eng.* **63**, 105519 (2023).
- NoParast, M., Hematian, M., Ashrafi, A., Amiri, M. J. T. & Azarijafari, H. Development of a non-dominated sorting genetic algorithm for implementing circular economy strategies in the concrete industry. *Sustainable Prod. Consum.* **27**, 933–946 (2021).
- Kurda, R., Silvestre, J. D. & de Brito, J. *Toxicity and environmental and economic performance of fly Ash and recycled concrete aggregates use in concrete: A review*. *Heliyon*, **4**(4). (2018).
- Oyebisi, S. et al. *Effects Cashew Nutshell Ash Therm. Sustain. Prop. Cem. Concrete Heliyon*, **8**(11). (2022).
- Rahman Sobuz, M. H. et al. Experimental and analytical studies of size effects on compressive ductility response of Ultra-High-Performance Fiber-Reinforced concrete. *Constr. Build. Mater.* **409**, 133864 (2023).
- Nadim, F. et al. *Effect of silica fume on the microstructural and mechanical properties of concrete made with 100% recycled aggregates*. Revista de la Construcción. *J. Constr.* **23** (2), 413–435 (2024).
- Celik, K. et al. High-volume natural volcanic Pozzolan and limestone powder as partial replacements for Portland cement in self-compacting and sustainable concrete. *Cem. Concr. Compos.* **45**, 136–147 (2014).

30. Islam, M. H. et al. Assessing the influence of banana leaf Ash as pozzolanic material for the production of green concrete: A mechanical and microstructural evaluation. *Materials* **17** (3), 720 (2024).
31. Chen, D. et al. A state of review on manufacturing and effectiveness of ultra-high-performance fiber reinforced concrete for long-term integrity of concrete structures. *Adv. Concrete Constr.* **17** (5), 293–310 (2024).
32. Hakeem, I. Y. et al. Restoration of load capacity and stiffness of continuous steel–concrete composite beams having web openings using externally applied FRP strips. *Struct. Concrete*. **25** (5), 3505–3530 (2024).
33. Thomas, M. *Optimizing the Use of Fly Ash in Concrete* Vol. 5420 (Portland Cement Association Skokie, IL, 2007).
34. Chi, M. C., Chi, J. H. & Wu, C. H. Effect of GGBFS on compressive strength and durability of concrete. *Adv. Mater. Res.* **1145**, 22–26 (2018).
35. Sobuz, M. H. R. et al. Development of Self-Compacting concrete incorporating rice husk Ash with waste galvanized copper wire Fiber. *Buildings* **12** (7), 1024 (2022).
36. Mishra, M. Review of experimental and FE parametric analysis of CFRP-Strengthened Steel-Concrete composite beams. *Journal of mechanical. J. Mech. Civil Industrial Eng.* **3** (3), 92–101 (2022).
37. Uddin, M. A., Jameel, M., Sobuz, H. R., Islam, M. S. & Hasan, N. M. S. Experimental study on strength gaining characteristics of concrete using Portland composite cement. *KSCE J. Civ. Eng.* **17** (4), 789–796 (2013).
38. Das, S. et al. Effects of incorporating hybrid fibres on rheological and mechanical properties of fibre reinforced concrete. *Constr. Build. Mater.* **262**, 120561 (2020).
39. Ahmed, E., Sobuz, H. R. & Sutan, N. M. Flexural performance of CFRP strengthened RC beams with different degrees of strengthening schemes. *Int. J. Phys. Sci.* **6** (9), 2229–2238 (2011).
40. Datta, S. D., Sobuz, M. H. R., Akid, A. S. M. & Islam, S. Influence of coarse aggregate size and content on the properties of recycled aggregate concrete using non-destructive testing methods. *J. Building Eng.* **61**, 105249 (2022).
41. Sobuz, M. H. R. et al. Evaluating the effects of recycled concrete aggregate size and concentration on properties of high-strength sustainable concrete. *J. King Saud Univ. - Eng. Sci.* **36** (3), 216–227 (2024).
42. Kharate, N., Anerao, P., Kulkarni, A. & Abdullah, M. Explainable AI techniques for comprehensive analysis of the relationship between process parameters and material properties in FDM-Based 3D-Printed biocomposites. *J. Manuf. Mater. Process.* **8** (4), 171 (2024).
43. Alamry, A., Islam, M. S., Hamdan, S. & Abdullah, M. *Enhancing the Performance of Hybrid bio-composites Reinforced with Natural Fibers by Using Coupling Agents* (Materials Research Ex, 2024).
44. Gorgun, E., Ali, A. & Islam, M. S. Biocomposites of Poly(Lactic Acid) and Microcrystalline Cellulose: Influence of the Coupling Agent on Thermomechanical and Absorption Characteristics. *ACS Omega*. **9** (10), 11523–11533 (2024).
45. Sobuz, M. H. R. et al. High-Strength Self-Compacting concrete production incorporating supplementary cementitious materials: experimental evaluations and machine learning modelling. *Int. J. Concrete Struct. Mater.* **18** (1), 67 (2024).
46. Alyami, M. et al. Estimating compressive strength of concrete containing rice husk Ash using interpretable machine learning-based models. *Case Stud. Constr. Mater.* **20**, e02901 (2024).
47. Shi, M. et al. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. *Mech. Syst. Signal Process.* **188**, 110022 (2023).
48. Shi, M. L., Lv, L. & Xu, L. A multi-fidelity surrogate model based on extreme support vector regression: fusing different fidelity data for engineering design. *Eng. Comput.* **40** (2), 473–493 (2023).
49. Al-Saraireh, M. & Ali, M. H. R. S. Eco-Friendly Ultra-High-Performance Fiber-Reinforced concrete production: A review. *Civil Eng. Archit.* **12** (4), 2952–2969 (2024).
50. Liu, C., Zhang, W., Liu, H., Lin, X. & Zhang, R. A compressive strength prediction model based on the hydration reaction of cement paste by rice husk Ash. *Constr. Build. Mater.* **340**, 127841 (2022).
51. Hamidian, P., Alidoust, P., Golafshani, E. M., Niaoul, K. P. & Behnood, A. Introduction of a novel evolutionary neural network for evaluating the compressive strength of concretes: A case of rice husk Ash concrete. *J. Building Eng.* **61**, 105293 (2022).
52. Mansour, W., Osama, B., Li, W., Wang, P. & Sobuz, M. H. R. Numerical response of Concrete-Filled steel tubular (CFST) columns externally strengthened with FRP composites subjected to Cyclic loading. *Int. J. Concrete Struct. Mater.* **18** (1), 82 (2024).
53. Rezaie-Balf, M. et al. Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: comparative assessment of a noise suppression hybridization approach. *J. Clean. Prod.* **271**, 122576 (2020).
54. Sun, K. et al. An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction. *Eng. Appl. Comput. Fluid Mech.* **15** (1), 251–271 (2021).
55. Ashrafiyan, A., Taheri Amiri, M. J. & Haghighi, F. Modeling the slump flow of self-compacting concrete incorporating Metakaolin using soft computing techniques. *J. Struct. Constr. Eng.* **6** (2), 5–20 (2019).
56. Ashrafiyan, A. et al. Classification-based regression models for prediction of the mechanical properties of roller-compacted concrete pavement. *Appl. Sci.* **10** (11), 3707 (2020).
57. Ashrafiyan, A., Gandomi, A. H., Rezaie-Balf, M. & Emadi, M. An evolutionary approach to formulate the compressive strength of roller compacted concrete pavement. *Measurement* **152**, 107309 (2020).
58. Saha, P., Sapkota, S. C., Das, S. & Kwatra, N. *Prediction of Fresh and Hardened Properties of self-compacting Concrete Using Ensemble Soft Learning Techniques*. 1–23 (Multiscale and Multidisciplinary Modeling, 2024).
59. Jabin, J. A., Khondoker, M. T. H., Sobuz, M. H. R. & Aditto, F. S. High-temperature effect on the mechanical behavior of recycled fiber-reinforced concrete containing volcanic pumice powder: an experimental assessment combined with machine learning (ML)-based prediction. *Constr. Build. Mater.* **418**, 135362 (2024).
60. Sobuz, M. H. R. et al. Optimization of recycled rubber self-compacting concrete: experimental findings and machine learning-based evaluation. *Heliyon* **10** (6), e27793 (2024).
61. Sobuz, M. H. R. et al. Enhancing sustainable concrete production by utilizing fly Ash and recycled concrete aggregate with experimental investigation and machine learning modeling. *J. Building Pathol. Rehabilitation*. **9** (2), 134 (2024).
62. Görgün, E. *Characterization of Superalloys by Artificial Neural Network Method*. in *Online International Symposium on Applied Mathematics and Engineering (ISAME22) January 21–23, 2022 Istanbul-Turkey*. (2022).
63. Amin, M. N. et al. *Split Tensile Strength. Prediction Recycled aggregate-based Sustainable Concrete Using Artif. Intell. Methods Mater.*, **15**(12): 4296. (2022).
64. Zheng, D. et al. *Flexural Strength. Prediction Steel fiber-reinforced Concrete Using Artif. Intell. Mater.*, **15**(15): 5194. (2022).
65. Sobuz, M. H. R., Khatun, M., Kabbo, M. K. I. & Sutan, N. M. An explainable machine learning model for encompassing the mechanical strength of polymer-modified concrete. *Asian J. Civil Eng.*, (2024).
66. Golafshani, E., Afzali, S. A. E., Chiniforush, A. A. & Ngo, T. Using ensemble machine learning and metaheuristic optimization for modelling the elastic modulus of geopolymer concrete. *Clean. Mater.* **13**, 100258 (2024).
67. Zou, B. et al. Artificial intelligence-based optimized models for predicting the slump and compressive strength of sustainable alkali-derived concrete. *Constr. Build. Mater.* **409**, 134092 (2023).
68. Shahmansouri, A. A., Bengar, H. A. & Ghanbari, S. Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *J. Building Eng.* **31**, 101326 (2020).
69. Gupta, P., Gupta, N., Saxena, K. K. & Goyal, S. Multilayer perceptron modelling of geopolymer composite incorporating fly Ash and GGBS for prediction of compressive strength. *Adv. Mater. Process. Technol.* **8** (sup3), 1441–1455 (2022).
70. Hasan, N. M. S. et al. *Eco-friendly concrete incorporating palm oil fuel ash: fresh and mechanical properties with machine learning prediction, and sustainability assessment*. *Heliyon*, **9**(11). (2023).

71. Datta, S. D. et al. Analysis of the characteristics and environmental benefits of rice husk Ash as a supplementary cementitious material through experimental and machine learning approaches. *Innovative Infrastructure Solutions*. **9** (4), 121 (2024).
72. Khan, M. I. & Abbas, Y. M. Intelligent data-driven compressive strength prediction and optimization of reactive powder concrete using multiple ensemble-based machine learning approach. *Constr. Build. Mater.* **404**, 133148 (2023).
73. Alabdullah, A. A. et al. Prediction of rapid chloride penetration resistance of Metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **345**, 128296 (2022).
74. Jagadeesh, P., de Prado-Gil, J., Silva-Monteiro, N. & Martínez-García, R. Assessing the compressive strength of self-compacting concrete with recycled aggregates from mix ratio using machine learning approach. *J. Mater. Res. Technol.* **24**, 1483–1498 (2023).
75. Alahmari, T. S., Ashraf, J., Sobuz, M. H. R. & Uddin, M. A. Predicting the compressive strength of fiber-reinforced self-consolidating concrete using a hybrid machine learning approach. *Innovative Infrastructure Solutions*. **9** (11), 446 (2024).
76. Saha, A., Aditto, F. S., Kundu, L., Sobuz, M. H. R. & Sunny, M. M. H. Analysis of waste glass as a partial substitute for coarse aggregate in self-compacting concrete: an experimental and machine learning study. *J. Building Eng.* **98**, 111112 (2024).
77. Mishra, M. Quantifying compressive strength in limestone powder incorporated concrete with incorporating various machine learning algorithms with SHAP analysis. *Asian J. Civil Eng.* (2024).
78. Haq, I. et al. Synergistic effect of fly Ash and stone dust on foam concrete under saline environment: mechanical, non-destructive and machine learning approaches. *Adv. Concrete Constr.* **18** (4), 237 (2024).
79. Sobuz, M. H. R. et al. Assessing the influence of sugarcane Bagasse Ash for the production of eco-friendly concrete: experimental and machine learning approaches. *Case Stud. Constr. Mater.* **20**, e02839 (2024).
80. Ashrafian, A., Behnood, A., Golareshani, E. M., Panahi, E. & Berenjian, J. Toward presenting an ensemble meta-model for evaluation of pozzolanic mixtures incorporating industrial by-products. *Struct. Concrete*. **25** (2), 1305–1323 (2024).
81. Farooq, F., Ahmed, W., Akbar, A., Aslam, F. & Alyousef, R. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *J. Clean. Prod.* **292**, 126032 (2021).
82. Shen, Z., Deifalla, A. F., Kamiński, P. & Dyczko, A. Compressive strength evaluation of ultra-high-strength concrete by machine learning. *Materials* **15** (10), 3523 (2022).
83. Alhakeem, Z. M. et al. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. *Materials* **15** (21), 7432 (2022).
84. Uddin, M. N., Ye, J., Deng, B., Li, L. & Yu, K. Interpretable machine learning for predicting the strength of 3D printed fiber-reinforced concrete (3DP-FRC). *J. Building Eng.* **72**, 106648 (2023).
85. Amin, M. N. et al. Predicting parameters and sensitivity assessment of nano-silica-based fiber-reinforced concrete: a sustainable construction material. *J. Mater. Res. Technol.* **23**, 3943–3960 (2023).
86. Khan, K. et al. Compressive strength Estimation of steel-fiber-reinforced concrete and Raw material interactions using advanced algorithms. *Polymers* **14** (15), 3065 (2022).
87. Song, H. et al. Predicting the compressive strength of concrete with fly Ash admixture using machine learning algorithms. *Constr. Build. Mater.* **308**, 125021 (2021).
88. Rahman, J., Ahmed, K. S., Khan, N. I., Islam, K. & Mangalathu, S. Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. *Eng. Struct.* **233**, 111743 (2021).
89. Friedman, J. H. *Greedy function approximation: a gradient boosting machine*. *Ann. Stat.*, pp. 1189–1232. (2001).
90. 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*. (2016).
91. Xi, B., Li, E., Fissah, Y., Zhou, J. & Segarra, P. LGBM-based modeling scenarios to compressive strength of recycled aggregate concrete with SHAP analysis. *Mech. Adv. Mater. Struct.* **31** (23), 5999–6014 (2024).
92. Hosseinzadeh, M., Dehestani, M. & Hosseinzadeh, A. Prediction of mechanical properties of recycled aggregate fly Ash concrete employing machine learning algorithms. *J. Building Eng.* **76**, 107006 (2023).
93. Ekanayake, I., Meddage, D. & Rathnayake, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Stud. Constr. Mater.* **16**, e01059 (2022).
94. Yang, L., Liu, S., Tsoka, S. & Papageorgiou, L. G. A regression tree approach using mathematical programming. *Expert Syst. Appl.* **78**, 347–357 (2017).
95. Song, Y. Y. & Ying, L. Decision tree methods: applications for classification and prediction. *Shanghai Archives Psychiatry*. **27** (2), 130 (2015).
96. Kingsford, C. & Salzberg, S. L. What are decision trees? *Nat. Biotechnol.* **26** (9), 1011–1013 (2008).
97. Chopra, P., Sharma, R. K., Kumar, M. & Chopra, T. Comparison of machine learning techniques for the prediction of compressive strength of concrete. *Adv. Civil Eng.* **2018** (1), 5481705 (2018).
98. Alahmari, T. S., Kabbo, M. K. I., Sobuz, M. H. R. & Rahman, S. A. Experimental assessment and hybrid machine learning-based feature importance analysis with the optimization of compressive strength of waste glass powder-modified concrete. *Mater. Today Commun.* **44**, 112081 (2025).
99. Sobuz, M. H. R. et al. *Microstructural Behavior and Explainable Machine Learning Aided Mechanical Strength Prediction and Optimization of Recycled glass-based Solid Waste Concrete*. e04305 (Case Studies in Construction Materials, 2025).
100. Sobuz, M. H. R. 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 Stud. Constr. Mater.* **22**, e04568 (2025).
101. Tran, V. Q., Dang, V. Q. & Ho, L. S. Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Constr. Build. Mater.* **323**, 126578 (2022).
102. Zhang, J., Li, D. & Wang, Y. Toward intelligent construction: prediction of mechanical properties of manufactured-sand concrete using tree-based models. *J. Clean. Prod.* **258**, 120665 (2020).
103. Shamsabadi, E. A. et al. Machine learning-based compressive strength modelling of concrete incorporating waste marble powder. *Constr. Build. Mater.* **324**, 126592 (2022).
104. Ahmed, A. H. A., Jin, W. & Ali, M. A. H. Prediction of compressive strength of recycled concrete using gradient boosting models. *Ain Shams Eng. J.* **15** (9), 102975 (2024).
105. Gulghane, A., Sharma, R. & Borkar, P. A formal evaluation of KNN and decision tree algorithms for waste generation prediction in residential projects: a comparative approach. *Asian J. Civil Eng.* **25** (1), 265–280 (2024).
106. Kashem, A., Karim, R., Das, P., Datta, S. D. & Alharthai, M. Compressive strength prediction of sustainable concrete incorporating rice husk Ash (RHA) using hybrid machine learning algorithms and parametric analyses. *Case Stud. Constr. Mater.* **20**, e03030 (2024).
107. Li, X. et al. A systematic review of waste materials in cement-based composites for construction applications. *J. Building Eng.* **45**, 103447 (2022).
108. Karim, R., Islam, M. H., Datta, S. D. & Kashem, A. Synergistic effects of supplementary cementitious materials and compressive strength prediction of concrete using machine learning algorithms with SHAP and PDP analyses. *Case Stud. Constr. Mater.* **20**, e02828 (2024).

Acknowledgements

This research was supported by the “University of Debrecen Program for Scientific Publication”. This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No.

(DGSSR-2023-02-02350). The work was carried out at the BIM laboratory of the Department of Building Engineering and Construction Management, Khulna University of Engineering and Technology, Khulna – 9203, Bangladesh. We express our sincere gratitude to Jouf University, Saudi Arabia, and the University of Debrecen, Hungary, for their support.

Author contributions

Md. Alhaz Uddin: Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Writing - review & editing. Md. Habibur Rahman Sobuz: Conceptualization, Methodology, Validation, Formal analysis, Supervision, Writing - original draft, Writing - review & editing. Md. Kawsarul Islam Kabbo: Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. Md. Kanan Chowdhury Tilak: Formal analysis, Data curation, Validation, Writing - original draft, Writing - review & editing. Ratan Lal: Methodology, Formal analysis, Data curation, Validation, Writing - original draft, Writing - review & editing. Md. Selim Reza: Formal analysis, Data curation, Validation, Writing - review & editing. Fahad Alsharari: Data curation, Validation, Writing - review & editing. Mohamed AbdelMongy: Data curation, Validation, Writing - review & editing. Masuk Abdullah: Data curation, Writing - review & editing.

Funding

Open access funding provided by University of Debrecen.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.H.R.S. or M.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025