Research Article

Ji Zhou, Qiong Tian*, Ayaz Ahmad*, and Jiandong Huang*

Compressive and tensile strength estimation of sustainable geopolymer concrete using contemporary boosting ensemble techniques

https://doi.org/10.1515/rams-2024-0014
received November 27, 2023; accepted March 28, 2024

Abstract: Geopolymer concrete (GPC) serves as an environmentally conscious alternative to traditional concrete, offering a sustainable solution for construction needs. The ability to make on-site changes is dependent on the concrete's strength after casting, which must be higher than the target value. To anticipate the concrete's strength before it is poured is, thus, quite helpful. Three ensemble machine learning (ML) approaches, including gradient boosting, AdaBoost regressor, and extreme gradient boosting, are presented in this work as potential methods for forecasting GPC's mechanical strength that incorporates corncob ash. To determine which modeling parameters are crucial, sensitivity analysis was employed. When the compressive strength and split-tensile strength of GPC were tested with ensemble ML models, R^2 values of more than 90% were discovered between the predicted and actual results. Statistics and a *k*-fold analysis based on the error and coefficient of determination were used to verify the developed models. Slag amount, curing age, and fine aggregate quantity were the three mix proportions that had the most impact on GPC's mechanical strength, as shown in the sensitivity analysis. The results of this study demonstrated that ensemble boosting approaches could reliably estimate GPC mechanical strength. Incorporating such procedures into GPC quality control can yield significant improvements.

Keywords: geopolymer concrete, mechanical strength, boosting ensembles

*** Corresponding author: Qiong Tian**, College of Civil and Environmental Engineering, Hunan University of Science and Engineering, Yongzhou, 425006, China, e-mail: hnkjxytq@huse.edu.cn
*** Corresponding author: Ayaz Ahmad**, Department of Civil Engineering, COMSATS University Islamabad, Abbottabad, 22060, Pakistan, e-mail: ayazahmad@cuiatd.edu.pk
*** Corresponding author: Jiandong Huang**, School of Civil Engineering, Guangzhou University, Guangzhou, 510006, China, e-mail: jiandong.huang@hotmail.com
Ji Zhou: College of Civil and Environmental Engineering, Hunan University of Science and Engineering, Yongzhou, 425006, China

1 Introduction

The long-term effects of concrete's over-reliance as a construction material on the environment are becoming increasingly obvious [1]. Carbon emissions will rise, and biodiversity will be lost faster if the world's demand for cement and concrete triples by 2050, according to predictions [2]. Portland cement (PC) uses a lot of energy and produces a lot of carbon dioxide; thus, scientists have been looking for alternatives [2]. PC, the main ingredient in conventional concrete manufacture, requires around 1.80 metric tons of inputs and produces about 0.8 metric tons of greenhouse gases during its creation [3]. Consequently, the cement industry's impact on global warming must be mitigated without delay [2]. Recycling materials from farms and factories into new construction supplies is one technical and systematic way to reduce material impact [4–7]. Recycled agricultural and industrial waste can be transformed into supplementary cementitious materials, which have positive impacts on the environment, economy, and society at large [8,9]. One effective, inexpensive, and long-term way to reduce one's carbon footprint is to use recycled materials instead of PC [10].

By utilizing recycled agro-industrial resources in place of traditional PC, eco-friendly geopolymer concrete (GPC) renders cementitious binder unnecessary [11,12]. It suggests that the activation of aluminosilicate-based raw materials involves the use of alkali hydroxide/alkali silicate [13–15]. Geopolymer (alumino-silicate) materials can be made from a variety of reused farming and industrial products, including ground granulated blast furnace slag (GGBFS), fly ash (FAS), metakaolin (MK), red mud (RM), silica fume (SF), and rice husk ash (RHA) [16]. GGBFS offers a cost-effective and environmentally friendly alternative for GPC, demonstrating promising attributes for

eco-friendly and economical concrete solutions [17,18]. Nevertheless, corncob ash (CCA) stands out as an innovative component in this particular setting. Due to the fact that it contains a larger percentage of silica, CCA can be utilized as a substitute of or as a supplement to more typical pozzolanic materials such as RHA and FAS products. To avoid problems with on-site baking of GPC, researchers are looking into creating eco-friendly concrete using CCA at ambient temperatures. Considering a structure's resistance to environmental and other stresses is crucial when predicting its service life; achieving strength specifications alone is not enough. Because of its increased mechanical strength and longevity, GPC is being considered as a possible substitute for regular concrete in eco-conscious regions. According to multiple sources, GPC's exceptional mechanical qualities and durability are caused by its one-of-a-kind chemical makeup [16,19,20]. Incorporating nano-silica and recycled plastic particles has recently improved GPC's performance [21–23].

GPC is a greener option to regular concrete since it substitutes recycled agricultural and industrial resources for cement [11,12,24]. According to the findings, alkali hydroxide/alkali silicate is the key ingredient for activating aluminosilicate-based raw materials [13]. Reprocessed agronomic and industrial materials such as FAS, RM, GGBFS, MK, RHA, and SF are promising as precursors to geopolymers (alumino-silicates) [16,25–29]. For environmentally conscious and budget-conscious concrete [17,18], increased stiffness [30], and exceptional resistance to chemical assaults [31,32], GGBFS is an attractive option to GPC due to its reasonable cost-benefit and minimal environmental impact. On the other hand, CCA is an entirely novel component. The increased silica content of CCA makes it a suitable supplement or replacement for traditional pozzolanic materials like RHA and FAS. Scientists are trying to figure out a way to make this green concrete without utilizing an oven, so that it would not have any of the issues that come with green precast concrete. Equally important is realizing that strength requirements are not the be-all and end-all of performance evaluations. When determining how long a building will last, it is important to take its resistance to environmental and other stresses into account. In ecologically delicate regions, GPC is a good substitute for regular concrete due to its higher mechanical strengths and durability. The unique chemical makeup of GPC is the reason for its exceptional mechanical capabilities and endurance, according to all of the aforementioned sources [16,19,20]. Used in conjunction with recycled plastic particles, nano-silica has recently improved GPC performance [21–23].

Engineers, scientists, researchers, and computer programmers' efforts to develop new products and improve existing ones are being profoundly influenced by mathematical and artificial intelligence (AI) developments [33–36]. The demand for engineers who possess the skill to seamlessly incorporate numerical and AI-driven models into their routine tasks is on the rise due to the necessity to address a diverse range of challenges encountered in the field [37–41]. There are still a number of downsides and performance concerns with AI-based systems despite their potential. Recognizing objects and keeping up with discussions are two examples of the things that humans take for granted [42]. This makes it difficult for modern AI to devise suitable substitutes for teaching computers intuition. AI systems have successfully tackled these challenges by leveraging machine learning (ML) techniques [42–46]. Computers can learn to function autonomously by analyzing massive datasets using ML techniques [47]. In order to use the method with the most accurate data, you must first recover the features that characterize it. An acronym for "feature extraction" describes this method. The subsequent step involves training a ML algorithm utilizing the provided sample data transmitting features and instructions for pattern separation [42,48]. To address ever-increasingly complicated problems, current civil engineering research must use statistical approaches and AI. Statistical approaches and AI are frequently used by civil engineers to forecast the CS of materials [27,49–52]. These methods have been applied to a number of challenging situations, including the prediction of slump and strength of self-compacting concrete, the calculation of the axial bearing of different columns, the prediction of shear behavior of beams, and the prediction of chloride penetration [53–56]. The number of possible test configurations is reduced, making future research more efficient and cost-effective, thanks to these predictions. From Artificial neural networks to Gaussian process regression, gradient boosting, adaptive boosting, regression trees, decision trees (DTs), expression trees, support vector machines, and extreme gradient boosting (XGB), a number of ML techniques have been employed for forecasting concrete strength [57,58]. Out of all the models tested, the ensemble boosting models performed the best in predicting GPC's mechanical properties.

Using experimental data and AI algorithms, this study evaluated the mechanical strength of GPC made from slag and CCA. Using the three ensemble-boosting ML approaches, the study was able to accomplish its goals. It is possible to select the most effective model from among ten distinct submodels produced by the ensemble approach. One way to examine the models' approximative correctness is to compare the predicted and actual results using *k*-fold analysis and statistical testing. Experiments are labor-intensive because of the many steps needed, such as collecting

materials, casting samples, curing samples to strengthen them, and finally analyzing the results. Modern modeling approaches, like ML, applied to the construction sector could greatly help alleviate these problems. Traditional testing methods have a hard time picking up on the cumulative effect of all the variables on GPC strength. To determine which variables were most important, this study used sensitivity analysis. It is possible to collect the data needed by ML approaches from existing research. This dataset can be utilized for a variety of purposes, including running ML algorithms, estimating material quality, and doing impact studies. Using an experimental dataset, the article verifies that the ensemble ML approaches are useful for forecasting GPC strength. To find out how the raw ingredients and curing time affected the results, a sensitivity analysis was done. In order to promote sustainable building approaches, this study's findings may increase the construction industry's use of GPC.

2 Research strategy

2.1 Data collection and valuation

The aim of this study is to use ML ensemble models such as gradient boosting regressor (GBR), AdaBoost regressor (ABR), and extreme gradient boosting regressor (XGBR) to predict the compressive strength (CS) and split-tensile strength (STS) of GPCs manufactured from CCA and slag [59]. The total number of data points obtained from the experimental inquiry was 260. With the help of ten input variables, including fine aggregate (FA), CCA, blast furnace slag (BFS), concrete grade (CG), water (W), curing day (CD), sodium silicate gel (SSG), molar concentration (MC), and STS, the CS and STS of GPC were predicted. Information gathering and organization were accomplished through data preparation. A common task while attempting the well-known approach of knowledge discovery from data is preparing the data for mining. One major obstacle can be circumvented in this way. Preparing data entails removing extraneous information and background noise. As can be seen from Table 1, the number of descriptive statistics for the inputs and outputs is shown. A validation process was also utilized in order to assess the validity of the models that were utilized. Figure 1(a) illustrates the frequency distributions of the different values. One method for determining the frequency distribution of an entire dataset is to add up the distributions of its individual components and then determine the overall frequency distribution. In order to find out how often certain values appear, a relative frequency distribution can be built.

Table 1: Statistical descriptions of input variables [60]

Descriptive statistics	BFS (kg·m^{-3})	CCA (kg·m^{-3})	FA (kg·m^{-3})	CA (kg·m^{-3})	W (kg·m^{-3})	SHP (kg·m^{-3})	SSG (kg·m^{-3})	CD (days)	MC (M)	CG (MPa)	CS (MPa)	STS (MPa)
Median	228.00	195.00	841.00	1045.00	35.16	23.44	146.40	42.00	14.00	30.00	36.04	3.60
Standard error	9.54	9.48	3.53	0.01	0.13	0.13	0.00	1.93	0.10	0.31	0.76	0.04
Mean value	218.65	215.24	818.12	1044.99	35.22	23.38	146.40	45.25	14.00	34.96	35.90	3.65
Kurtosis	−1.12	−1.12	−1.41	127.47	−1.53	−1.53	127.47	−1.34	−1.53	−2.02	−0.74	0.58
Standard deviation	153.90	152.78	56.90	0.09	2.15	2.15	0.01	31.20	1.65	5.01	12.19	0.66
Range	488.00	488.00	171.00	1.00	5.22	5.22	0.10	83.00	4.00	10.00	53.42	4.57
Mode	0.00	0.00	841.00	1045.00	37.86	20.74	146.40	7.00	12.00	30.00	29.04	3.60
Skewness	0.12	0.13	−0.14	−11.34	0.04	−0.04	−11.34	0.25	0.00	0.02	0.17	0.20
Sample variance	23685.6	23341.7	3237.34	0.01	4.63	4.63	0.00	973.43	2.72	25.10	148.64	0.43
Minimum	0.00	0.00	728.00	1044.00	32.64	20.74	146.30	7.00	12.00	30.00	10.67	1.97
Maximum	488.00	488.00	899.00	1045.00	37.86	25.96	146.40	90.00	16.00	40.00	64.09	6.54
Sum	56850.0	55962.0	212712.0	271698.0	9157.44	6078.5	38063.8	11765	3640.0	9090.0	9334.0	949.4
Count	260.00	260.00	260.00	260.00	260.00	260.00	260.00	260.00	260.00	260.00	260.00	260.00

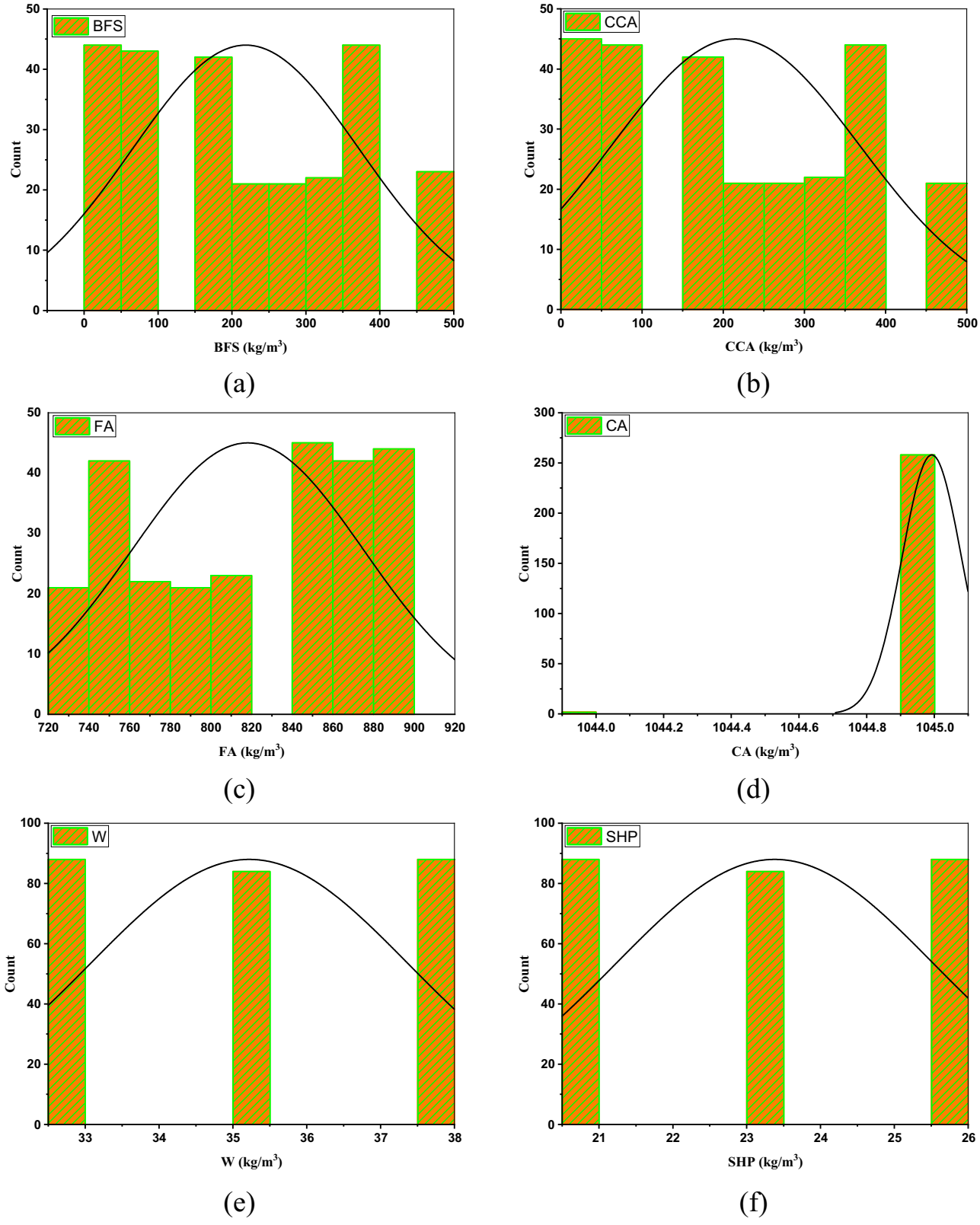
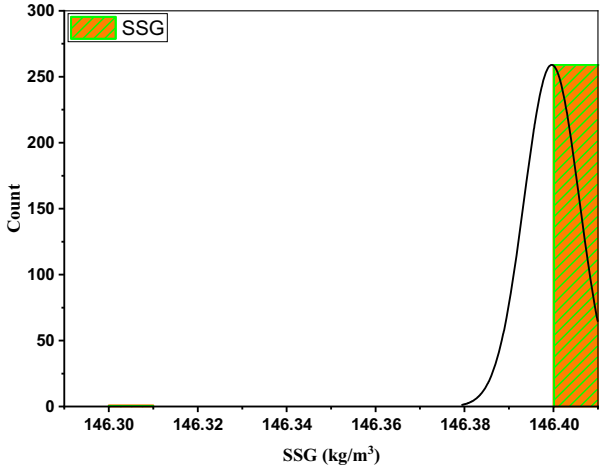
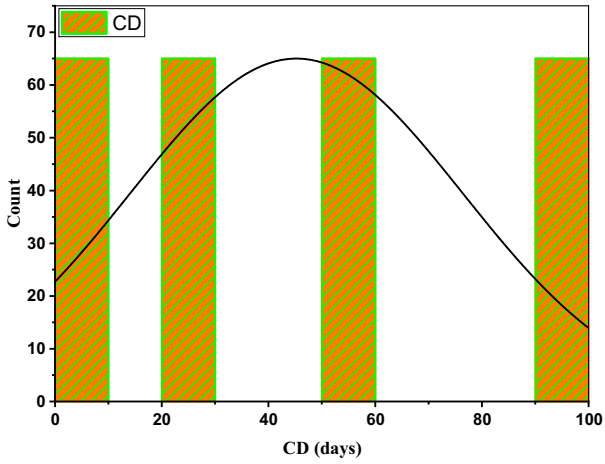


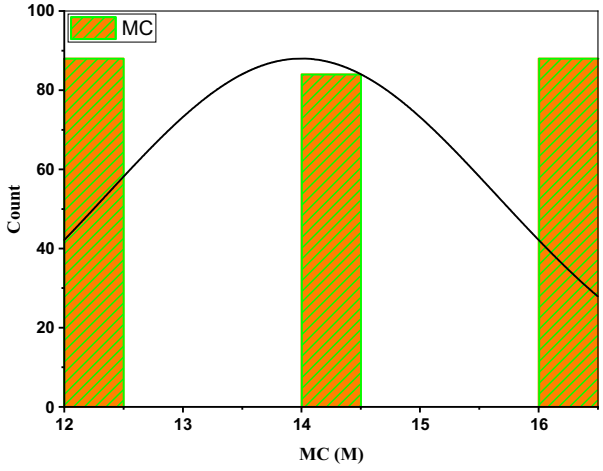
Figure 1: Database input/output frequency distribution: (a) BFS, (b) CCA, (c) FA, (d) coarse aggregate (CA), (e) W, (f) sodium silicate pellets (SHP), (g) SSG, (h) CD, (i) MC, (j) CG, (k) CS, and (l) STS [60].



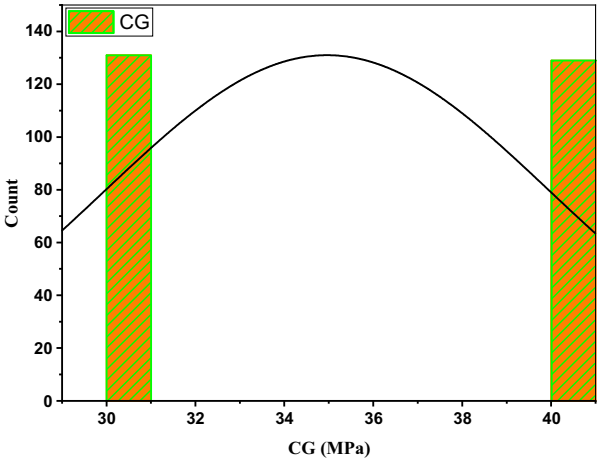
(g)



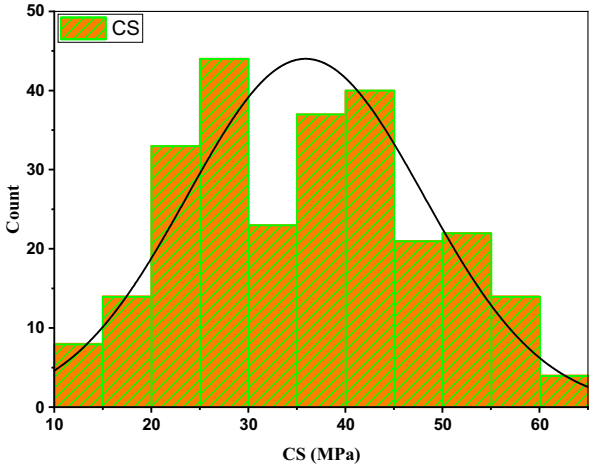
(h)



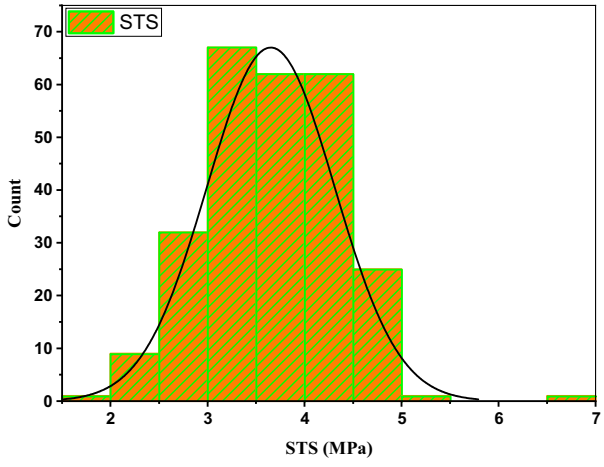
(i)



(j)



(k)



(l)

Figure 1: *Continued*

2.2 Simulations of ML

Research into GPC's mechanical properties took place in a controlled laboratory setting. Ten separate components were needed to make CS and STS. State-of-the-art ML techniques like GBR, ABR, and XGBR are utilized to forecast the GPC's CS and STS. The most common scenario for using ML algorithms is comparing the results to the inputs. Half of the data were used for testing ML models, while the other half were used for training. The data were partitioned at random using Python programming. The capacity to train ML models on subsets of data and then assess their generalizability on other subsets is known as data splitting, and it is a crucial part of ML model evaluation. By dividing the data in this way, we may prevent the model from being too specialized to our training set and evaluate its predictive power on new, unknown data. Other researchers conducting similar nature studies have also used data splitting [61–63]. The model's dependability was shown by the result's R^2 score. The R^2 score, which shows the amount of variation explained by the model, is one reliable statistic for measuring the prediction accuracy of a ML model. A low value for R^2 implies a significant discrepancy between the actual findings and the forecasts [64]. The model was proven accurate by multiple investigations, which included statistical tests and evaluations of errors. The research strategy's flowchart is displayed in Figure 2.

A model's learning process can be fine-tuned using hyperparameters. For supervised ML training (regression and classification), hyperparameter values are required [65]. To determine the optimal hyperparameter values, the M–L package's default setup or the user-adjusted trial-and-error method is utilized. User trial and error to obtain appropriate hyperparameter settings is arduous and time-consuming [66]. Hyperparameter optimization and tuning can save users time and effort when choosing the optimal ones for their model [67]. A key aspect of ML models is selecting the hyperparameter value that ensures both minimal loss and maximum accuracy [68]. A basic notion of the model's performance was obtained by testing and ranking all potential combinations of hyperparameters according to their effectiveness on each fold. The hyperparameters that were set throughout the modeling phase, along with their ideal and range values, are displayed in Table 2.

2.2.1 GBR

Friedman proposed employing ensemble methods for classifying and predicting data [69]. GBR, like other boosting methods, is limited in application to regression but otherwise the same. Some of the limitations of GBR are its computational complexity, its susceptibility to noisy data, and

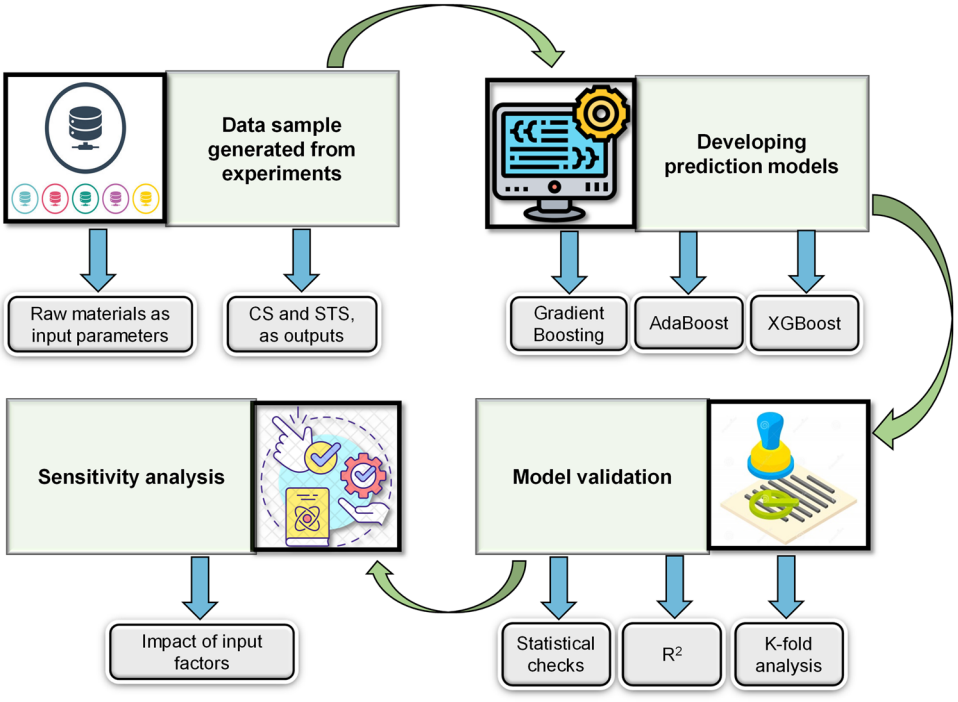


Figure 2: Flow diagram of the research strategy.

Table 2: Setting hyperparameters for ML models

Parameters	ABR		GBR		XGBR	
	Range	Optimal value	Range	Optimal value	Range	Optimal value
No. of estimators	10–200	70	10–200	150	50–1000	250
Learning rate	0.01–0.50	0.30	0.01–0.50	0.30	0–1	0.4
Max. depth	—	—	1–5	3	1, ∞	10
Max. features	—	—	0.8–1.0	1	0.8–1.0	1
Min. sample leaf	—	—	1–4	3	1–4	2
Min. sample split	—	—	2–10	6	2–10	6
colsample_bytree					0–1	0.5
Gamma					0, ∞	1
Reg_alpha					0, ∞	1
Reg_lambda					0, ∞	1

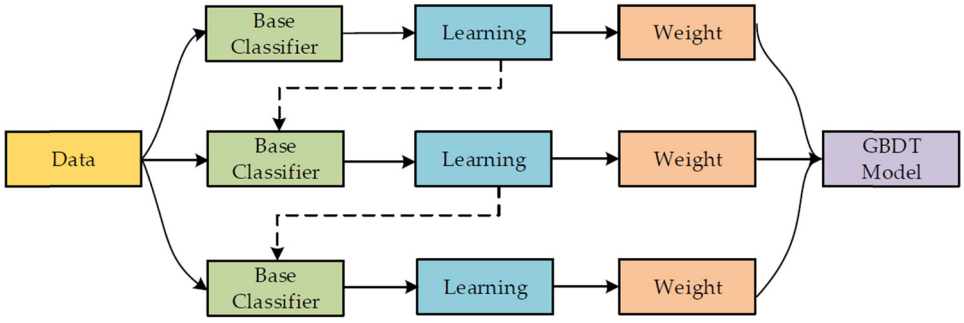


Figure 3: GBR for regressor training: A temporal study [70].

the necessity of meticulous hyperparameter adjustment. Using ensemble learning, sequential training, weighted instances, and aggregation of weak learners, GBR is similar to other boosting approaches such as ABR and XGBR. As shown in Figure 3, the technique randomly selects training set repetitions to compare to the reference model. By resampling with replacement from the original dataset, the GBR approach uses bootstrap sampling to randomly choose training set repeats. The model's performance can be evaluated across various training data configurations by comparing these repetitions to a reference model. By studying how different training sets affect the GBR's prediction abilities, this bootstrap-based method improves stability. To prevent over-fitting and improve GBR's accuracy and performance, randomly subsampling the training data is one option. Randomly subsampling GBR training data, commonly *via* bootstrap sampling, improves model training accuracy by diversifying training. This diversity helps the model fit varied data patterns. This reduces over-fitting by making the model less susceptible to noise and outliers in any subset, enabling robust learning and better performance on new data. Adjusting the shrinkage rate and n-tree size allows for fine-tuning of the GBR approach. An adequately large n-trees dataset is required due to the fact that the learning rate or shrinkage factor influences each expansion tree independently. Adjusting the GBR shrinking rate affects learning step size and model complexity-training speed balance. Adding trees fine-tunes the model's complexity, affecting its data pattern recognition. GBR model performance and stability improve with a big dataset for n-trees. A huge dataset provides various training instances for each ensemble tree, helping the algorithm generalize. This diversity helps capture data patterns and makes the GBR model more resilient and accurate.

2.2.2 ABR

ML specialists often employ the ensemble method for training several models with a single learning algorithm [71]. The two methods for developing models, GBR and ABR, are different. By shifting weights to concentrate on hard-to-predict samples, ABR sequentially highlights misclassified occurrences. GBR, on the other hand, optimizes

overall predictive performance by minimizing residuals by fitting each new model to the ensemble's errors. The ensemble is a collection of independent classification algorithms that are combined into a single decision. Students from all around the world are working towards the same goal and band together to find the solution. The ABR implements a managed ML approach called ensemble knowledge. Because the weights adjusted after each incidence is processed, adaptive boosting is another name for it; it gives greater weight to cases that were incorrectly labeled. After processing each instance, it modifies weights by giving misclassified examples a larger weight, increasing their influence in subsequent model iterations. By adjusting the weights, ABR is able to improve the model's overall performance by concentrating on cases that are hard to forecast. To lessen bias and variability, supervised ML frequently uses boosting techniques. To help a struggling student, these strategies are employed. During the training phase, you can use as many DTs as you want with the input data. Data that were wrongly classified throughout the entire original model is given priority throughout the construction of the model. Another model utilizes nothing but these numbers as input. When enough new primary school students have been generated, the process outlined above will begin again. The ABR is particularly effective in improving the enactment of DTs on two-fold arrangement issues. Its implementation of other machine-learning techniques is further enhanced by its usage. A doubter can get knowledge from it as well. Predicting the mechanical properties of concrete is one area where ensemble methods find extensive application in civil engineering. The comprehensive procedure for guessing the best ABR algorithmic outcome is shown in Figure 4.

2.2.3 XGBR

The XGBR method, created by Yeh, is a powerful tool for data science researchers because of its tree-based ensemble information strategy [73]. A variety of functions are used by the GBR architecture, which XGBR is built upon, to estimate outcomes in accordance with Eq. (1) [69].

$$\bar{y}_i = y_i^0 + \eta \sum_{K=1}^{n} f_k(U_i), \tag{1}$$

where null hypothesis y_i^0 represents the amount of expertise to enhance the accuracy and prevent over-fitting of a model through "linking" new trees; $\bar{y}_i$ exemplifies the result that can be anticipated using ith data and U_i as the restriction vector; η indicates the estimator sum in relation to tree topologies that are discrete as opposed to all f_k where k ranges from 1 to n. A major obstacle in ML is developing models with minimal overfitting of data. Subsequent trees are designed to rectify the residuals from this null hypothesis, which acts as a starting point. The model's predictions are made more accurate through this iterative corrective procedure. Regularization strategies, like tree trimming and managing the depth of the trees, are used by XGBR to prevent over-fitting. In order to promote a more generalized model, it also includes a shrinkage parameter that reduces the impact of each unique tree. The kth forecaster is connected to the method at the kth level and the estimated kth $\boldsymbol{y_i^{-(k-1)}}$ with the conforming produced f_k adjacent to the kth identical forecaster is found using Eq. (2). Whereas, where y_i^{-k} represents the predicted value at the kth level, the kth forecaster is connected to the method at the kth level through the addition of the product of the learning rate η and the kth

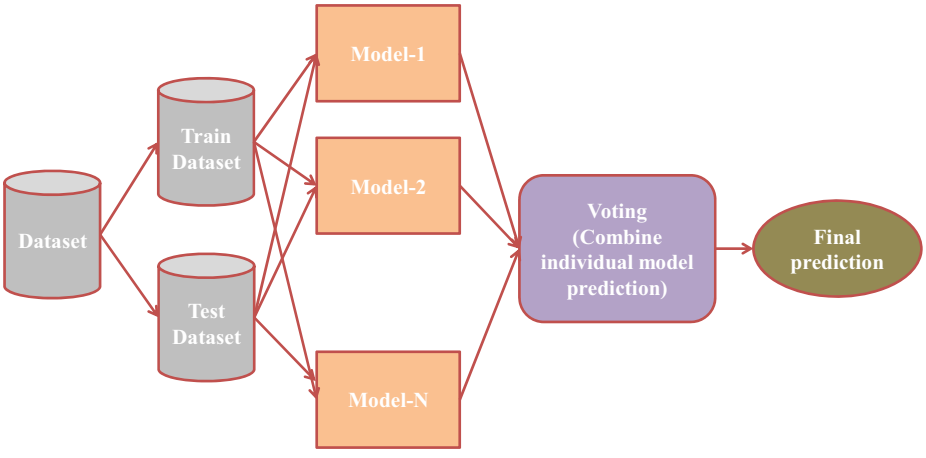


Figure 4: Schematic representation of the AdaBoost method for predicting forecasts [72].

weak learner's prediction f_k. This equation illustrates the iterative nature of the boosting algorithm, where each new weak learner is trained to correct the residuals of the combined ensemble up to the k^{-1}th level.

$$y_i^{-k} = y_i^{-(k-1)} + \eta f_k. \tag{2}$$

As shown in Eq. (3), the reduction of the kth truthful job of a tree results in a leaf with a weight of f_k.

$$f_{\text{obj}} = \gamma Z + \sum_{a=1}^{Z}\left[g_a\omega_a + \frac{1}{2}(h_a + \lambda)\omega_a^2\right], \tag{3}$$

where Z is the segment of leaf lumps, γ is the difficulty factor, λ is a persistent coefficient, and ω_a^2 is the weight of a leaf; these regulating limitations are utilized to avoid overfitting and enhance the model. Across the whole dataset, h_a and g_a provide a concise summary of the original gradient leaf loss function and its predecessor. Cutting a single leaf into many smaller leaves is required to construct the kth tree. Advanced factors are employed to produce such a framework, as shown in Eq. (4).

$$G = \frac{1}{2}\left[\frac{O_L^2}{P_L + \lambda} + \frac{O_R^2}{P_R + \lambda} + \frac{(O_L + O_R)^2}{P_L + P_R + \lambda}\right], \tag{4}$$

where P_R and O_R stand for the right leaf, whereas P_L and O_L stand for the left leaf, and G stands for gain variables. Typically, we ignore the gain parameter and only use the division criterion. An indirect regulatory variable denoted by λ is under the control of the gain. One way to stop the leaf convolution process is to increase the regularization value, which can significantly lower the gain parameter. However, including training data would cause the model's performance to decline. Figure 5 shows the top-level hierarchy of the XGB tree algorithm.

2.3 Validation of models

The accuracy of the ML models was verified using mathematical methods and k-fold procedures. One popular way to test how well a method works is with the k-fold strategy, which involves randomly splitting the dataset into ten categories [75]. As can be observed in Figure 6, ML simulations are taught utilizing nine different groups, but only one is used for verification. In cases with low error and high R^2, ML methods perform well. It takes ten iterations of the process to achieve the target outcome. The model's already impressive accuracy is further enhanced by this technique. Reducing the impact of variability in a single train-test split helps to generate a more robust evaluation of a ML model's performance, offering a more trustworthy estimate of its generalization performance. To compare the implementation of several ML algorithms, statistical error measures were utilized. These included root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). The ML algorithms' estimates were statistically tested using Eqs (5)–(7), which were obtained from previous works [76,77].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - T_i|, \tag{5}$$

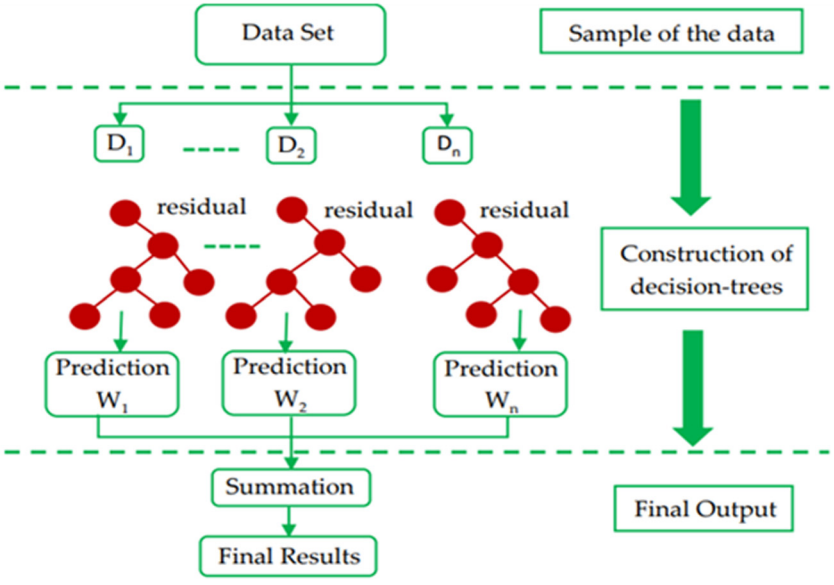


Figure 5: A straightforward flowchart outlining the XGBR process [74].

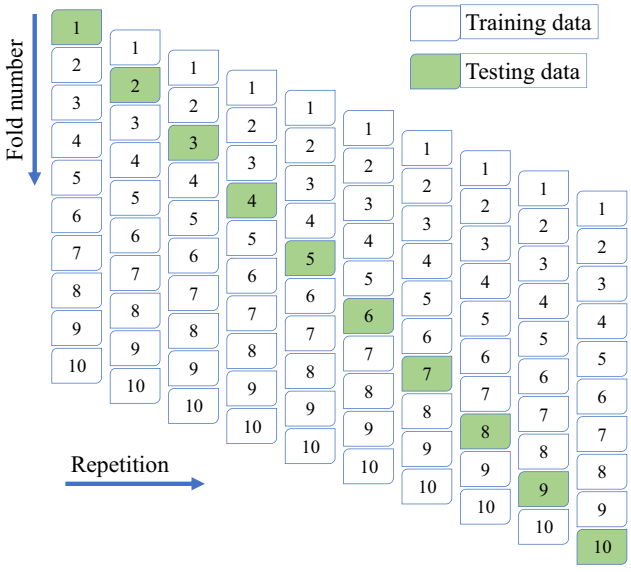


Figure 6: K-method operational procedure [83].

$$\text{RMSE} = \sqrt{\sum \frac{(P_i - T_i)^2}{n}}, \tag{6}$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|P_i - T_i|}{T_i}, \tag{7}$$

$$a20\text{-index} = \frac{m20}{n}, \tag{8}$$

where n is the number of observations, P_i are the expected results, and T_i are the measured values.

According to Eq. (8), retrieved from previous studies [78–80], these factors in an anticipated or experimental value range from 0.80 to 1.20, where M is the dataset size and m20 is the entry count. Based on the prediction model, a20-index values of 1% would be ideal. The 20-index that has been suggested provides the benefit of a physical engineering method by showing what proportion of samples match expected values within a ±20% uncertainty range of experimental data.

A model's predictive power can be best assessed with statistical validation and by consulting Taylor's diagram. This chart shows the trajectory of the models' departure from the truth, which can be used as a benchmark to assess their credibility and accuracy [81,82]. Three indicators of a model's proper placement are the RMSE, the correlation coefficient (radial line), and the standard deviation (circle centered at the real value point). Whichever model has the best track record of accurately predicting outcomes is the most reliable one [81].

3 Outcomes and evaluations

3.1 CS-ML models

3.1.1 CS-GBR model

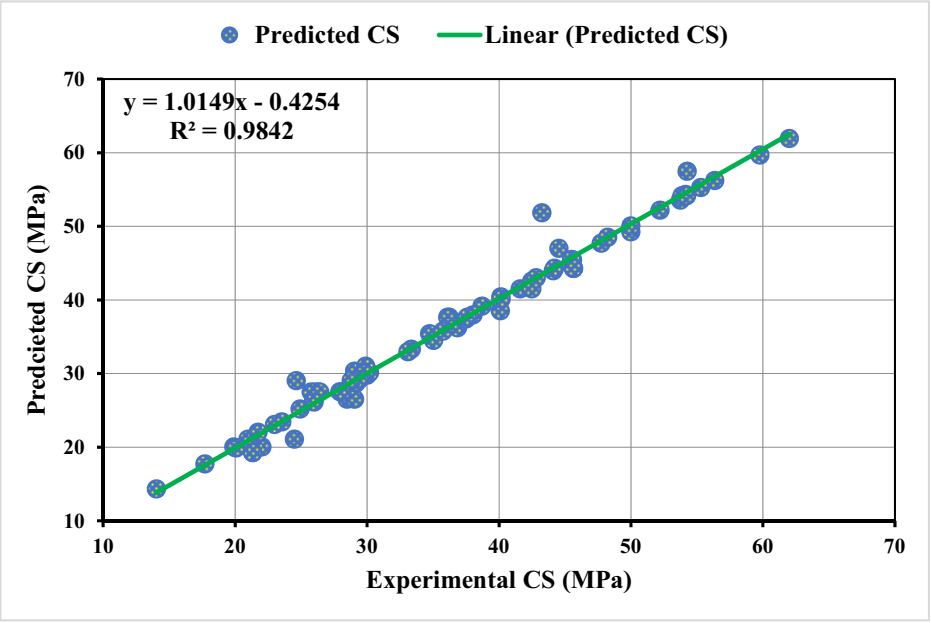
An approximation of the GPC CS using the GBR model is illustrated in Figure 7. Figure 7(a) provides a visual representation of the relationship between the actual and expected CS. The GBR model's CS predictions were remarkably close to the observed values. A GBR technique R^2 of 0.9842 indicated substantially greater accuracy in determining the GPC CS. In Figure 7(b), we can see the spread of discrepancies (errors) between the experimental and GBR-predicted values. On average, the incorrect readings were close to 0.80 MPa, with a standard variation of 0.00–8.58 MPa. And it was found that 56 of those numbers were less than 1 MPa, 18 were in the 1–3 MPa range, and 4 were beyond 3 MPa. A GBR model can be used to forecast the CS of GPC by examining the distribution of mistakes. This is the case even if the only reason this is possible is due to the improper partitioning of the data.

3.1.2 CS-ABR model

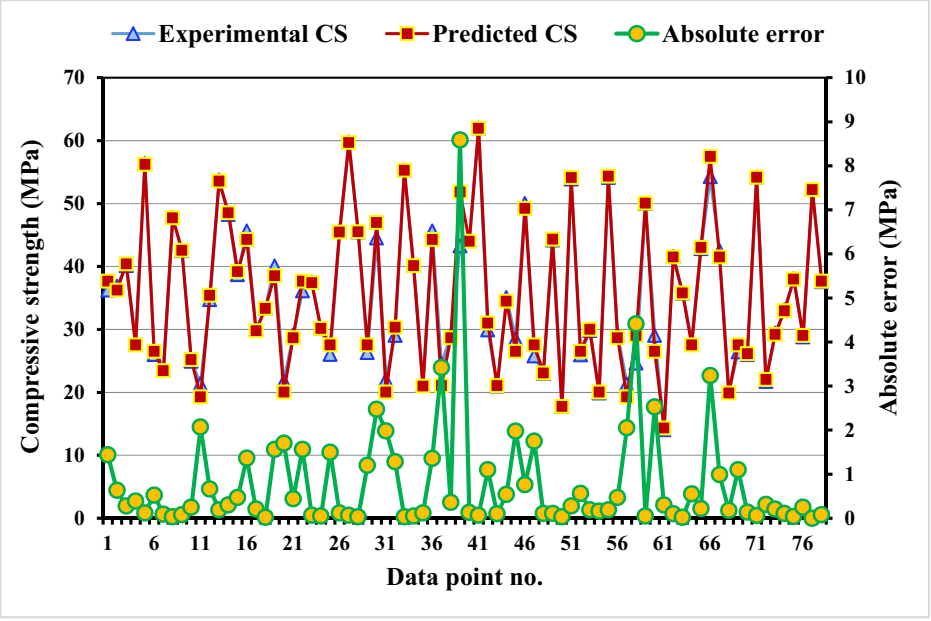
Figure 8 shows that the ABR model can be used to approximate GPC's CS. The graphic representation of the link between the actual and expected CS may be found in Figure 8. The ABR model closely matched the predicted and actual CS levels. With an R^2 value of 0.9937, the ABR method was significantly more effective than previous approaches in calculating the CS of GPC. The error range between the experimental values and the anticipated values using the ABR technique is displayed in Figure 8(b). The average error value was about 0.55 MPa on a scale that ran from 0.00 to 3.35 MPa. According to the error distribution, 63 of the readings were found to be less than 1 MPa, 13 to be between 1 and 3 MPa, and 1 to be above 3 MPa. The ABR model is preferable to the GBR model because of its smaller error margin.

3.1.3 CS-XGBR model

Figure 9 shows that the XGBR ensemble technique, which consists of ten sub-models, yielded obvious and positive results due to the low error distribution of the test data



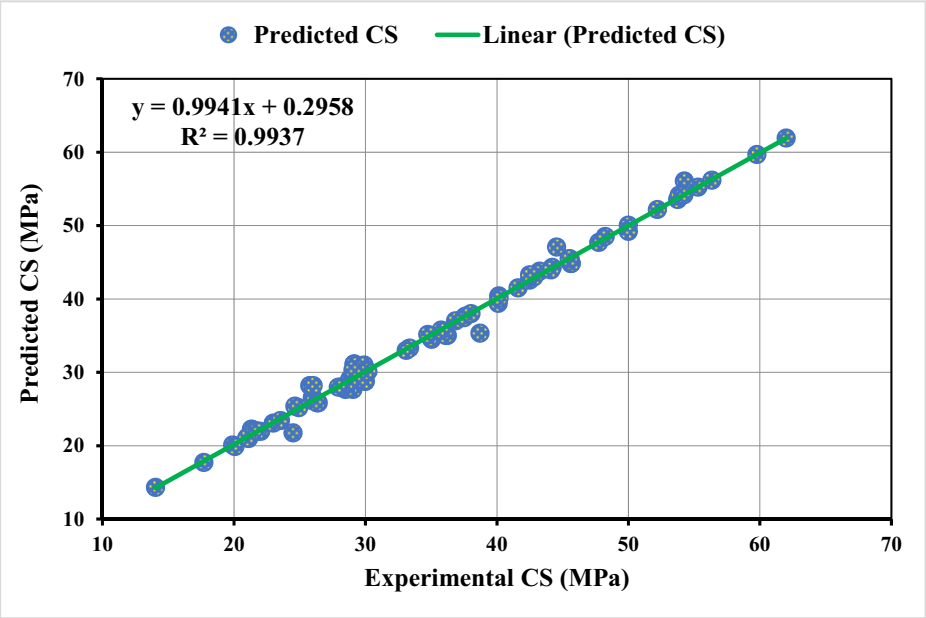
(a)



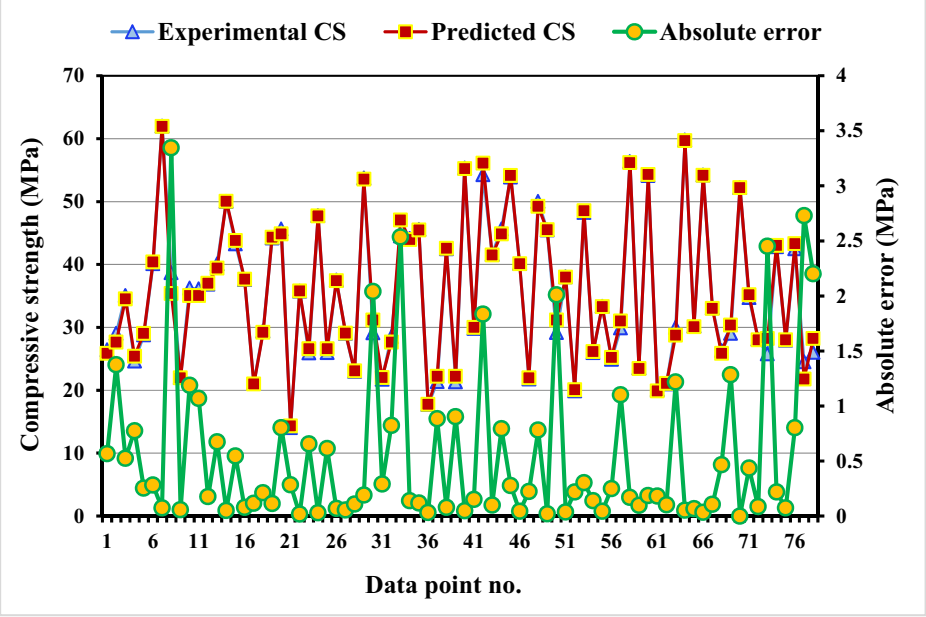
(b)

Figure 7: The CS-GBR model: (a) the link between the expected and experimental CS and (b) the dispersion of the errors and the predicted CS.

and the good regression correlation (R^2 = 0.9937). The test and the anticipated CS are shown to be related in Figure 9(a). The XGBR procedure outperformed its predictions when compared to the actual outcomes. The XGBR method's higher R^2 score of 0.9937 shows that it is more accurate. The distribution of the actual, estimated, and incorrect values produced by the XGBR approach is depicted in Figure 9(b). The XGBR method produced somewhat more precise findings than the GBR and ABR models; 64 errors were less than 1 MPa, 13 were between 1 and 3 MPa, and 1 was greater than 3 MPa. An inaccuracy ranging from 0.55 to 3.345 MPa was recorded as the standard deviation. Predictions of the CS of GPC concrete using the XGBR method are likely to be more accurate than those using the GBR or ABR methods. Nonetheless, the GBR and



(a)



(b)

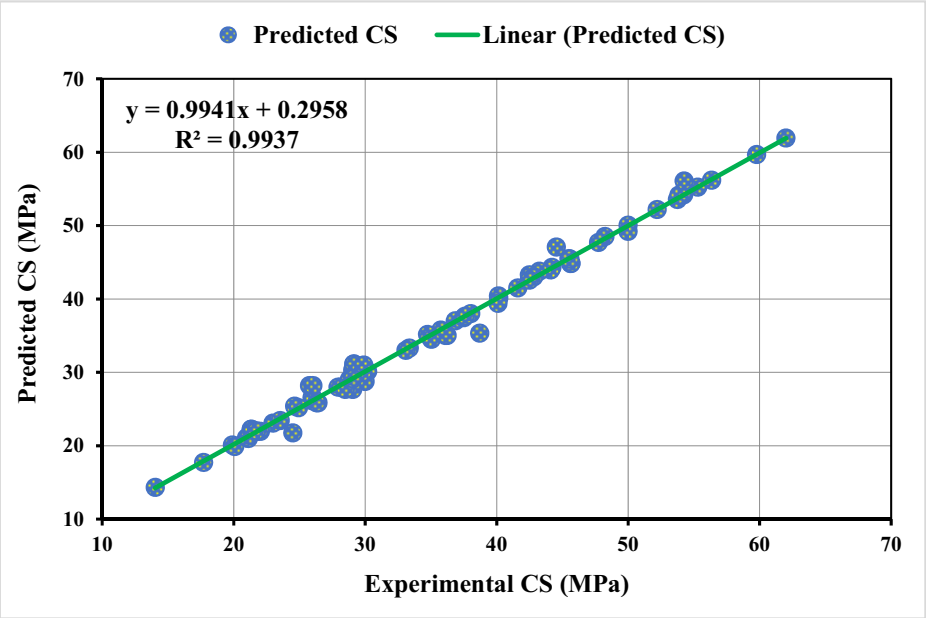
Figure 8: The CS-ABR model: (a) the link between the expected and experimental CS and (b) the dispersion of the errors and the predicted CS.

ABR models' precision was impressive. That being said, any model can be used to assess the CS of GPC.

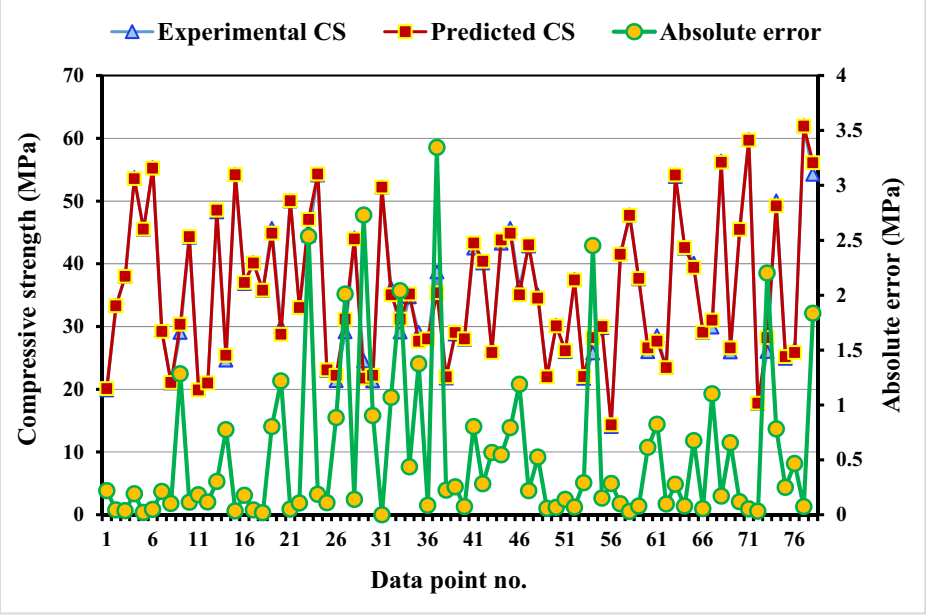
3.2 STS-ML models

3.2.1 STS-GBR model

As shown in Figure 10, the GBR model can be used to estimate the STS of GPC. The agreement between observed and predicted STS is graphically represented in Figure 10(a). The GBR model's estimations and the measured values of STS were found to be in good agreement. The STS of GPC was successfully estimated using the GBR method, with an R^2 value of 0.9387, suggesting significantly higher accuracy. Figure 10(b) shows the range of discrepancies (errors) between the experimental and GBR-predicted values. The standard deviation of the erroneous values was 0.131 MPa, with the range being from 0.005 to



(a)



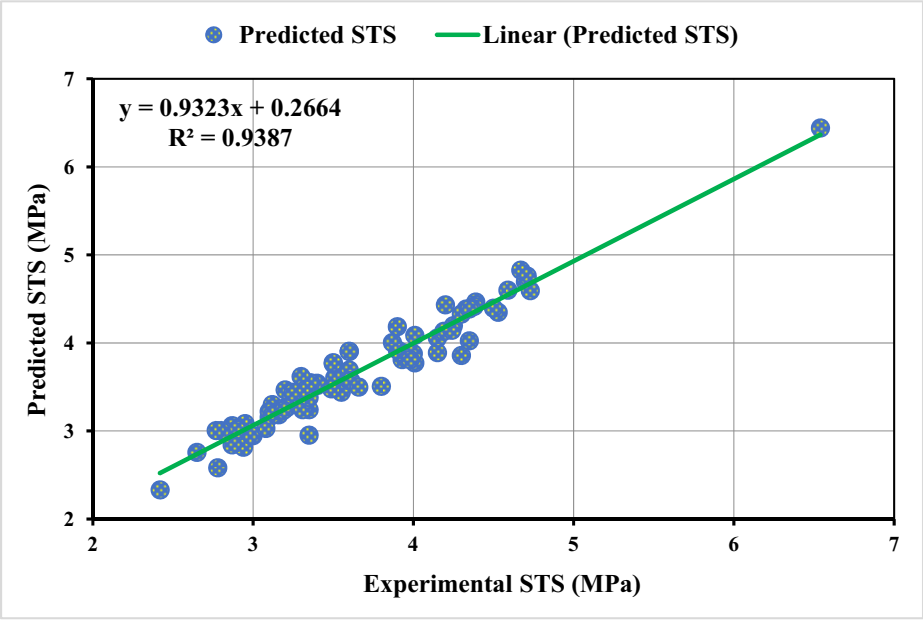
(b)

Figure 9: The CS-XGBR model: (a) the link between the expected and experimental CS and (b) the dispersion of the errors and the predicted CS.

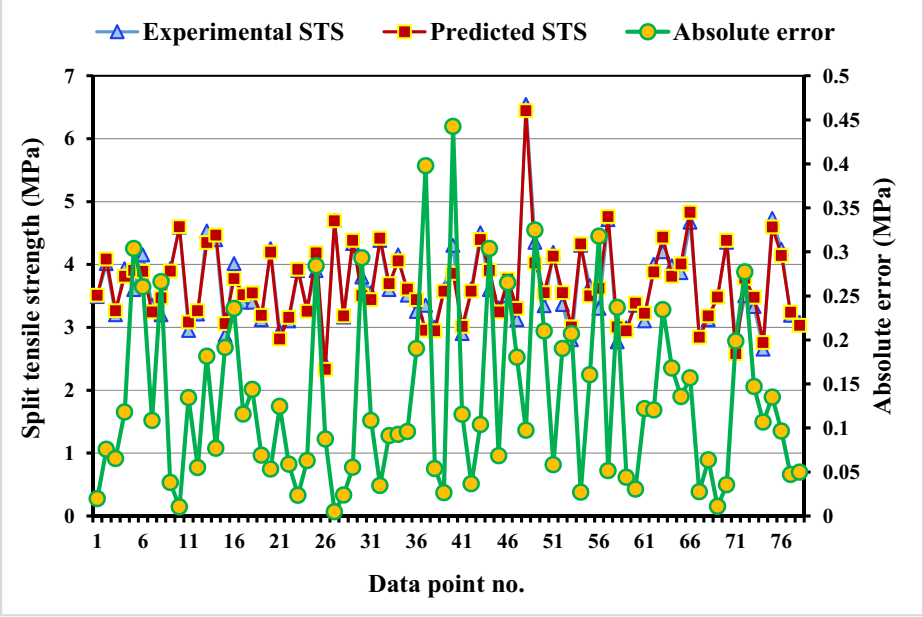
0.442 MPa. As can be seen from the error distribution, 37 values were found to be lower than 0.1 MPa, 35 to be in the range of 0.1–0.3 MPa, and 6 to be higher than 0.3 MPa. It is evident that the STS of GPC may be predicted using a GBR model, even though this is likely attributable to improper data partitioning.

3.2.2 STS-ABR model

As shown in Figure 11, the ABR model can be used to estimate the STS of GPC. Figure 11(a) shows a pictorial depiction of the agreement between observed and predicted STS. The predicted and observed values of STS agreed



(a)



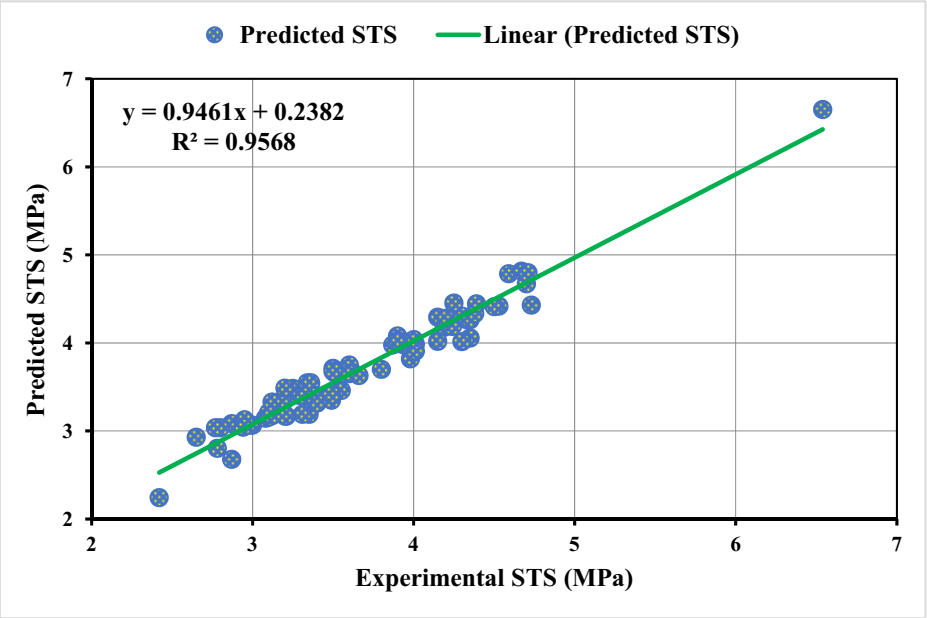
(b)

Figure 10: The STS-GBR model: (a) the link between the expected and experimental STS and (b) the dispersion of the errors and the predicted STS.

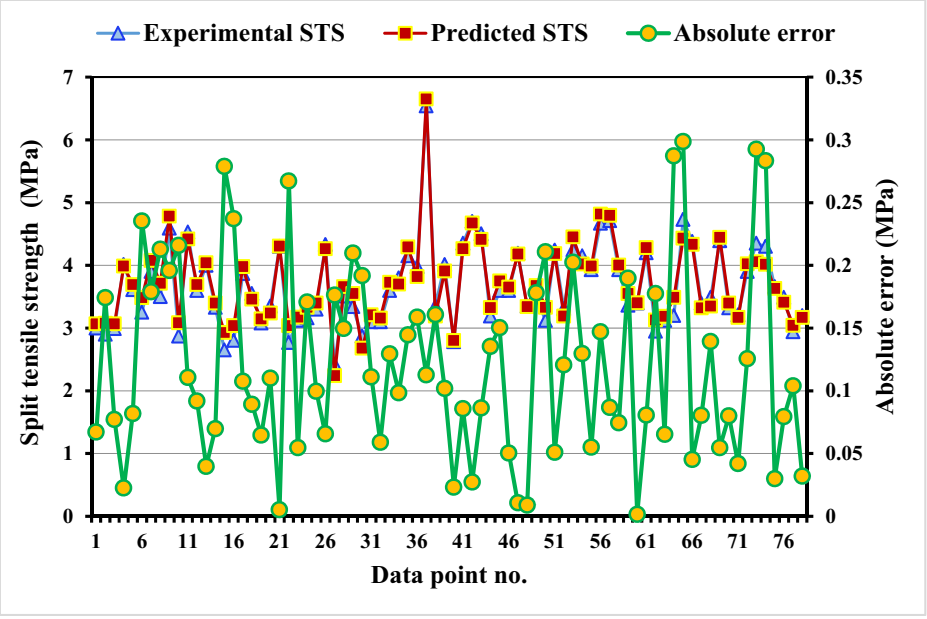
closely with each other, as predicted by the ABR model. Compared to the STS-GBR model, the STS of GPC was effectively predicted using the ABR method, with an R^2 value of 0.9568. This figure shows the experimental and projected values using the ABR approach, together with the range of errors (Figure 11(b)). With a standard deviation of 0.121 MPa, the inaccurate readings varied between 0.001 and 0.299 MPa. In addition, 37 figures were found to be below 0.1 MPa, 41 to be between 0.1 and 0.3 MPa, and 0 to be above 0.3 MPa when the spread of the inaccuracies was examined. In terms of STS prediction models, the ABR model is preferable to the GBR model due to its smaller error margin.

3.2.3 STS-XGBR model

Figure 12 shows that the XGBR ensemble method with ten sub-models produces clear and favorable results due to the



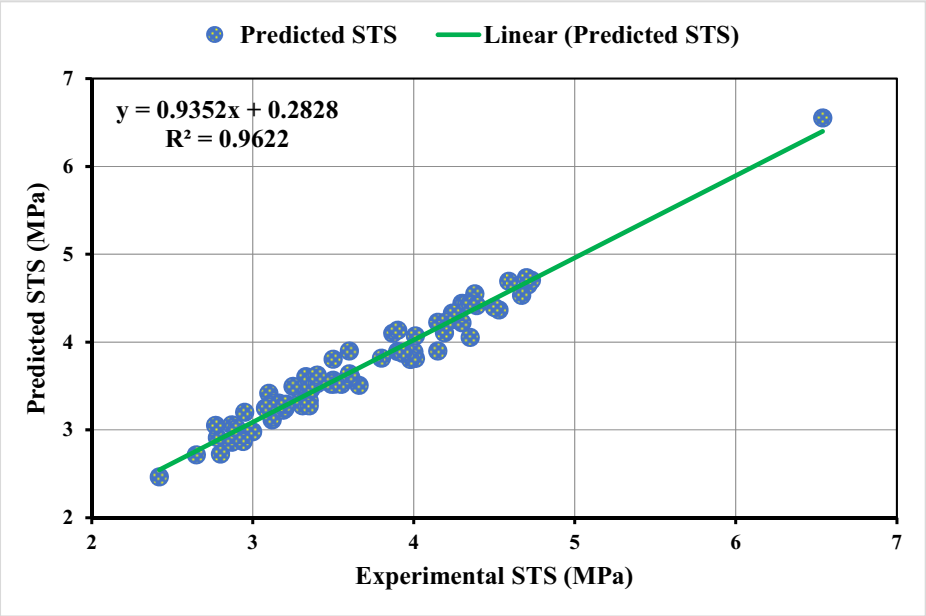
(a)



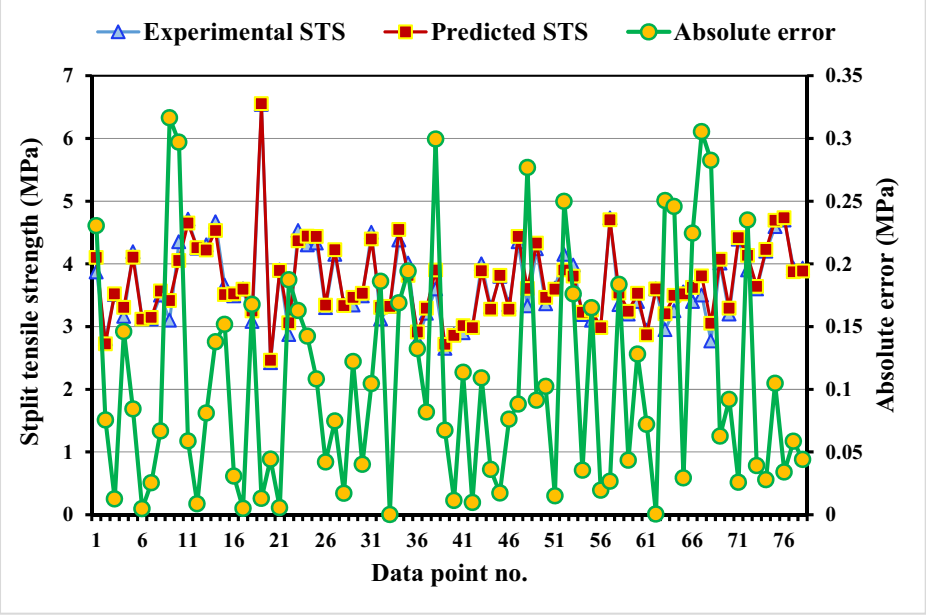
(b)

Figure 11: The STS-ABR model: (a) the link between the expected and experimental STS and (b) the dispersion of the errors and the predicted STS.

low test data error distribution and the high regression correlation of R^2 = 0.9622. Predicted STS is shown to be related to the test in Figure 12(a). When comparing observed and predicted results, the XGBR method proved to be more accurate. The XGBR method's higher R^2 score (0.9622) indicates that it is more accurate. Figure 12(b) shows the distribution of accurate, approximated, and incorrect figures using the XGBR approach. There were 44 errors below 0.1 MPa, 32 errors between 0.1 and 0.3 MPa, and 2 errors beyond 0.3 MPa in the results obtained using the somewhat more accurate XGBR approach. A maximum of 0.316 MPa was recorded, with an average inaccuracy of 0.105 MPa. When it comes to estimating the STS of GPC concrete, the XGBR method is expected to perform better than both the



(a)



(b)

Figure 12: The STS-XGBR model: (a) the link between the expected and experimental STS and (b) the dispersion of the errors and the predicted STS.

GBR and ABR approaches. Having said that, the ABR and GBR models were incredibly accurate. Hence, any model may be used to test the GPC's STS.

3.3 Substantiation of models

Table 3 shows the computed errors (MAE, RMSE, and MAPE) as a result of applying the aforementioned Eqs (5)–(7) to the STS-approximation and CS models. Predictions of CS using GBR, ABR, and XGBR had MAEs of 0.802, 0.557, and 0.557 MPa, respectively. GBR, ABR, and XGBR all improved performance by an average of 2.60, 1.80, and 1.80%, respectively, according to the MAPE metric. Additional investigation showed that XGBR had an RMSE of 0.915 MPa, GBR had an RMSE of 1.494 MPa, and ABR had an RMSE of 0.915 MPa. However, similar tendencies were found in STS prediction models for MAE, RMSE, and MAPE

Table 3: Evaluation of errors by statistical procedures

ML technique	CS				STS			
	MAPE (%)	MAE (MPa)	RMSE (MPa)	a20 index	MAPE (%)	MAE (MPa)	RMSE (MPa)	a20 index
GBR	2.60	0.802	1.494	0.90	3.70	0.131	0.164	0.88
ABR	1.80	0.557	0.915	0.94	3.50	0.121	0.143	0.92
XGBR	1.80	0.557	0.915	1.00	3.00	0.105	0.137	1.00

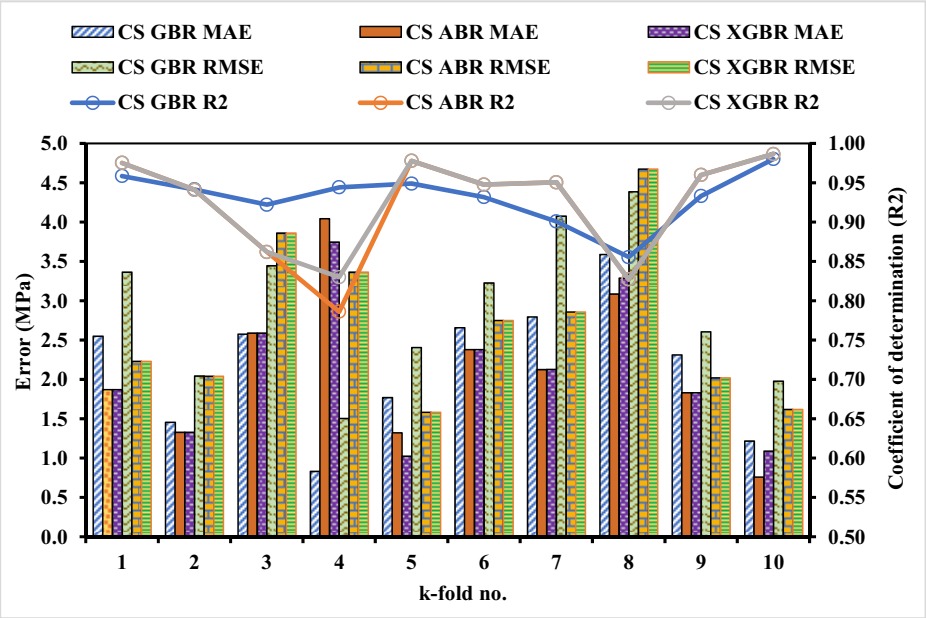
Table 4: Measures of accuracy (RMSE, R^2, and MAE) as per *k*-fold analysis

Property	ML model	Parameters	*k*-fold									
			1	2	3	4	5	6	7	8	9	10
CS	GBR	MAE	2.55	1.45	2.57	0.83	1.77	2.66	2.79	3.59	2.31	1.22
		RMSE	3.36	2.04	3.44	1.50	2.41	3.23	4.08	4.39	2.61	1.98
		R^2	0.96	0.94	0.92	0.94	0.95	0.93	0.90	0.86	0.93	0.98
	ABR	MAE	1.87	1.33	2.59	4.04	1.32	2.38	2.13	3.09	1.83	0.76
		RMSE	2.23	2.04	3.86	3.36	1.58	2.75	2.86	4.67	2.02	1.62
		R^2	0.98	0.94	0.86	0.00	0.98	0.95	0.95	0.83	0.96	0.99
	XGBR	MAE	1.87	1.33	2.59	3.74	1.02	2.38	2.13	3.29	1.83	1.09
		RMSE	2.23	2.04	3.86	3.36	1.58	2.75	2.86	4.67	2.02	1.62
		R^2	0.98	0.94	0.86	0.00	0.98	0.95	0.95	0.83	0.96	0.99
STS	GBR	MAE	0.34	0.33	0.26	0.31	0.58	0.35	0.25	0.13	0.29	0.43
		RMSE	0.58	0.58	0.58	0.36	0.73	0.31	0.26	0.28	0.18	0.33
		R^2	0.55	0.52	0.65	0.43	0.50	0.38	0.66	0.86	0.93	0.54
	ABR	MAE	0.65	0.31	0.23	0.11	0.37	0.37	0.18	0.12	0.16	0.19
		RMSE	0.66	0.35	0.34	0.11	0.65	0.43	0.19	0.12	0.20	0.19
		R^2	0.02	0.48	0.51	0.95	0.62	0.45	0.89	0.93	0.81	0.87
	XGBR	MAE	0.23	0.24	0.29	0.19	0.40	0.20	0.24	0.21	0.17	0.35
		RMSE	0.27	0.21	0.32	0.24	0.36	0.53	0.25	0.28	0.28	0.25
		R^2	0.76	0.44	0.14	0.76	0.65	0.38	0.68	0.96	0.49	0.45

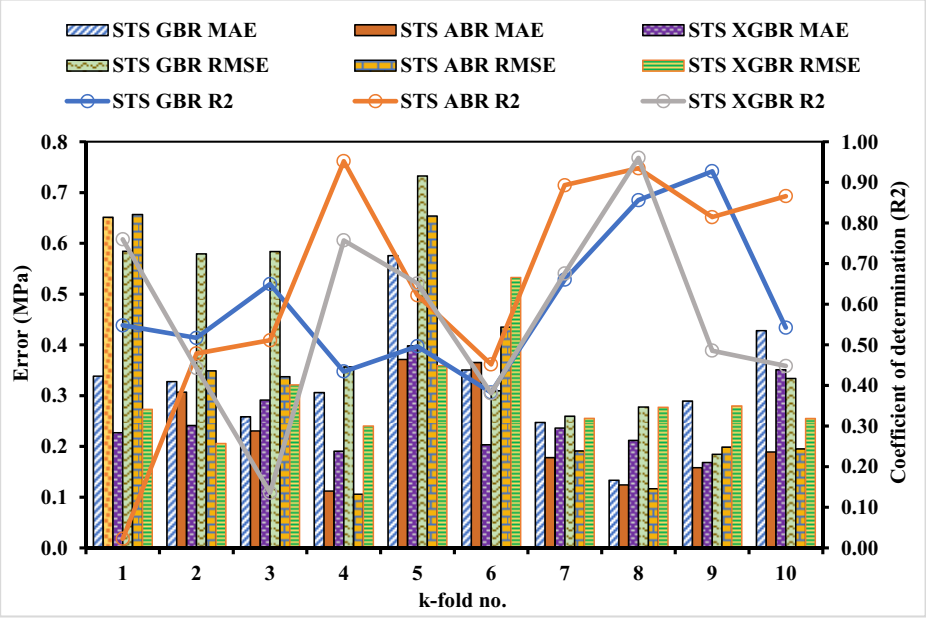
as seen in CS prediction models. The XGBR technique outperforms the GBR and ABR models, as shown by these data. To validate the *k*-fold approach, the results of determining R^2, RMSE, and MAE are provided in Table 4. Results from *k*-fold tests of several ML methods for STS and CS prediction are shown in Figure 13. An estimated CS with MAE values ranging from 0.83 to 3.59 (mean value: 2.17 MPa) was generated by the GBR method. An MAE of 2.13 MPa was observed in the ABR model, ranging from 0.76 MPa to 4.04 MPa. However, XGBR had an MAE of 2.13 MPa, ranging from 1.02 to 3.74 MPa. The RMSE values for the GBR, ABR, and XGBR methods were 2.90, 2.70, and 2.70 MPa, respectively. Nonetheless, the maximum R^2 values for GBR, ABR, and XGBR are 0.98, 0.99, and 0.99, respectively. STS prediction *k*-fold analysis results were also seen, with MAE and RMSE values decreasing dramatically from GBR to ABR to XGBR and R^2 increasing marginally across the same series of models. The maximum R^2 and lowest error rate may be found in the best XGBR model for predicting GPC's CS and STS. According to the results of the error analysis and *k*-fold R^2 measurements, the XGBR model is more accurate. However, both the GBR and ABR models performed admirably well. Because of this, GBR, ABR, and XGBR models may offer a more accurate way to evaluate the CS and STS of GPC. Moreover, Figure 14 demonstrates how Taylor's diagram can be used to depict the accuracy of prediction of ensemble models. The R^2 values for XGBR models are quite close to 1.0, which confirms their superior accuracy in predicting GPC's CS and STS compared to the other models.

3.4 Sensitivity analysis

This research primarily aims to examine the effect of different input features on the GPC predictions of STS and CS. The inputs have a strong correlation with the projected outcomes [72]. Offering a glimpse into the future of concrete, Figure 15 elucidates the influence of each constituent



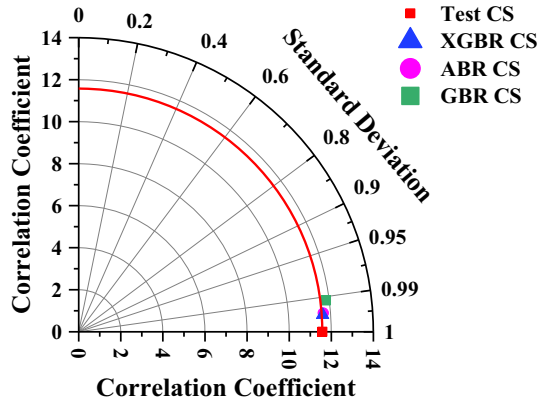
(a)



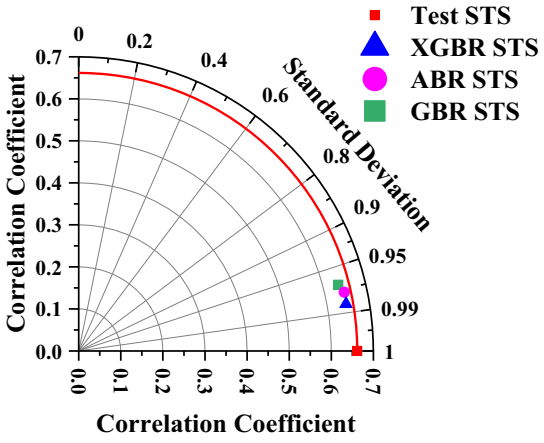
(b)

Figure 13: The results of the *k*-fold correlation: (a) CS and (b) STS.

on the mechanical properties of GPC. Concerning CS and STS, BFS emerged as the most pivotal factor, accounting for 65% of the influence, followed by curing duration at 25%. Factors such as FA, CG, CCA, water content, SHP, and MC contributed to a lesser extent, with percentages ranging from 0.50 to 3.30%. On the other hand, the sensitivity study showed that CA and SSG had no impact on GPC's mechanical characteristics, suggesting that the parameter dataset was quite inconsistent. This observation underscores the relationship between the number of model parameters and the quantity of data points utilized in the sensitivity analysis. The study's findings were sensitive to concealed input characteristics, such as the amounts of the concrete mix, which became obvious after applying the ML technique in the



(a)



(b)

Figure 14: Taylor's diagram. (a) CS and (b) STS.

analysis. The significance of the input variables was determined using Eqs (9) and (10):

$$N_i = f_{\max}(x_i) - f_{\min}(x_i), \tag{9}$$

$$S_i = \frac{N_i}{\sum_{j-i}^{n} N_j}, \tag{10}$$

where $f_{\max}(x_i)$ and $f_{\min}(x_i)$ are the highest and lowest possible predicted values for the ith output, respectively.

4 Discussion

In order to guarantee that the predictions are particular to GPC, the ensemble ML models used in this study can produce ten independent sub-models with fixed input parameters. The models reliably forecast strengths due to their shared testing process and consistent unit metrics. Models must produce independent sub-models to understand mix design and input parameter effects. Extra parameters in the assessment beyond the ten inputs may render the predicted models useless. If the data are not according to the model's specifications, it might not work as intended. Unreliable results may be produced by models due to erroneous or altered units of the input parameters. The accuracy of the models relies on consistent unit sizes.

Enhancing energy efficiency, predicting material strength, ensuring quality, assessing risks, and implementing predictive maintenance represent just a subset of the numerous applications of ML models within the

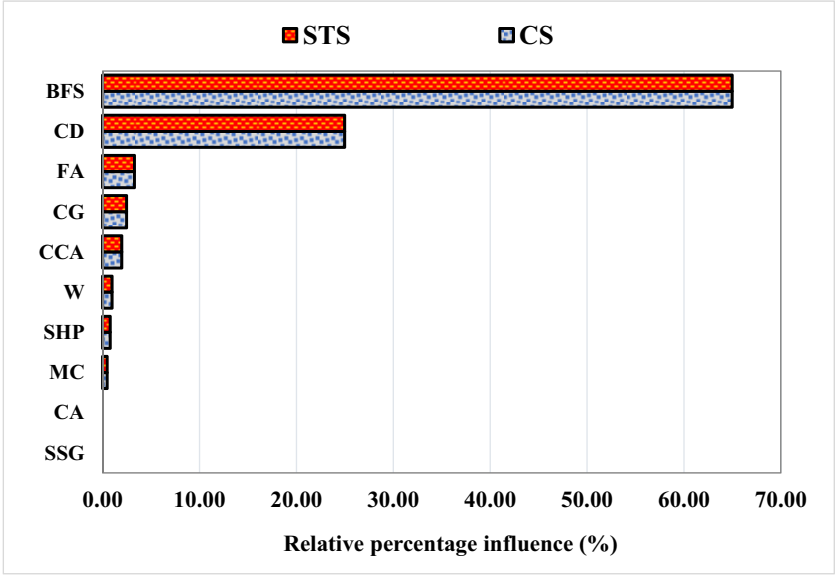


Figure 15: CS/STS sensitivity analysis.

construction industry. However, these models encounter certain limitations, such as reliance on human intervention, utilization of imperfect data, and the potential for inaccuracies in model outputs. To overcome these challenges, avenues for further research could include refining ML-based solutions through advancements in the Internet of Things, development of hybrid model approaches, utilization of explainable AI techniques, integration of sustainability considerations, and tailoring data collection and dissemination strategies to suit industry-specific needs. Project completion times could be reduced, and worker and environmental health and safety could be enhanced through data-driven decision-making, greater interpretability, and openness in the construction industry. This study's findings have the potential to promote GPC use among builders, leading to greener building practices. In addition, new areas of GPC, such as those involving nanotechnology-infused geopolymer and 3D geopolymer printing, could be investigated using ML approaches [84–89].

5 Conclusion

This work predicted the mechanical properties of GPC manufactured from slag and CCA using three ensemble ML models: GBR, ABR, and XGBR. The final models were trained and validated using CS and STS. Noteworthy study findings are as follows:

- The study discovered that XGBR models had exceptionally high data prediction accuracy when it came to forecasting the CS and STS of GPC.
- All three ensemble ML models (XGBR, ABR, and GBR) that were created to predict CS and STS for GPC reached R^2 values greater than 0.90.
- The constructed models were evaluated statistically using MAE, RMSE, and MAPE. When ML models have lower error values, it verifies their accuracy. It was established that XGBR models were exceptionally accurate in predicting GPC's CS and STS, which were further supported by the lower error rates.
- k-fold analysis (MAE, RMSE, and R^2) revealed that the XGBR model outperformed the GBR and ABR models when it came to accuracy.
- According to sensitivity analysis, BFS and CD had the strongest positive correlation with GPC CS and STS.

It is possible to understand the development of GPC by referring to the findings of the study, which are founded on robust forecasting frameworks. Researchers can speedily evaluate, improve, and rationalize GPC mixture proportioning using the study approach outlined in this article.

Acknowledgments: This research was supported by the Natural Science Foundation of Hunan (Grant no. 2023JJ50418) and the Hunan Provincial Transportation Technology Project (Grant no. 202109). The authors are grateful for this support.

Funding information: Natural Science Foundation of Hunan (Grant no. 2023JJ50418) and the Hunan Provincial Transportation Technology Project (Grant no. 202109).

Author contributions: J.Z.: conceptualization, formal analysis, data acquisition, methodology, and writing – original draft. Q.T.: supervision, project administration, resources, visualization, and writing – reviewing and editing. A.A.: conceptualization, software, methodology, supervision, and writing – original draft. J.H.: investigation, funding acquisition, validation, resources, and writing – reviewing and editing. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Ghosh, A. and G. D. Ransinchung. Application of machine learning algorithm to assess the efficacy of varying industrial wastes and curing methods on strength development of geopolymer concrete. *Construction and Building Materials*, Vol. 341, 2022, id. 127828.
[2] Belaïd, F. How does concrete and cement industry transformation contribute to mitigating climate change challenges? *Resources, Conservation and Recycling Advances*, Vol. 15, 2022, id. 200084.
[3] Andrew, R. M. Global CO_2 emissions from cement production, 1928–2018. *Earth System Science Data*, Vol. 11, 2019, pp. 1675–1710.
[4] Puertas, F., J. A. Suárez-Navarro, M. M. Alonso, and C. Gascó. NORM waste, cements, and concretes. A review. *Materiales de Construcción*, Vol. 71, 2021, id. e259.
[5] Khan, M., A. Rehman, and M. Ali. Efficiency of silica-fume content in plain and natural fiber reinforced concrete for concrete road. *Construction and Building Materials*, Vol. 244, 2020, id. 118382.
[6] Khan, M., M. Cao, A. Hussain, and S. H. Chu. Effect of silica-fume content on performance of $CaCO_3$ whisker and basalt fiber at matrix interface in cement-based composites. *Construction and Building Materials*, Vol. 300, 2021, id. 124046.
[7] Khan, M. and C. McNally. A holistic review on the contribution of civil engineers for driving sustainable concrete construction in the

built environment. *Developments in the Built Environment*, Vol. 16, 2023, id. 100273.
[8] Elmagarhe, A., Q. Lu, M. Alharthai, M. Alamri, and A. Elnihum. Performance of porous asphalt mixtures containing recycled concrete aggregate and fly ash. *Materials*, Vol. 15, 2022, id. 6363.
[9] Schaubroeck, T., T. Gibon, E. Igos, and E. Benetto. Sustainability assessment of circular economy over time: Modelling of finite and variable loops and impact distribution among related products. *Resources, Conservation and Recycling*, Vol. 168, 2021, id. 105319.
[10] Nurruddin, M. F., H. Sani, B. S. Mohammed, and I. Shaaban. Methods of curing geopolymer concrete: A review. *International Journal of Advanced and Applied Sciences*, Vol. 5, 2018, pp. 31–36.
[11] Oyebisi, S., A. Ede, F. Olutoge, and D. Omole. Geopolymer concrete incorporating agro-industrial wastes: Effects on mechanical properties, microstructural behaviour and mineralogical phases. *Construction and Building Materials*, Vol. 256, 2020, id. 119390.
[12] Oyebisi, S., A. Ede, F. Olutoge, and B. Ngene. Assessment of activity indexes on the splitting tensile strengthening of geopolymer concrete incorporating supplementary cementitious materials. *Materials Today Communications*, Vol. 24, 2020, id. 101356.
[13] Davidovits, J. Geopolymers: inorganic polymeric new materials. *Journal of Thermal Analysis and Calorimetry*, Vol. 37, 1991, pp. 1633–1656.
[14] Ahmad, M. R., C. S. Das, M. Khan, and J.-G. Dai. Development of low-carbon alkali-activated materials solely activated by flue gas residues (FGR) waste from incineration plants. *Journal of Cleaner Production*, Vol. 397, 2023, id. 136597.
[15] Riaz Ahmad, M., M. Khan, A. Wang, Z. Zhang, and J.-G. Dai. Alkali-activated materials partially activated using flue gas residues: An insight into reaction products. *Construction and Building Materials*, Vol. 371, 2023, id. 130760.
[16] Singh, B., G. Ishwarya, M. Gupta, and S. K. Bhattacharyya. Geopolymer concrete: A review of some recent developments. *Construction and Building Materials*, Vol. 85, 2015, pp. 78–90.
[17] Lenka, B. P., R. K. Majhi, S. Singh, and A. N. Nayak. Eco-friendly and cost-effective concrete utilizing high-volume blast furnace slag and demolition waste with lime. *European Journal of Environmental and Civil Engineering*, Vol. 26, 2022, pp. 5351–5373.
[18] Majhi, R. K. and A. N. Nayak. Production of sustainable concrete utilising high-volume blast furnace slag and recycled aggregate with lime activator. *Journal of Cleaner Production*, Vol. 255, 2020, id. 120188.
[19] Zakka, W. P., N. H. A. S. Lim, and M. C. Khun. A scientometric review of geopolymer concrete. *Journal of Cleaner Production*, Vol. 280, 2021, id. 124353.
[20] Farooq, F., X. Jin, M. F. Javed, A. Akbar, M. I. Shah, F. Aslam, and et al. Geopolymer concrete as sustainable material: A state of the art review. *Construction and Building Materials*, Vol. 306, 2021, id. 124762.
[21] Ahmed, H. U., A. A. Mohammed, and A. S. Mohammed. Effectiveness of silicon dioxide nanoparticles (Nano SiO_2) on the internal structures, electrical conductivity, and elevated temperature behaviors of geopolymer concrete composites. *Journal of Inorganic and Organometallic Polymers and Materials*, Vol. 33, 2023, pp. 1–21.
[22] Ahmed, H. U., A. S. Mohammed, and A. A. Mohammed. Engineering properties of geopolymer concrete composites incorporated recycled plastic aggregates modified with nano-silica. *Journal of Building Engineering*, Vol. 75, 2023, id. 106942.
[23] Ahmed, H. U., A. A. Mohammed, and A. S. Mohammed. Effectiveness of nano-SiO_2 on the mechanical, durability, and microstructural behavior of geopolymer concrete at different curing ages. *Archives of Civil and Mechanical Engineering*, Vol. 23, 2023, pp. 1–28.
[24] Lao, J.-C., B.-T. Huang, L.-Y. Xu, M. Khan, Y. Fang, and J.-G. Dai. Seawater sea-sand Engineered Geopolymer Composites (EGC) with high strength and high ductility. *Cement and Concrete Composites*, Vol. 138, 2023, id. 104998.
[25] Pazouki, G. Fly ash-based geopolymer concrete's compressive strength estimation by applying artificial intelligence methods. *Measurement*, Vol. 203, 2022, id. 111916.
[26] He, J., Y. Jie, J. Zhang, Y. Yu, and G. Zhang. Synthesis and characterization of red mud and rice husk ash-based geopolymer composites. *Cement and Concrete Composites*, Vol. 37, 2013, pp. 108–118.
[27] Peng, Y. and C. Unluer. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Construction and Building Materials*, Vol. 316, 2022, id. 125785.
[28] Shahmansouri, A. A., M. Yazdani, S. Ghanbari, H. A. Bengar, A. Jafari, and H. F. Ghatte. Artificial neural network model to predict the compressive strength of eco-friendly geopolymer concrete incorporating silica fume and natural zeolite. *Journal of Cleaner Production*, Vol. 279, 2021, id. 123697.
[29] Zhang, C., Z. Zhu, F. Liu, Y. Yang, Y. Wan, W. Huo, et al. Efficient machine learning method for evaluating compressive strength of cement stabilized soft soil. *Construction and Building Materials*, Vol. 392, 2023, id. 131887.
[30] Revilla-Cuesta, V., V. Ortega-López, M. Skaf, and J. M. Manso. Deformational behavior of self-compacting concrete containing recycled aggregate, slag cement and green powders under compression and bending: Description and prediction adjustment. *Journal of Building Engineering*, Vol. 54, 2022, id. 104611.
[31] Ortega-López, V., F. Faleschini, C. Pellegrino, V. Revilla-Cuesta, and J. M. Manso. Validation of slag-binder fiber-reinforced self-compacting concrete with slag aggregate under field conditions: Durability and real strength development. *Construction and Building Materials*, Vol. 320, 2022, id. 126280.
[32] Majhi, R. K., A. N. Nayak, and B. B. Mukharjee. Characterization of lime activated recycled aggregate concrete with high-volume ground granulated blast furnace slag. *Construction and Building Materials*, Vol. 259, 2020, id. 119882.
[33] Wang, H., X. Zhang, and S. Jiang. A laboratory and field universal estimation method for tire–pavement interaction noise (TPIN) based on 3D image technology. *Sustainability*, Vol. 14, 2022, id. 12066.
[34] Li, D., J.-H. Nie, H. Wang, and W.-X. Ren. Loading condition monitoring of high-strength bolt connections based on physics-guided deep learning of acoustic emission data. *Mechanical systems and signal processing*, Vol. 206, 2024, id. 110908.
[35] Huang, H., M. Guo, W. Zhang, J. Zeng, K. Yang, and H. Bai. Numerical investigation on the bearing capacity of RC columns strengthened by HPFL-BSP under combined loadings. *Journal of Building Engineering*, Vol. 39, 2021, id. 102266.
[36] Lin, J.-X., G. Chen, H.-S. Pan, Y.-C. Wang, Y.-C. Guo, and Z.-X. Jiang. Analysis of stress-strain behavior in engineered geopolymer composites reinforced with hybrid PE-PP fibers: A focus on cracking characteristics. *Composite Structures*, Vol. 323, 2023, id. 117437.
[37] Zhang, X., G. Zhou, X. Liu, Y. Fan, E. Meng, J. Yang, et al. Experimental and numerical analysis of seismic behaviour for recycled aggregate concrete filled circular steel tube frames. *Computers and Concrete*, Vol. 31, 2023, id. 537.

[38] Zhang, X., X. Liu, S. Zhang, J. Wang, L. Fu, J. Yang, et al. Analysis on displacement-based seismic design method of recycled aggregate concrete-filled square steel tube frame structures. *Structural Concrete*, Vol. 24, 2023, pp. 3461–3475.
[39] He, H., E. Shuang, T. Wen, J. Yao, X. Wang, C. He, et al. Employing novel N-doped graphene quantum dots to improve chloride binding of cement. *Construction and Building Materials*, Vol. 401, 2023, id. 132944.
[40] Cao, J., H. He, Y. Zhang, W. Zhao, Z. Yan, and H. Zhu. Crack detection in ultrahigh-performance concrete using robust principal component analysis and characteristic evaluation in the frequency domain. *Structural Health Monitoring*, Vol. 23, 2023, id. 14759217231178457.
[41] Wang, M., X. Yang, and W. Wang. Establishing a 3D aggregates database from X-ray CT scans of bulk concrete. *Construction and Building Materials*, Vol. 315, 2022, id. 125740.
[42] Avci, O., O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical systems and signal processing*, Vol. 147, 2021, id. 107077.
[43] Huang, J., M. Zhou, J. Zhang, J. Ren, N. I. Vatin, and M. M. S. Sabri. Development of a new stacking model to evaluate the strength parameters of concrete samples in laboratory. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, Vol. 46, 2022, pp. 4355–4370.
[44] Huang, J. and J. Xue. Optimization of SVR functions for flyrock evaluation in mine blasting operations. *Environmental Earth Sciences*, Vol. 81, 2022, id. 434.
[45] Huang, J., M. Zhou, J. Zhang, J. Ren, N. I. Vatin, and M. M. S. Sabri. The use of GA and PSO in evaluating the shear strength of steel fiber reinforced concrete beams. *KSCE Journal of Civil Engineering*, Vol. 26, 2022, pp. 3918–3931.
[46] Huang, J., J. Zhang, X. Li, Y. Qiao, R. Zhang, and G. S. Kumar. Investigating the effects of ensemble and weight optimization approaches on neural networks' performance to estimate the dynamic modulus of asphalt concrete. *Road Materials and Pavement Design*, Vol. 24, 2023, pp. 1939–1959.
[47] Wang, R., J. Zhang, Y. Lu, and J. Huang. Towards Designing Durable Sculptural Elements: Ensemble Learning in Predicting Compressive Strength of Fiber-Reinforced Nano-Silica Modified Concrete. *Buildings*, Vol. 14, 2024, id. 396.
[48] Marani, A. and M. L. Nehdi. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. *Construction and Building Materials*, Vol. 265, 2020, id. 120286.
[49] Khan, M., J. Lao, and J.-G. Dai. Comparative study of advanced computational techniques for estimating the compressive strength of UHPC. *Journal of Asian Concrete Federation*, Vol. 8, 2022, pp. 51–68.
[50] Essam, A., S. A. Mostafa, M. Khan, and A. M. Tahwia. Modified particle packing approach for optimizing waste marble powder as a cement substitute in high-performance concrete. *Construction and Building Materials*, Vol. 409, 2023, id. 133845.
[51] Nguyen, N.-H., T. P. Vo, S. Lee, and P. G. Asteris. Heuristic algorithm-based semi-empirical formulas for estimating the compressive strength of the normal and high performance concrete. *Construction and Building Materials*, Vol. 304, 2021, id. 124467.
[52] Emad, W., A. S. Mohammed, A. Bras, P. G. Asteris, R. Kurda, Z. Muhammed, et al. Metamodel techniques to estimate the compressive strength of UHPFRC using various mix proportions and a high range of curing temperatures. *Construction and Building Materials*, Vol. 349, 2022, id. 128737.
[53] Timur Cihan, M. Prediction of concrete compressive strength and slump by machine learning methods. *Advances in Civil Engineering*, Vol. 2019, 2019, pp. 1–11.
[54] Raza, A., Q. Z. Khan, and Ahmad. A. Prediction of axial compressive strength for FRP-confined concrete compression members. *KSCE Journal of Civil Engineering*, Vol. 24, 2020, pp. 2099–2109.
[55] Mansour, M. Y., M. Dicleli, J.-Y. Lee, and J. Zhang. Predicting the shear strength of reinforced concrete beams using artificial neural networks. *Engineering Structures*, Vol. 26, 2004, pp. 781–799.
[56] Tamimi, A. K., J. A. Abdalla, and Z. I. Sakka. Prediction of long term chloride diffusion of concrete in harsh environment. *Construction and Building Materials*, Vol. 22, 2008, pp. 829–836.
[57] Nazar, S., J. Yang, A. Ahmad, and S. F. A. Shah. Comparative study of evolutionary artificial intelligence approaches to predict the rheological properties of fresh concrete. *Materials Today Communications*, Vol. 32, 2022, id. 103964.
[58] Moein, M. M., A. Saradar, K. Rahmati, S. H. G. Mousavinejad, J. Bristow, V. Aramali, and et al. Predictive models for concrete properties using machine learning and deep learning approaches: A review. *Journal of Building Engineering*, Vol. 63, 2022, id. 105444.
[59] Oyebisi, S. and T. Alomayri. Artificial intelligence-based prediction of strengths of slag-ash-based geopolymer concrete using deep neural networks. *Construction and Building Materials*, Vol. 400, 2023, id. 132606.
[60] Zhou, J., Q. Tian, S. Nazar, and J. Huang. Hyper-tuning gene expression programming to develop interpretable prediction models for the strength of corncob ash-modified geopolymer concrete. *Materials Today Communications*, Vol. 38, 2024, id. 107885.
[61] Cao, Q., X. Yuan, M. Nasir Amin, W. Ahmad, F. Althoey, and F. Alsharari. A soft-computing-based modeling approach for predicting acid resistance of waste-derived cementitious composites. *Construction and Building Materials*, Vol. 407, 2023, id. 133540.
[62] Chen, Z., M. N. Amin, B. Iftikhar, W. Ahmad, F. Althoey, and F. Alsharari. Predictive modelling for the acid resistance of cement-based composites modified with eggshell and glass waste for sustainable and resilient building materials. *Journal of Building Engineering*, Vol. 76, 2023, id. 107325.
[63] Khan, K., W. Ahmad, M. N. Amin, and A. F. Deifalla. Investigating the feasibility of using waste eggshells in cement-based materials for sustainable construction. *Journal of Materials Research and Technology*, Vol. 23, 2023, pp. 4059–4074.
[64] Lee, B. C. and D. M. Brooks. Accurate and efficient regression modeling for microarchitectural performance and power prediction. *ACM SIGOPS operating systems review*, Vol. 40, 2006, pp. 185–194.
[65] Yang, L. and A. Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, Vol. 415, 2020, pp. 295–316.
[66] Singh, S., S. K. Patro, and S. K. Parhi. Evolutionary optimization of machine learning algorithm hyperparameters for strength prediction of high-performance concrete. *Asian Journal of Civil Engineering*, 2023, pp. 1–23.
[67] Thornton, C., F. Hutter, H. H. Hoos, and K. Leyton-Brown. *Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms*. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, USA, 2013, pp. 847–855.
[68] Yeh, I. C. and L.-C. Lien. Knowledge discovery of concrete material using Genetic Operation Trees. *Expert Systems with Applications*, Vol. 36, 2009, pp. 5807–5812.

[69] Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, Vol. 29, 2001, pp. 1189–1232.
[70] Yao, M., Y. Zhu, J. Li, H. Wei, and P. He. Research on predicting line loss rate in low voltage distribution network based on gradient boosting decision tree. *Energies*, Vol. 12, 2019, id. 2522.
[71] Ribeiro, M. H. D. M. and L. dos Santos Coelho. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, Vol. 86, 2020, id. 105837.
[72] Ahmad, A., W. Ahmad, K. Chaiyasarn, K. A. Ostrowski, F. Aslam, P. Zajdel, and et al. Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. *Polymers*, Vol. 13, 2021, id. 3389.
[73] Yeh, I. C. Prediction of strength of fly ash and slag concrete by the use of artificial neural networks. *J Chin Inst Civil Hydraul Eng*, Vol. 15, 2003, pp. 659–663.
[74] Amjad, M., I. Ahmad, M. Ahmad, P. Wróblewski, P. Kamiński, and U. Amjad. Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. *Applied Sciences*, Vol. 12, 2022, id. 2126.
[75] Ahmad, A., K. Chaiyasarn, F. Farooq, W. Ahmad, S. Suparp, and F. Aslam. Compressive Strength Prediction via Gene Expression Programming (GEP) and Artificial Neural Network (ANN) for Concrete Containing RCA. *Buildings*, Vol. 11, 2021, id. 324.
[76] Farooq, F., W. Ahmed, A. Akbar, F. Aslam, and R. Alyousef. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*, Vol. 292, 2021, id. 126032.
[77] Aslam, F., F. Farooq, M. N. Amin, K. Khan, A. Waheed, A. Akbar, et al. Applications of gene expression programming for estimating compressive strength of high-strength concrete. *Advances in Civil Engineering*, Vol. 2020, 2020, pp. 1–23.
[78] Apostolopoulou, M., P. G. Asteris, D. J. Armaghani, M. G. Douvika, P. B. Lourenço, L. Cavaleri, et al. Mapping and holistic design of natural hydraulic lime mortars. *Cement and concrete research*, Vol. 136, 2020, id. 106167.
[79] Asteris, P. G., M. Koopialipoor, D. J. Armaghani, E. A. Kotsonis, and P. B. Lourenço. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Computing and Applications*, Vol. 33, 2021, pp. 13089–13121.
[80] Asteris, P. G., P. B. Lourenço, M. Hajihassani, C.-E. N. Adami, M. E. Lemonis, A. D. Skentou, et al. Soft computing-based models for the prediction of masonry compressive strength. *Engineering Structures*, Vol. 248, 2021, id. 113276.
[81] Band, S. S., E. Heggy, S. M. Bateni, H. Karami, M. Rabiee, S. Samadianfard, et al. Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression. *Engineering Applications of Computational Fluid Mechanics*, Vol. 15, 2021, pp. 1147–1158.
[82] Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of geophysical research: atmospheres*, Vol. 106, 2001, pp. 7183–7192.
[83] Amin, M. N., W. Ahmad, K. Khan, S. Nazar, A. M. A. Arab, and A. F. Deifalla. Evaluating the relevance of eggshell and glass powder for cement-based materials using machine learning and SHapley Additive exPlanations (SHAP) analysis. *Case Studies in Construction Materials*, Vol. 19, 2023, id. e02278.
[84] Shilar, F. A., S. V. Ganachari, V. B. Patil, T. M. Y. Khan, N. M. Almakayeel, and S. Alghamdi. Review on the relationship between nano modifications of geopolymer concrete and their structural characteristics. *Polymers*, Vol. 14, 2022, id. 1421.
[85] Shilar, F. A., S. V. Ganachari, and V. B. Patil. Advancement of nano-based construction materials-A review. *Construction and Building Materials*, Vol. 359, 2022, id. 129535.
[86] Shilar, F. A., S. V. Ganachari, V. B. Patil, B. E. Bhojaraja, T. M. Y. Khan, and N. Almakayeel. A review of 3D printing of geopolymer composites for structural and functional applications. *Construction and Building Materials*, Vol. 400, 2023, id. 132869.
[87] Chen, Y., K. Xia, Z. Jia, Y. Gao, Z. Zhang, and Y. Zhang. Extending applicability of 3D-printable geopolymer to large-scale printing scenario via combination of sodium carbonate and nano-silica. *Cement and Concrete Composites*, Vol. 145, 2024, id. 105322.
[88] Saadati, F. and E. N. Kani. Phosphorous slag-based geopolymer cement incorporate with mullite for 3D printing application. *Construction and Building Materials*, Vol. 406, 2023, id. 133444.
[89] Yao, X., X. Lyu, J. Sun, B. Wang, Y. Wang, M. Yang, et al. AI-based performance prediction for 3D-printed concrete considering anisotropy and steam curing condition. *Construction and Building Materials*, Vol. 375, 2023, 130898.