



OPEN

# Data-driven prediction on critical mechanical properties of engineered cementitious composites based on machine learning

Shuangquan Qing<sup>1✉</sup> & Chuanxi Li<sup>1,2</sup>

The present study introduces a novel approach utilizing machine learning techniques to predict the crucial mechanical properties of engineered cementitious composites (ECCs), spanning from typical to exceptionally high strength levels. These properties, including compressive strength, flexural strength, tensile strength, and tensile strain capacity, can not only be predicted but also precisely estimated. The investigation encompassed a meticulous compilation and examination of 1532 datasets sourced from pertinent research. Four machine learning algorithms, linear regression (LR), K nearest neighbors (KNN), random forest (RF), and extreme gradient boosting (XGB), were used to establish the prediction model of ECC mechanical properties and determine the optimal model. The optimal model was utilized to employ SHapley Additive exPlanations (SHAP) for scrutinizing feature importance and conducting an in-depth parametric analysis. Subsequently, a comprehensive control strategy was devised for ECC mechanical properties. This strategy can provide actionable guidance for ECC design, equipping engineers and professionals in civil engineering and material science to make informed decisions throughout their design endeavors. The results show that the RF model demonstrated the highest prediction accuracy for compressive strength and flexural strength, with  $R^2$  values of 0.92 and 0.91 on the test set. The XGB model outperformed in predicting tensile strength and tensile strain capacity, with  $R^2$  values of 0.87 and 0.80 on the test set, respectively. The prediction of tensile strain capacity was the least accurate. Meanwhile, the MAE of the tensile strain capacity was a mere 0.84%, smaller than the variability (1.77%) of the test results in previous research. Compressive strength and tensile strength demonstrated high sensitivity to variations in both water-cement ratio ( $W$ ) and water reducer ( $WR$ ). In contrast, flexural strength exhibited high sensitivity solely to changes in  $W$ . Conversely, the sensitivity of tensile strain capacity to input features was moderate and consistent. The mechanical attributes of ECC emerged from the combined effects of multiple positive and negative features. Notably,  $WR$  exerted the most significant influence on compressive strength among all features, whereas polyethylene (PE) fiber emerged as the primary driver affecting flexural strength, tensile strength, and tensile strain capacity.

**Keywords** Engineered cementitious composites, Machine learning, Mechanical properties, Random forest, XGBoost

Engineered cementitious composites (ECCs) represent a class of ultra-high-performance cementitious composites that have significantly transformed the landscape of civil engineering. Since its inception by Li and colleagues in 1992, ECC has garnered substantial interest owing to its notable deformability and superior crack control capabilities<sup>1,2</sup>. Numerous experimental studies on mechanical properties of ECC have been conducted worldwide, such as uniaxial compression<sup>3</sup>, uniaxial tension<sup>4</sup>, flexural<sup>5</sup>, shear<sup>6</sup>, fatigue<sup>7</sup> and impact<sup>8</sup>, as well as a large number of structural or engineering applications<sup>9–15</sup>. The importance of comprehending and forecasting

<sup>1</sup>Department of Civil Engineering, Changsha University of Science & Technology, Changsha 410114, China. <sup>2</sup>State Key Laboratory of Featured Metal Materials and Life-Cycle Safety for Composite Structures, Nanning 530004, China. ✉email: doc\_qing@stu.csust.edu.cn

the mechanical characteristics of ECCs cannot be overstressed, as they directly influence the design and functionality of civil engineering structures.

Prior studies have often necessitated numerous experimental trials to attain an optimal ECC mix ratio. This process not only consumes considerable time but also demands substantial financial resources. Li introduced strength and energy criteria as a means to ensure ECC materials achieve a stable state of cracking, thereby mitigating potential errors in testing<sup>16</sup>. Based on numerous test results, empirical ranges for the strength and energy criteria have been established for various types of fibers<sup>16,17</sup>. The double criteria are generally obtained by means of a three-point flexural test and a single-crack tensile test<sup>18,19</sup>. However, the double criterion can only be used to determine whether a given mixture ratio can attain multiple cracks and cannot predict the mechanical properties of the mixture ratio.

Machine learning belongs to a sub-direction of artificial intelligence, including supervised and unsupervised<sup>20</sup>. Machine learning has a wide range of applications in the strength prediction of civil engineering materials. Hu used six machine learning models to predict the compressive strength of FRP-concrete confined column and compared them with prevailing mathematical or empirical models. SHAP analysis was utilized to ascertain feature importance and model sensitivity. The findings indicate that the XGB model exhibited superior robustness and accuracy<sup>21</sup>. At the same time, the problem of small experimental datasets was solved by using an innovative method of data augmentation, facilitating the achievement of a higher prediction accuracy for the punching shear strength of steel fiber-reinforced concrete slabs<sup>22</sup>. Jayasinghe used eight machine-learning models to develop a framework for predicting the shear capacity of recycled aggregate concrete (RAC) beams, and the results revealed that the XGBoost model had the best prediction effect, with the coefficient of determination  $R^2$  on the test set reaching 0.95 (slender beam) and 0.78 (deep beam)<sup>23</sup>. You utilized 48 datasets and three machine learning models to predict the bond strength between UHPC and deformed reinforcing bars. Among these models, the Adaptive Neuro-Fuzzy Interface System (ANFIS) emerged as the most accurate prediction model. The coefficient of determination  $R^2$  on the test set reached 0.98<sup>24</sup>. Similarly, Zhang predicted FRP and concrete interfacial bond strength with six machine learning models. Here, the XGBoost model performed best, and the root mean square error was reduced by 73% compared with the original empirical formula<sup>25</sup>. Liu predicted the creep compliance of concrete with four machine-learning models and found that the XGBoost was the optimal prediction model, with the  $R^2$  on the test set reaching 92%<sup>26</sup>. Mai predicted the compressive strength of fiber-reinforced self-compacting concrete with three machine-learning models. Among the three models, the categorical gradient boosting (CGB) model was the best, and the coefficient of determination was 0.986 for the test dataset<sup>27</sup>. Zhu used an artificial neural network (ANN) and support vector machine (SVM) to predict the 7-day compressive strength of ultrahigh-performance concrete (UHPC). The findings indicate that the ANN model yields superior prediction results compared to the SVM model<sup>28</sup>.

Several scholars have recently used machine learning to predict ECC strength. For instance, Shanmugasundaram used the ANN model to predict the compressive strength of ECC mixed with fly ash and GGBS, respectively. Here, the minimum and maximum differences between the predicted and experimental values were 0.317% and 11.4%, respectively<sup>29</sup>. Mahjoubi employed three machine learning models to predict the compressive strength, tensile strength, and ductility of SHCC. The results showed that XGBoost performed best, and the coefficients of determination on the test set were 0.95, 0.97 and 0.93, respectively<sup>30</sup>. Meng employed an artificial neural network algorithm to forecast both the compressive strength and tensile strength of PVA-ECC. The coefficients of determination on the test set were notably high at 0.98 for compressive strength and 0.99 for tensile strength, indicating significant prediction accuracy. However, the dataset in the described study was relatively small, with compressive strength ranging from 21.3 to 75.2 MPa and tensile strength ranging from 1.56 to 5.81 MPa<sup>31</sup>.

Moreover, the dataset used for machine learning is limited in scope, as it does not encompass high-strength and ultra-high-strength datasets. The algorithms with better results are XGBoost and artificial neural networks. However, artificial neural networks are prone to overfitting, and their training process is intricate. Despite this, there is a noticeable scarcity of predictions for the tensile strain capacity and flexural strength of ECC materials. Existing studies primarily focus on predicting compressive and tensile strength, predominantly within the PVA-ECC domain. The preparation of ECC may use PE fibers, PP fibers, and PVA fibers, which are also divided into oil-film PVA fibers and oil-free PVA fibers. The influence of specimen size on ECC mechanical properties remains unexplored in machine learning literature, indicating a gap in comprehensive understanding within prior research.

Based on a more extensive range of data sets, the aim of the present study was to predict critical mechanical properties of ECC, including compressive strength, flexural strength, tensile strength, and tensile strain capacity. At the same time, the importance of the characteristics affecting the mechanical properties of ECC was explored. Finally, a design strategy for adjusting ECC components was proposed according to the specified mechanical properties.

## Data processing

### Data collection and preprocessing

In the present study, the dataset uniformly sets the cement ratio as 1, while expressing the proportions of other components relative to cement. To use machine learning to predict the mechanical properties of ECC, a total of 1532 sets of data were collected from the existing literature<sup>3,32–68</sup>, including 501 sets of compressive strength, 224 sets of flexural strength, 404 sets of tensile strength, and 403 sets of tensile strain capacity. The experimental group containing CGP was completed by the present author. Owing to space limitations, the experiment details are not provided, but the data can be published, and the relevant experiments are to be published in other articles. There were a total of 13 input features and four output labels. The selection of input features was based on the research conclusions of the literature referenced in the database. The input features were water-cement ratio (W), silica

fume (SF), slag (SG), fly ash (FA), coal gangue powder (CGP), silica sand (SS), fiber content (FCO), water reducer (WR), fiber class (FC), fiber aspect ratio (FAR), compressive specimen size (CSS), cement grade (CG), and flexural specimen size (FSS). The output labels were compressive strength (CS: 7.4–166 MPa), flexural strength (FS: 2.6–34.6 MPa), tensile strength (TS: 0.95–18 MPa), and tensile strain capacity (TSC: 0.002% ~ 12.3%). Among them, FC was divided into surface oil-coating polyvinyl alcohol fiber (PVA-O), surface oil-free polyvinyl alcohol fiber (PVA-UO), ultra-high molecular weight polyethylene fiber (PE), fiber-free (NF) (There was a test group without fiber to analyze the effect of fiber), and polypropylene fiber (PP); CG was divided into PO.42.5 and PO.52.5; CSS was divided into 50 mm, 70 mm, and 100 mm; FSS was divided into three cases: four-point flexural plate (PLATE), three-point flexural beam (3-XL), and four-point flexural beam (4-DL). Therefore, 19 input features were included in the present study. The test conditions from which the data set was derived needed to be specified. In the present study, only the datasets under 28-day standard curing conditions were considered, utilizing variable speed mixing methods. However, the input features did not encompass properties of raw materials, such as particle size distribution, particle shape, and chemical composition, which could potentially have influenced the mechanical properties of ECC. Nevertheless, these factors were omitted due to the challenge of quantitatively describing and simplifying them within the machine learning model.

The machine learning data set was derived from the reference literature, and the mix ratios with missing values were excluded when collecting data. Before training the model, the classification features were processed by means of one-hot encoding, and the input features of the non-tree model were normalized using min–max scaling.

### Data analysis

Figure 1 shows the relationship of all continuous characteristic parameters and compressive strength. Figure 2 shows the correlation heat map between all features (including classification features). However, the main focus was primarily on the correlation between each feature and the mechanical properties. Correlations with coefficients between  $-0.4$  and  $0.4$  were disregarded as they were considered insignificant for the analysis<sup>26</sup>. According to Figs. 1 and 2, there was a negative correlation between (W, FA) and CS and a positive correlation between (CG, SF) and CS; a negative correlation between FAR and FS; a positive correlation between W and FS; a negative correlation between W and TS; and a positive correlation between (FAR, FC, CG, SF) and TS. There was no negative correlation between feature parameters and TSC, but a positive correlation between (FAR, FC) and TSC.

### Methodology

The technology route of the present study can mainly be divided into three steps: the first step is data collection and division. The training set accounted for 80%, and the test set accounted for 20%. In the second step, the tenfold cross-validation was used to optimize the hyperparameters and reduce overfitting, as shown in Fig. 3. The third step involved predicting the test set and analyzing the importance of features. Figure 4 shows the specific technology route.

## Machine learning

### Machine learning models

Numerous machine learning algorithms are available for predicting the mechanical properties of concrete. In the present study, four representative algorithms were employed to predict the mechanical properties of ECC: linear regression (LR), K nearest neighbors (KNN), random forest (RF), and extreme gradient boosting (XGB).

#### LR model

LR is the simplest and most basic supervised learning algorithm in machine learning. It has good interpretability and is widely used in engineering. The calculating function can be expressed as follows:

$$h_{\theta}(x_1, x_2, \dots, x_n) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n, h_{\theta}(x_1, x_2, \dots, x_n) \text{ is a mechanical indicator, } \theta_0, \theta_1, \dots, \theta_n \text{ are regression parameters, and } x_1, x_2, \dots, x_n \text{ are characteristic parameters.}$$

#### KNN model

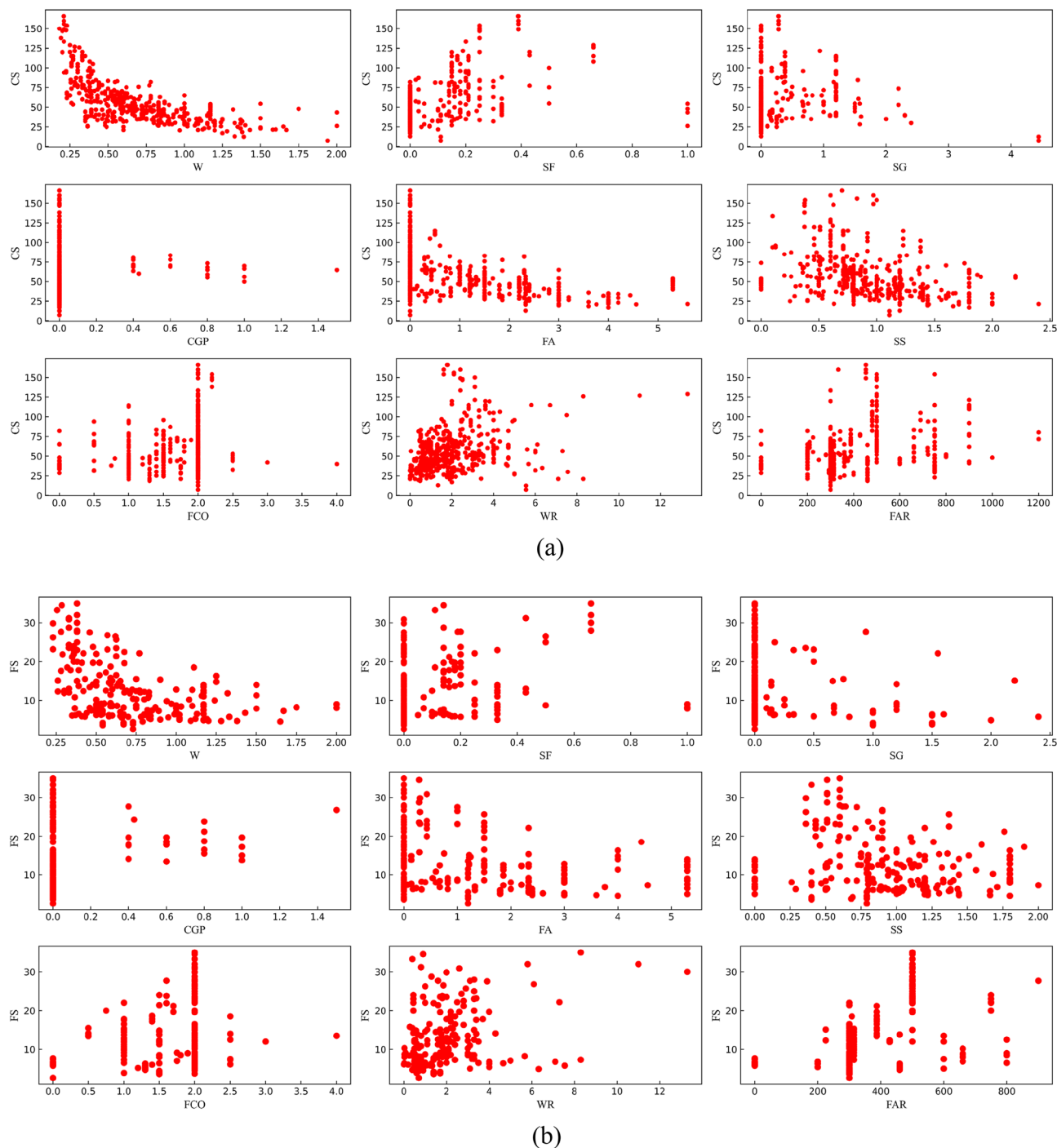
KNN is also a simple algorithm in machine learning. It calculates the K nearest sample points based on Euclidean distance and derives the final result by averaging these K sample points.

Supposing there are two points P and Q,  $P = \{p_1, p_2, \dots, p_n\}$ ,  $Q = \{q_1, q_2, \dots, q_n\}$ ,  $n = 1, 2, 3, \dots$ , the Euclidean distance between P and Q is  $d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_1)^2 + \dots + (p_n - q_n)^2}$ .

#### RF model

RF is an ensemble algorithm belonging to the tree model. A random forest comprises multiple decision trees, and there is no correlation between each decision tree in the forest<sup>69</sup>. The main advantages of random forests lie in their ability to handle high-dimensional data without the need for dimensionality reduction or feature selection. Additionally, they can provide insight into feature importance. However, one disadvantage is that random forest models are generally less powerful for regression problems compared to classification problems.

The working mechanism of the RF algorithm is as follows: firstly, M samples ( $M \leq N$ ) are randomly drawn from the training set with a size of N for each tree; if a total of K times are randomly selected, then the randomly selected training set is put into K trees for separate training; in the final prediction, the final output of the test set is determined by each decision tree in the forest. For the regression problem, the mean output values of all decision trees are used as the final output. Figure 5 shows the learning flow chart of RF.

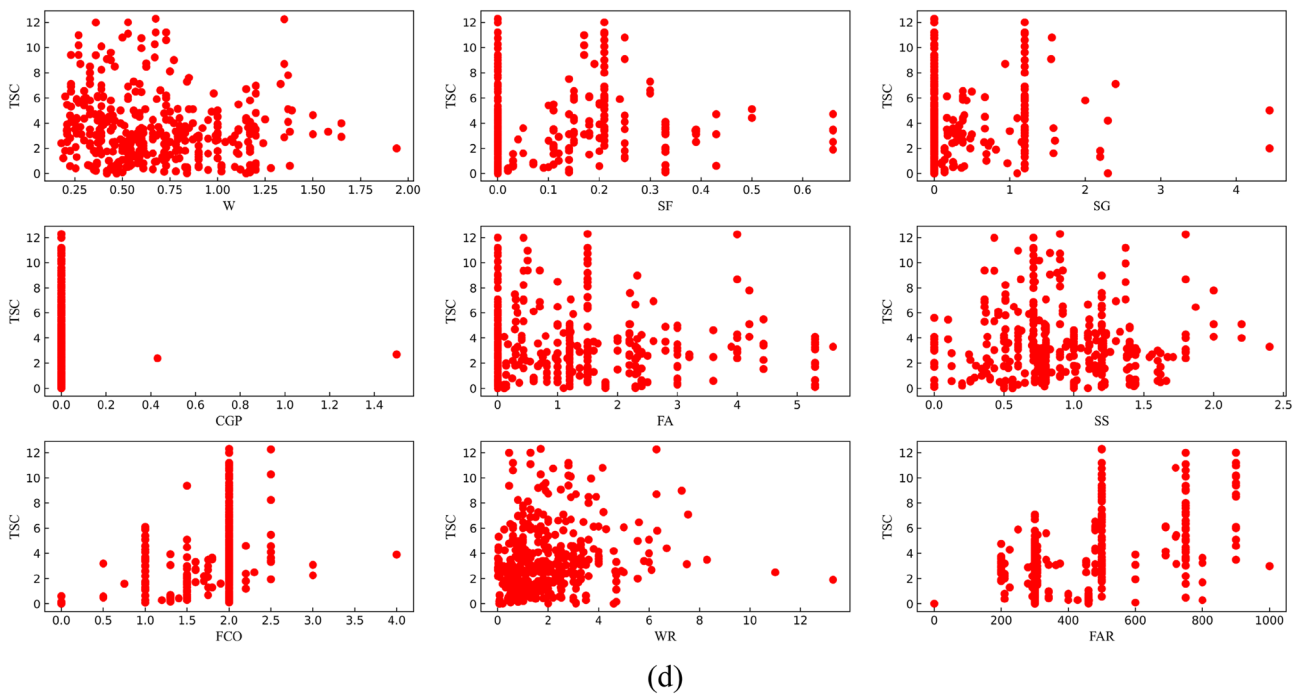
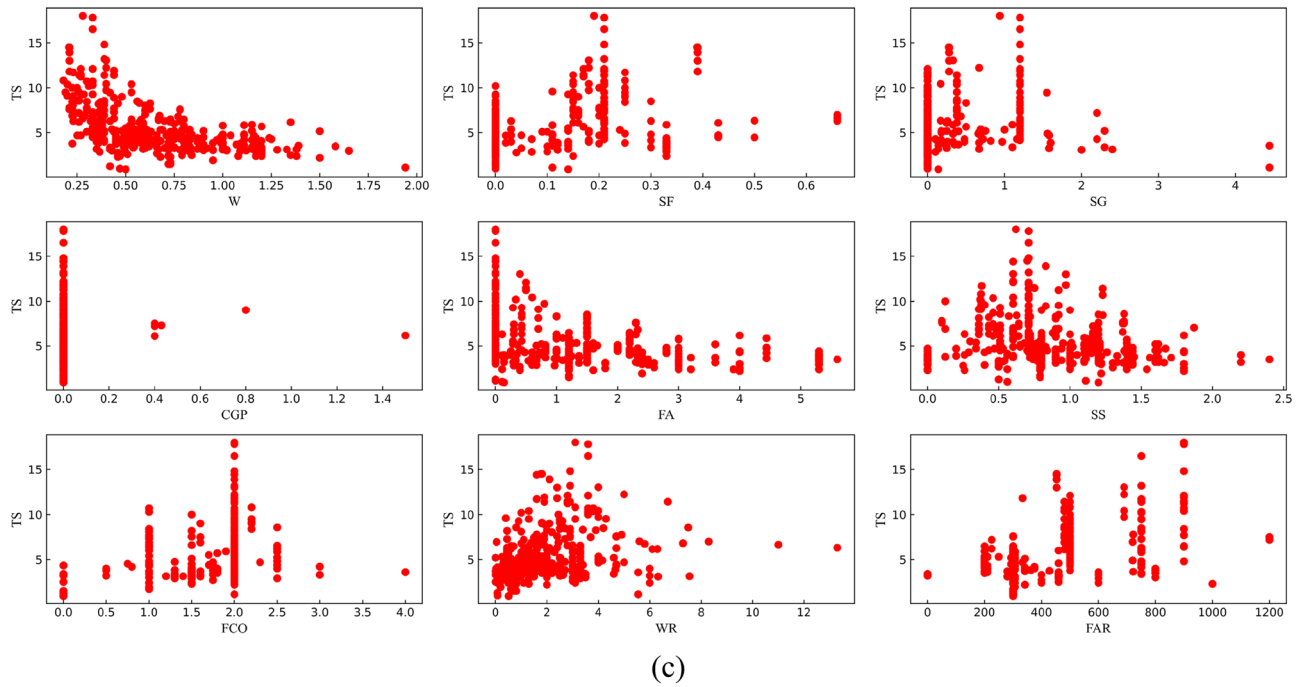


**Figure 1.** The relationship between features and mechanical properties [(a) CS, (b) FS, (c) TS, (d) TSC].

#### XGB model

An essential difference between XGB and RF lies in how their decision trees are constructed. In XGB, the decision trees are interrelated, offering certain advantages. The XGB algorithm boasts high accuracy and rapid training speed, and it exhibits robustness to missing values. Nevertheless, it comes with the drawback of high space complexity, necessitating more storage memory<sup>70</sup>.

The working mechanism of the XGB algorithm is as follows: firstly, a weak learner  $W_1$  is trained on the training set by using the initial weight, and the weight of the training sample is updated according to the residual left by the weak learner 1. The weight is increased when the residual is large, and the weight is reduced when the residual is small. After updating the weights based on the residuals left by the weak learner 1, the new adjusted weights are used to train the weak learner 2. The described process is repeated until the number of weak learners reaches the preset value. Finally, all weak learners are integrated through a set strategy to obtain the final strong learner. Figure 6 shows the learning flow chart of XGB.

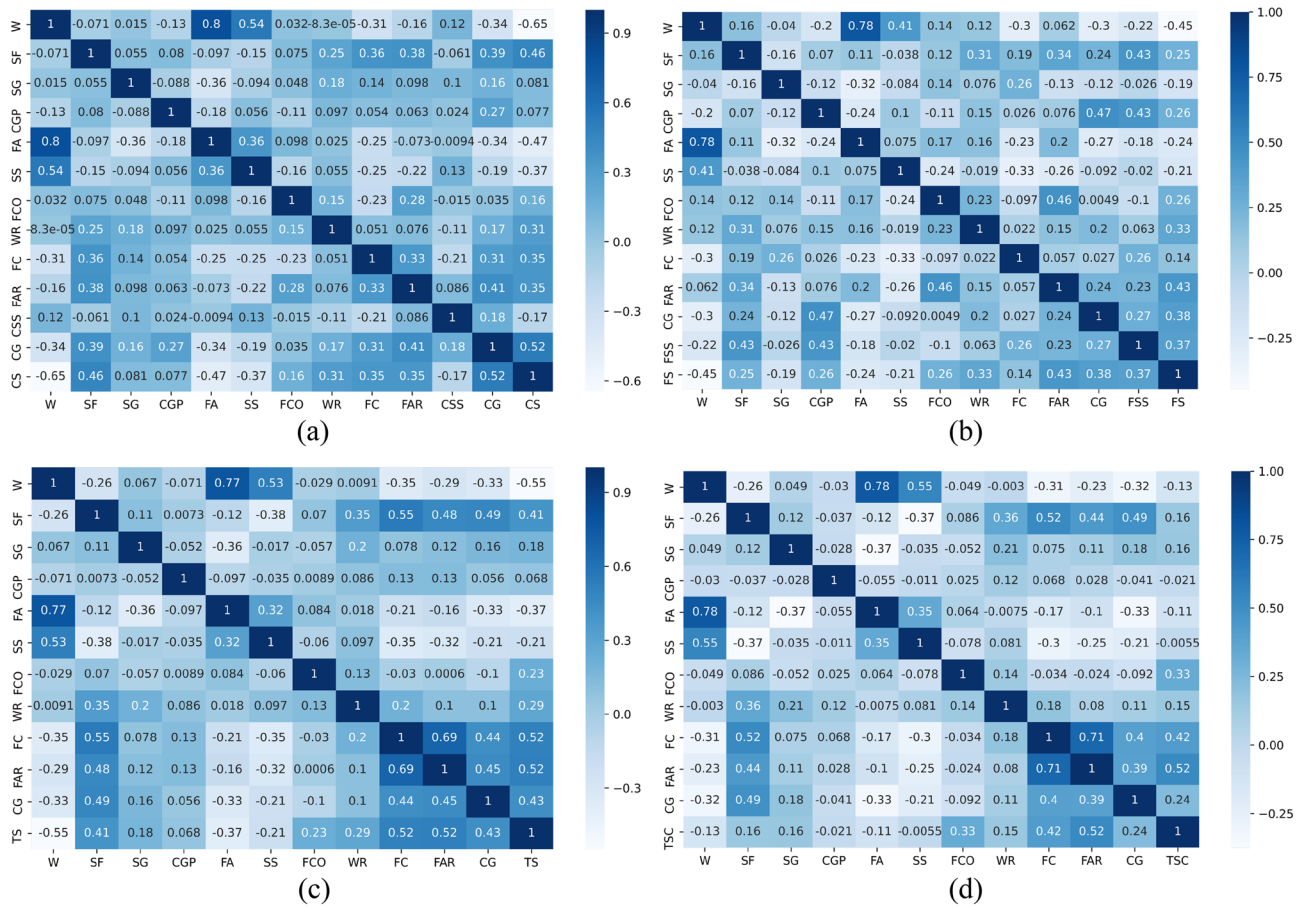


**Figure 1.** (continued)

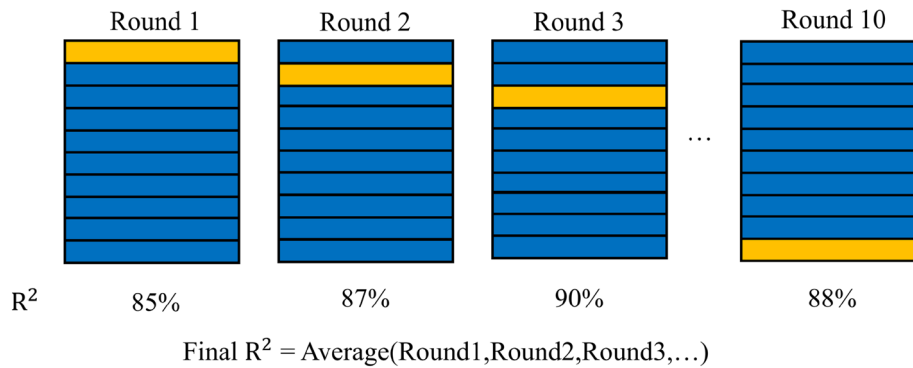
**Evaluation indicators**

In the present study, the four evaluation indications most widely used in machine learning were adopted to compare the prediction accuracy of different models: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ( $R^2$ ). The detailed expressions are shown in Eqs. (1–4), representing the predicted value, actual value, and the average of the actual value.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \tag{1}$$



**Figure 2.** Correlation heat map between all features [(a) CS, (b) FS, (c) TS, (d) TSC].

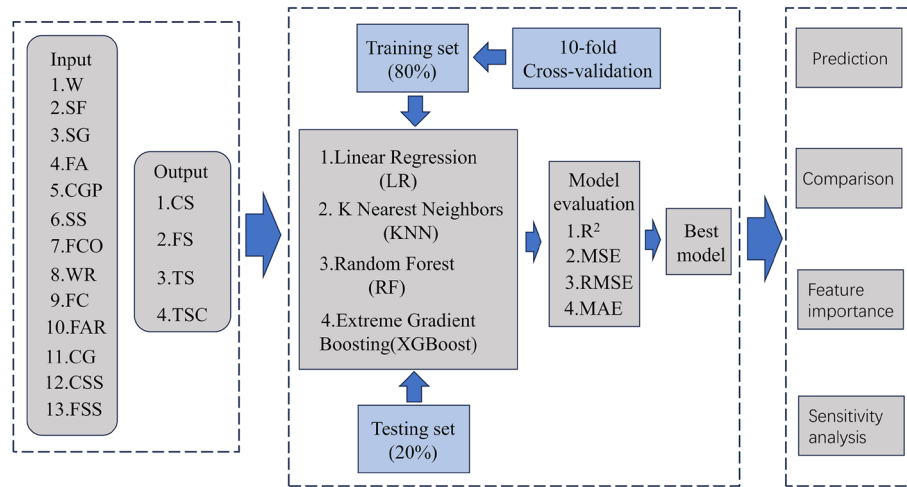


**Figure 3.** 10-fold cross-validation.

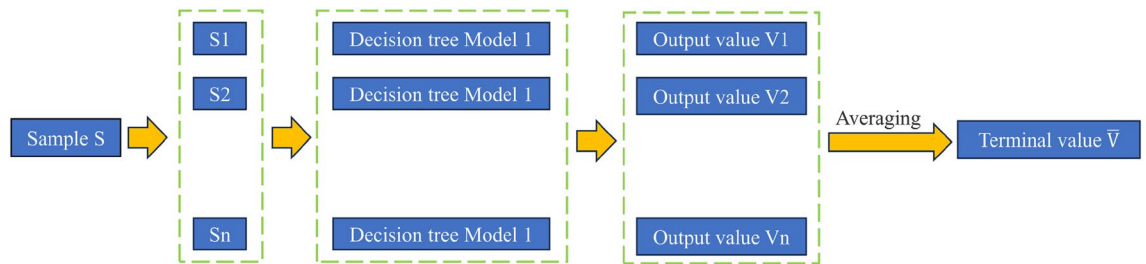
$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{2}$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \tag{3}$$

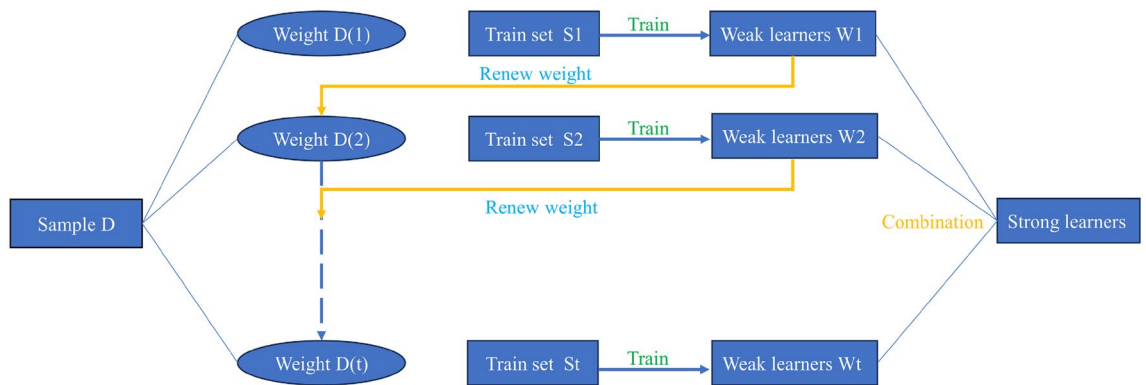
$$R^2 = 1 - \frac{\sum (\hat{y}_i - y_i)}{\sum (\hat{y}_i - \bar{y}_i)} \tag{4}$$



**Figure 4.** Machine learning technology route for the prediction of ECC mechanical properties.



**Figure 5.** Process of random forest algorithm.



**Figure 6.** Process of XGB algorithm.

## Results and discussion

### Compressive strength

Table 1 lists the optimal parameter values of the model after hyperparameter optimization, and the definition of each parameter is also shown, in which the LR model has no hyperparameter. Table 2 shows the three evaluation indicator values of the four models after hyperparameter optimization on the training and test sets. An observation can be made that the  $R^2$  values of the LR model on the train and test set were 0.72 and 0.70, respectively, which was the lowest among the four models, while the RMSE and MAE of the LR model on the training and test set were the largest. These indications suggest that the LR model is experiencing underfitting. The  $R^2$  of the KNN model was significantly higher than that of the LR model, and the RMSE and MAE values were also significantly reduced.

The RF and XGB models were superior to the LR and KNN models. The RF model had the best generalization ability, and its  $R^2$  was the highest among the four models, at 0.99 and 0.92 on the training and test sets,

Model	Parameter	Definition	Value
KNN	n_neighbors	Number of nearest neighbors	3
RF	n_estimators	Number of decision trees	19
	max_depth	Maximum depth of each decision tree	15
	max_features	Maximum number of features	8
XGB	n_estimators	Number of decision trees	64
	min_child_weight	Minimum sum of instance weight	6
	max_depth	Maximum depth of each decision tree	8
	gamma	Minimum amount of loss required for further branching	0.3

**Table 1.** The optimal parameter values of the model (CS).

Model	Train set			Test set		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
LR	0.72	14.43	10.55	0.70	15.28	11.24
KNN	0.89	8.98	6.15	0.86	10.43	7.52
RF	0.99	3.23	2.34	0.92	7.90	5.40
XGB	1.00	1.41	1.01	0.89	9.50	6.23

**Table 2.** Three evaluation indicators on the training set and the test set (CS).

respectively. The XGB model exhibited signs of overfitting, evident from the R<sup>2</sup> of 1.00 on the training set and R<sup>2</sup> of 0.89 on the test set. As such, the XGB model was deemed inappropriate under the division of the compressive strength data set in the present study.

From the R<sup>2</sup> comparison of the aforementioned models on the training set and the test set, it can be found (Table 2) that the R<sup>2</sup> values of the LR and KNN models on the two data sets were small. Both models exhibited underfitting in compressive strength prediction. The LR model reached its limit state as it lacked hyperparameters to optimize, and thus was unable to enhance its accuracy in predicting the compressive strength of ECC further. While the KNN model's prediction accuracy for compressive strength may not have matched that of XGB and RF models, it offered simplicity with only one hyperparameter, resulting in faster training speed. Despite this, the KNN model demonstrated acceptable performance on the test set. Overall, in terms of compressive strength prediction accuracy, the RF model emerged as the top performer.

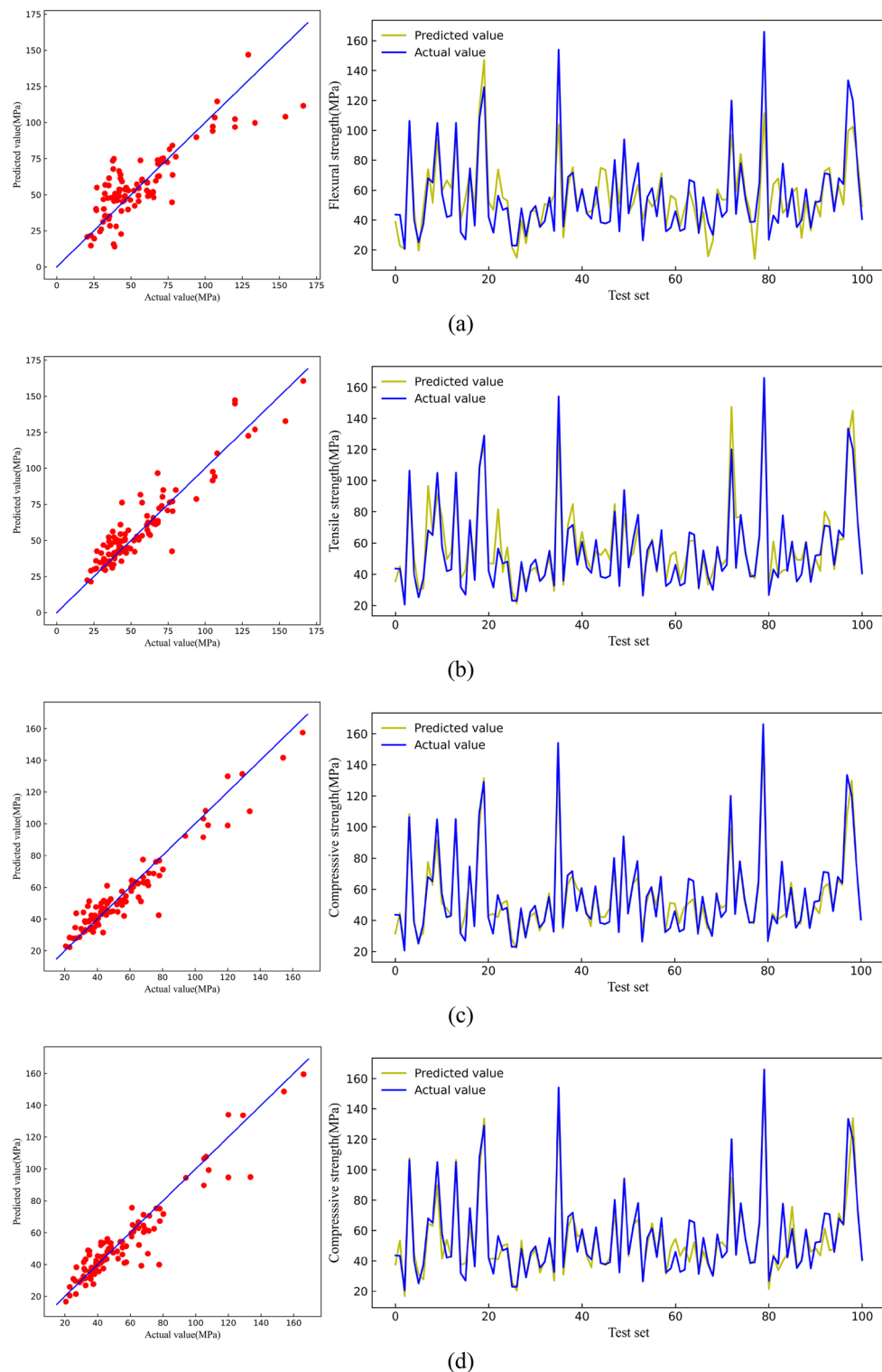
Figure 7 shows the predicted and actual compressive strength values on the test set. The predicted values of the LR model, KNN model, XGB model, and RF model gradually became more aggregated on both sides of the straight line, indicating that the RF model has the best generalization level on the compressive strength test set.

While ensemble learning methods often exhibit exceptional accuracy, they lack transparency in explaining their output, leading to what is known as the “black box” of machine learning. To address this challenge, the SHapley Additive exPlanations (SHAP) approach was introduced. SHAP is a mathematical framework developed to elucidate the underlying prediction mechanisms of machine learning models. This methodology originated from Shapley game theory and was initially proposed by Lundberg and Lee<sup>71</sup>.

Figure 8 shows the results of the feature importance analysis of the RF model (analyzing the optimal model). The SHAP value of a feature indicates its contribution to the target value. When the SHAP value is closer to 0, it suggests that the feature has a smaller impact on the target value. Conversely, when the SHAP value is farther from 0, it indicates a more significant contribution of the feature to the target value. An observation can be made from Fig. 8 that the water-cement ratio, silica fume, and water reducer contributed the most to the prediction of compressive strength. In contrast, the size of the compressive specimen, PVA fiber, and CGP contributed the least to the prediction of compressive strength. Although the feature importance of Fig. 8 can illustrate the contribution of feature parameters, it cannot indicate whether the feature had a positive or negative impact on the prediction results. However, the Global SHAP value in Fig. 9 can demonstrate how each feature positively or negatively affected the compressive strength. In Fig. 9, each point represents a sample, with red representing a high eigenvalue and blue representing a low eigenvalue. Taking 'W' as an example, the high eigenvalue (red) made a negative contribution to the model output, and the low eigenvalue (blue) made a positive contribution to the model output. A lower eigenvalue (blue) would cause a higher SHAP value, indicating that the water-cement ratio and compressive strength had a negative impact. Contrarily, the high eigenvalue (red) positively contributed to the model output, and the low eigenvalue (blue) negatively contributed to the model output. A higher eigenvalue (red) would cause a higher SHAP value, indicating that silica fume positively affected compressive strength. This pattern applies to other features analyzed similarly.

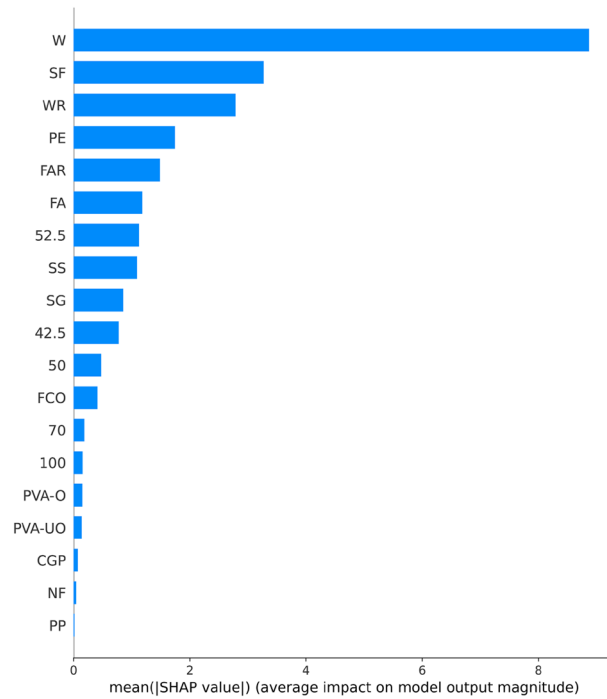
SHAP offers the capability to observe both local (individual) interpretability as well as global interpretability, a feature not achieved by traditional variable importance algorithms. In Fig. 10, the SHAP values of a randomly selected single sample from the training model are depicted. The colors, blue and red, represent negative and positive contributions, respectively. Each feature is represented by a bar, with the length indicating



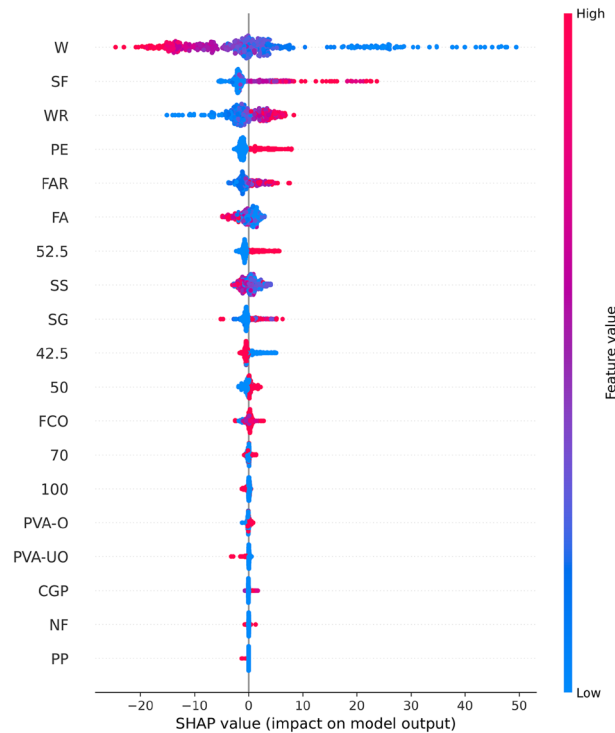


**Figure 7.** Comparison between predicted value and actual value of CS [(a) LR, (b) KNN, (c) RF, (d) XGB].

its contribution to the prediction result. Red bars signify features that increase the prediction value, while blue bars indicate features that decrease the prediction value. Notably, in this specific example, quartz sand exhibited the longest blue bar, suggesting its substantial role in decreasing compressive strength, while silica fume had the longest red bar, suggesting its substantial role in increasing compressive strength. Additionally, SHAP can illustrate partial dependence plots, which differ from traditional ones by using SHAP values on the y-axis rather than the target value. Figure 11 shows the SHAP partial dependence plots. As shown, the compressive strength was positively correlated with silica fume and water reducing agent and negatively correlated with water

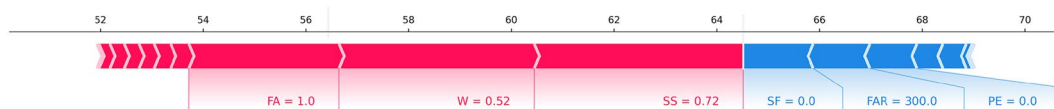


**Figure 8.** Feature importance of RF modeling.

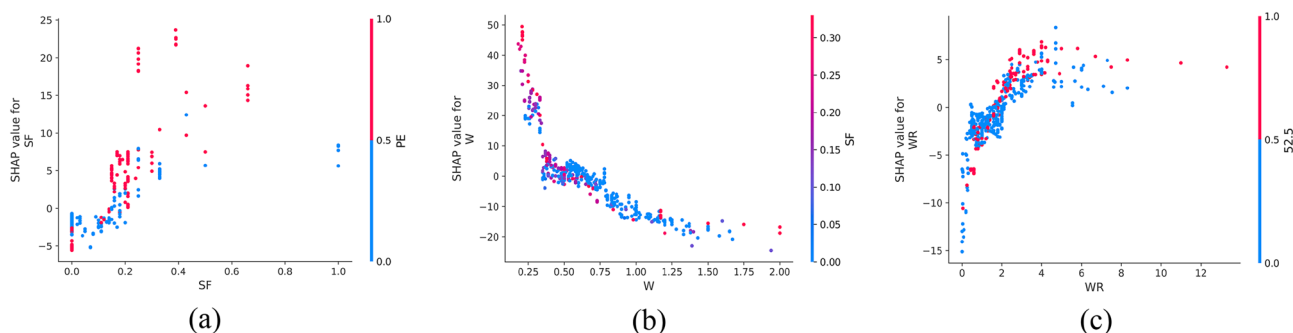


**Figure 9.** Global SHAP values of RF modeling.

cement ratio. In addition, the relationship between silica fume and compressive strength was almost linear. In contrast, the relationship between the water-cement ratio of the water-reducing agent and compressive strength was virtually logarithmic. From Fig. 11b, it can be observed from Fig. 11b that when  $W < 0.5$ , the SHAP value decreased with the increase in  $W$ , and the compressive strength was highly sensitive to the change of  $W$ . When  $W > 0.5$ , although  $W$  was still inversely correlated with compressive strength, the sensitivity was significantly



**Figure 10.** SHAP values of an individual sample.



**Figure 11.** SHAP partial dependency plots for the CS.

weaker than  $W < 0.5$ . According to Fig. 11c, it can be found that when  $WR < 0.1$ , the SHAP value increased with the increase in  $W$ , and the compressive strength was highly sensitive to the change of  $WR$ ; when  $WR > 0.1$ , the sensitivity of compressive strength to  $WR$  was significantly weakened.

### Flexural strength

Table 3 shows the optimal parameter values of the model after hyperparameter optimization. Table 4 shows the evaluation indicators of the four models on the training and test sets. An observation can be made that the  $R^2$  values of the LR model and KNN model on the training set were smaller than those on the test set, which suggests the presence of data leakage, potentially leading to overly optimistic results<sup>72</sup>. The RF and XGB models performed better, with  $R^2$  of 0.97 on the training set and  $R^2$  of 0.91 on the test set. The RMSE and MAE of the RF model on the training set were 1.16 and 0.80, and the RMSE and MAE on the test set were 2.32 and 1.62, smaller than the XGB.

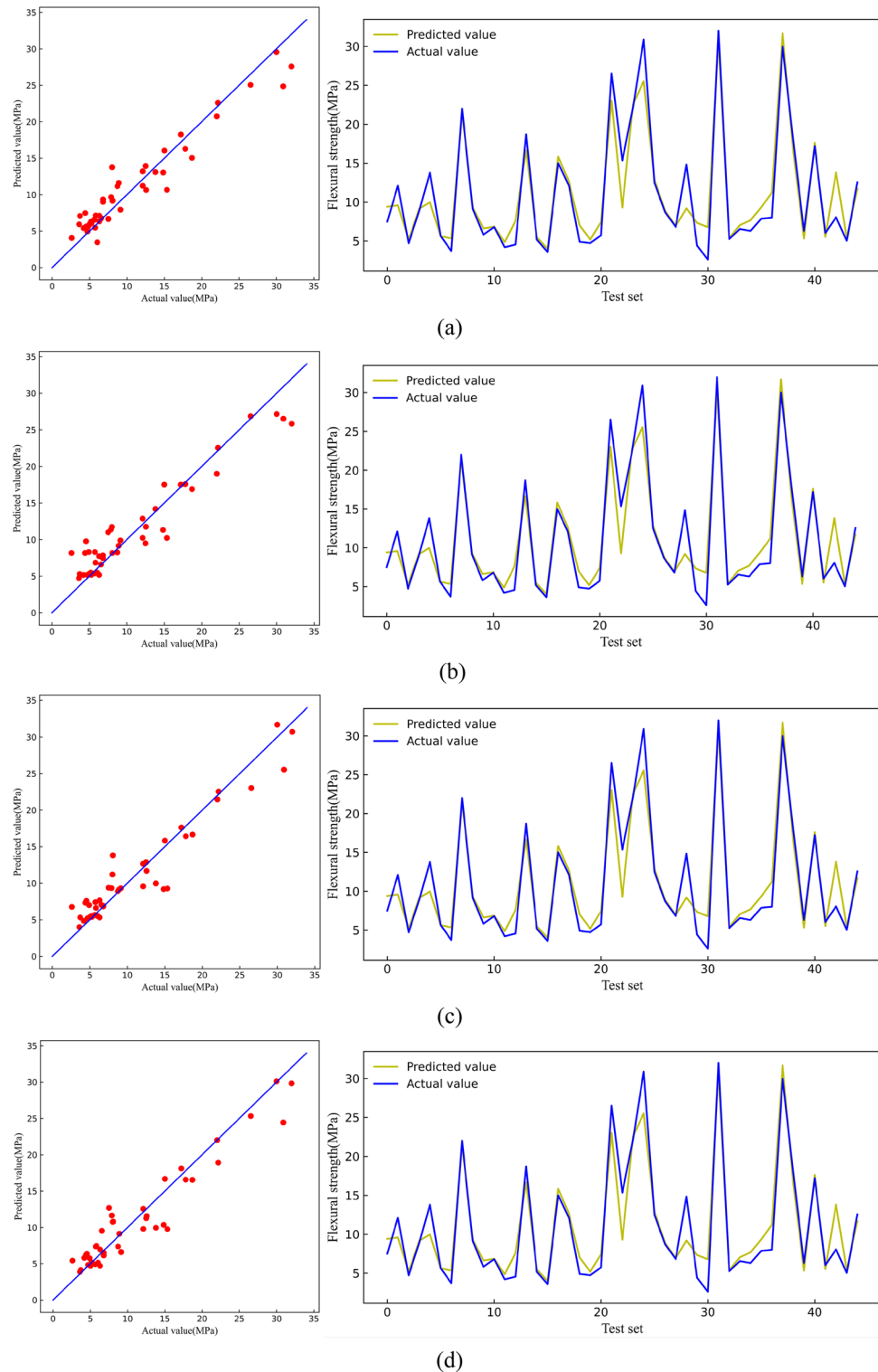
Figure 12 shows the predicted and actual flexural strength values on the test set. The four models had good prediction results on the test set, but the LR and KNN models exhibited data leakage phenomena that could not

Model	Parameter	Definition	Value
KNN	n_neighbors	Number of nearest neighbors	6
RF	n_estimators	Number of decision trees	229
	max_depth	Maximum depth of each decision tree	18
	max_features	Maximum number of features	6
XGB	n_estimators	Number of decision trees	13
	min_child_weight	Minimum sum of instance weight	4
	max_depth	Maximum depth of each decision tree	9
	reg_alpha	Penalty coefficient of L1 Regularization	0.5

**Table 3.** The optimal parameter values of the model(FS).

Model	Train set			Test set		
	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE
LR	0.82	3.12	2.27	0.92	2.21	1.73
KNN	0.85	2.83	2.00	0.90	2.48	1.82
RF	0.97	1.16	0.80	0.91	2.32	1.62
XGBoost	0.97	1.33	0.87	0.91	2.32	1.73

**Table 4.** Three evaluation indicators on the training set and the test set (FS).



**Figure 12.** Comparison between predicted value and actual value of FS [(a) LR, (b) KNN, (c) RF, (d) XGB].

be used to make accurate predictions. Thus, only the generalization ability of the RF and XGB models on the test set was deemed practical, and the RF model was the most superior in terms of predicting flexural strength.

In summary, under the premise of the present study’s flexural strength data set, the LR and KNN models exhibited data leakage, and thus were not feasible for ECC flexural strength prediction. The XGB and RF models demonstrated exceptional prediction accuracy for the flexural strength of ECC, effectively addressing a gap in the literature where previous studies did not predict the flexural strength of ECC.

Figure 13 shows the importance of analyzing RF model characteristics of flexural strength. It indicates that PE fiber, water-cement ratio, and fiber aspect ratio contributed the most to the prediction of flexural strength. In contrast, other fiber types contributed the least to the prediction of compressive strength.

Figure 14 shows that PE fiber positively affected flexural strength, while the water-cement ratio had a negative effect. Other characteristics can also be analyzed according to the Global SHAP value. Figure 15 shows the SHAP

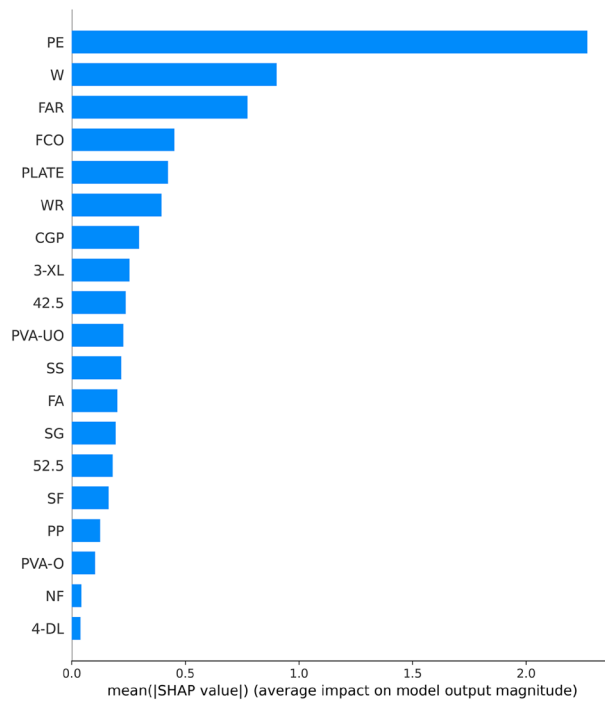


Figure 13. Feature importance of RF modeling.

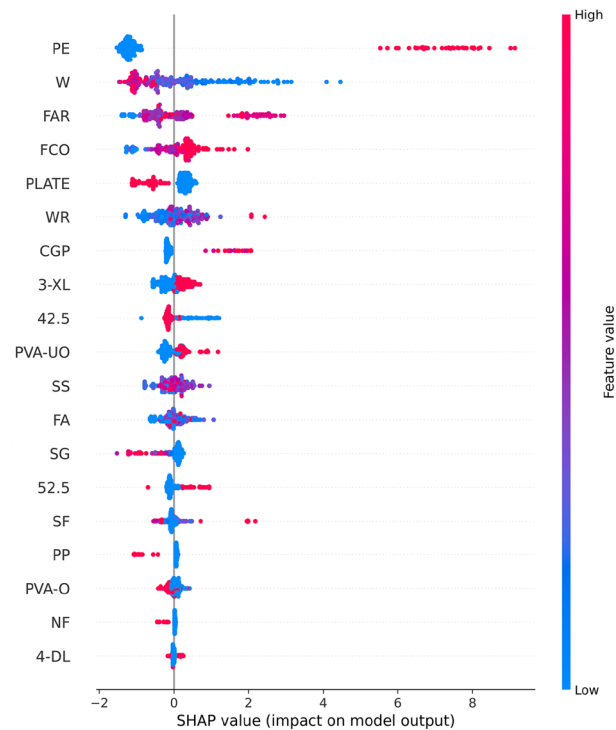
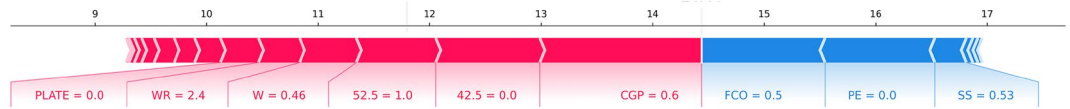


Figure 14. Global SHAP values of RF modeling.



**Figure 15.** SHAP values of an individual sample.

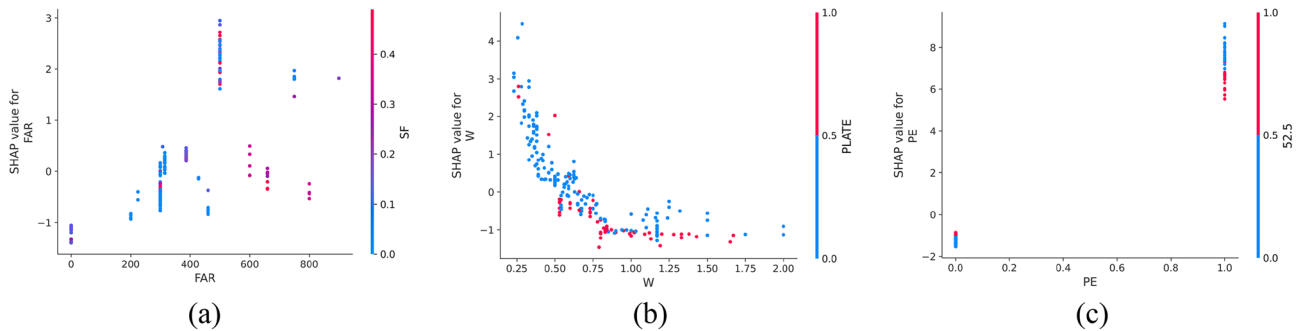
values of a random single sample of the training model. The most extensive red bar was coal gangue powder, which would increase the flexural strength for the specific sample of coal gangue powder. The most extensive blue bar was fiber content, which would reduce the flexural strength in this particular sample of fiber content.

In Fig. 16, the SHAP dependency partial plots reveal several key observations regarding the relationship between input features and flexural strength. Notably, flexural strength exhibited a positive correlation with fiber aspect ratio and the presence of PE fiber, while showing a negative correlation with the water-cement ratio. Moreover, the relationship between fiber aspect ratio and flexural strength appeared to be roughly linear, indicating that flexural strength increased with higher fiber aspect ratios. Conversely, the relationship between water-cement ratio and flexural strength appeared to be almost logarithmic. It can be found from Fig. 16b that when  $W < 0.75$ , the SHAP value decreased rapidly with the increase in  $W$ , indicating that the flexural strength was highly sensitive to the change of  $W$  in this range. When  $W > 0.75$ , although  $W$  was still inversely correlated with flexural strength, the sensitivity had almost disappeared.

### Tensile strength

Table 5 lists the optimal parameter values of the model after hyperparameter optimization. Table 6 shows the evaluation indicator values of the four models on the training and test sets. The  $R^2$  values of the LR model on the training and test sets were only 0.71 and 0.63, which shows that the LR model could not predict the tensile strength well. The  $R^2$  values of the KNN model on the training and test sets were 0.89 and 0.75, which indicates overfitting. The  $R^2$  values of the RF model on the training and test sets were 0.96 and 0.84; although RF exhibited slight overfitting, the prediction accuracy on the test set was relatively good. The  $R^2$  values of the XGB model on the training and test sets were 0.97 and 0.87.  $R^2$  was the largest among the four models, and the RMSE and MAE on the test set were only 1.04 and 0.68, the smallest among the four models.

From the collected data sets, it is evident that the variability of tensile strength in ECC surpassed that of compressive strength. In Table 6, it is apparent that both LR and KNN models exhibited underfitting. In the dataset used in the present study, the LR and KNN models failed to accurately predict the tensile strength of ECC. This could be attributed to the models' simplicity and the considerable variability in tensile strength test values. As such, machine learning struggled to fully capture the underlying relationships within the data, necessitating



**Figure 16.** SHAP partial dependency plots for the FS.

Model	Parameter	Definition	Value
KNN	n_neighbors	Number of nearest neighbors	2
RF	n_estimators	Number of decision trees	9
	max_depth	Maximum depth of each decision tree	15
XGB	n_estimators	Number of decision trees	37
	min_child_weight	Minimum sum of instance weight	2
	max_depth	Maximum depth of each decision tree	5
	gamma	Minimum amount of loss required for further branching	0.3
	reg_alpha	Penalty coefficient of L1 regularization	0.05

**Table 5.** The optimal parameter values of the model(TS).

Model	Train set			Test set		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
LR	0.71	1.46	1.06	0.63	1.73	1.25
KNN	0.89	0.89	0.60	0.75	1.41	0.91
RF	0.96	0.56	0.38	0.84	1.13	0.74
XGB	0.97	0.44	0.32	0.87	1.04	0.68

**Table 6.** Three evaluation indicators on the training set and the test set (TS).

improvements in model complexity. This is why RF and XGB models achieved more accurate predictions. Despite their increased complexity, tree models yielded superior prediction results.

Figure 17 shows the predicted and actual tensile strength values on the test set. The RF and XGB data were considerably aggregated near the line, indicating exceptional generalization ability of the two models for tensile strength. Comparatively, the XGB model was superior.

In the dataset used in the present study, the LR and KNN models failed to predict the tensile strength of ECC accurately. This could be attributed to the models' simplicity and the considerable variability in tensile strength test values. As such, machine learning struggled to fully capture the underlying relationships within the data, necessitating improvements in model complexity. This is why RF and XGB models achieved more accurate predictions. Despite their increased complexity, tree models yielded superior prediction results.

Figure 18 shows analysis of the XGB model's feature importance for tensile strength. PE fiber, water-cement ratio, and water reducer contributed the most to predicting tensile strength, while other fibers contributed minimally.

Figure 19 shows that PE fiber positively affected tensile strength, while the water-cement ratio had a negative effect. Other characteristics could also be analyzed according to the Global SHAP value. Figure 20 shows the SHAP value of a random single sample of the training model. The most extensive red bar was slag. For this specific sample, the slag would increase the tensile strength. The most extensive blue bar was PE fiber. In this particular sample, PE fiber would reduce the tensile strength.

Figure 21 shows the SHAP dependency partial plots. An observation can be made that the tensile strength was positively correlated with the water-reducing agent and whether PE fiber was present or negatively correlated with the water-cement ratio. In addition, the water-cement ratio and water-reducing agent were almost logarithmically related to the tensile strength. Figure 21a shows that when  $W < 0.5$ , the SHAP value decreased rapidly with the increase in  $W$ , indicating that the tensile strength was highly sensitive to the change of  $W$  in this range. When  $W > 0.5$ , although  $W$  was still inversely related to the tensile strength, the sensitivity became considerably weak. An observation can be made from Fig. 21b that when  $WR < 2$ , the SHAP value increased rapidly with the increase in  $WR$ , indicating that the tensile strength was highly sensitive to the change of  $WR$  in this range. When  $WR > 2$ , the sensitivity of tensile strength to  $WR$  changes in this range became significantly weaker.

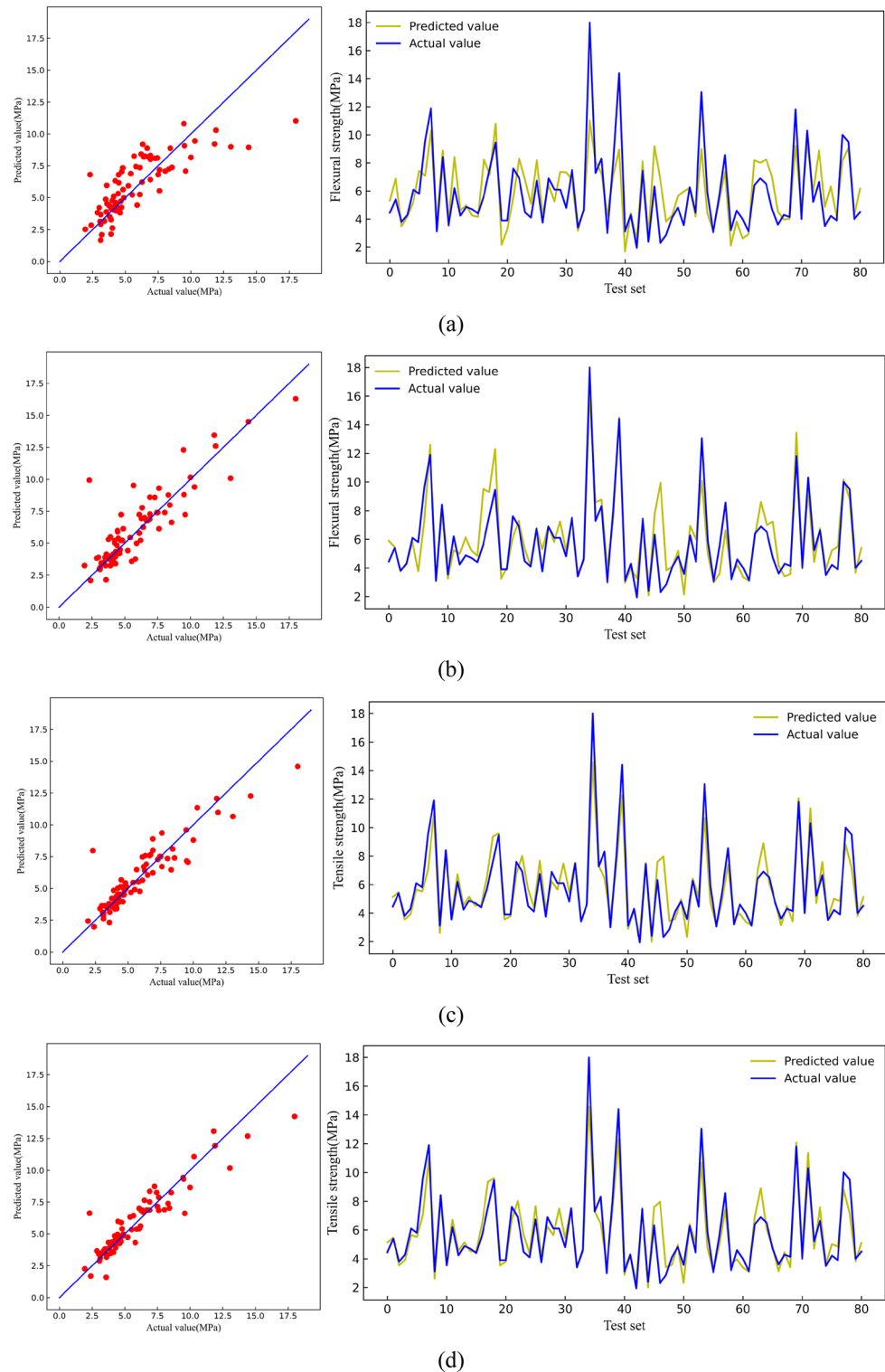
### Tensile strain capacity

Table 7 lists the optimal parameter value of the model after hyperparameter optimization. Table 8 shows the evaluation indicator values of the four models on the training and test sets. The LR model was the weakest since the  $R^2$  values were only 0.63 and 0.61 on the training and test sets. The  $R^2$  values for the KNN model were 0.85 and 0.70 on the training and test sets, respectively, indicating a slight degree of overfitting. However, the model's generalization ability on the test set remained moderate. The  $R^2$  of the RF model reached 0.92 on the test set, but was only 0.72 on the test set, indicating overfitting of the model. The XGB model was the best, although slight overfitting existed. The  $R^2$  values of the XGB model were 0.95 and 0.80 on the training and test set. As shown in previous research<sup>73</sup>, the variation range of tensile strain capacity in the test was  $[\pm 0.56\%, \pm 1.77\%]$ . However, the MAE on the test set of the XGB model in the present study was only 0.84%, which is entirely within a reasonable error range.

Figure 22 shows the predicted values and the actual values of tensile strain capacity on the test set. Among them, the XGB model prediction data was relatively more aggregated near the line and had the best effect on the prediction of tensile strain capacity. Notably, the prediction of tensile strain capacity could not be better than the other three mechanical indicators. This discrepancy can be attributed to the close relationship between tensile strain capacity and the uniform defects within the matrix of ECC. Many test groups do not introduce uniform artificial defects, leading to significant variability in tensile strain capacity test values<sup>74–76</sup>.

In Fig. 23, the analysis of the importance of features for the XGB model reveals that PE fiber, fiber content, and oil film PVA fiber made the most significant contributions to predicting tensile strain capacity, while other fibers contributed minimally. Figure 24 shows that PE fiber, fiber content, and oil film PVA fiber positively impacted tensile stress capacity. Additionally, other characteristics can be further analyzed based on Global SHAP values.

Figure 25 shows the SHAP values of a random single sample of the training model. The most extensive red bar was fly ash. For this specific sample, fly ash would increase the tensile strain capacity. The most extensive blue bar was the fiber content. In this particular sample, the fiber content would reduce the tensile strain capacity. Figure 26 shows the SHAP dependency partial plots. An observation can be made that the tensile strain capacity was positively correlated with the water-cement ratio, fiber content, and whether there was PE fiber. In addition, the three features were roughly linear with the tensile strain capacity, and the sensitivity between tensile strain capacity and input features was moderate and stable.



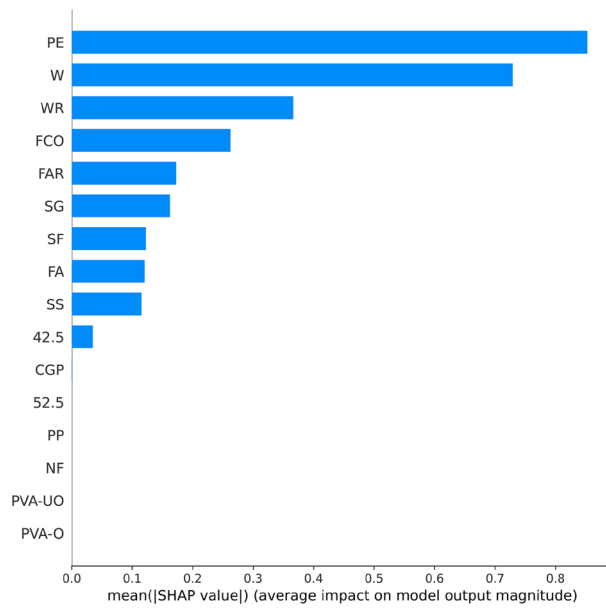
**Figure 17.** Comparison between predicted value and actual value of TS [(a) LR, (b) KNN, (c) RF, (d) XGB].

**Control strategy of ECC mechanical indicator**

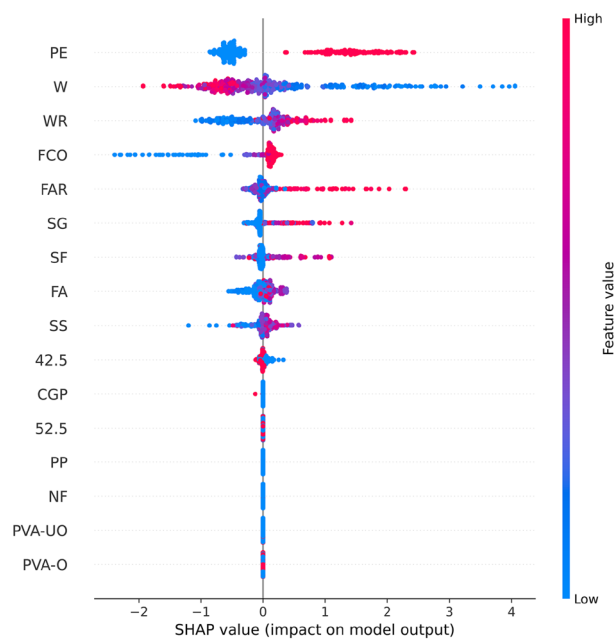
According to the feature importance and SHAP analysis, there are different strategies to improve each mechanical indicator of ECC.

Reducing the water-cement ratio, increasing silica fume content, and selecting PE fiber are the most effective strategies for compressive strength. For flexural strength and tensile strength, selecting PE fiber, reducing the water-cement ratio, and increasing the fiber aspect ratio are the most effective methods. For tensile strain capacity, increasing the fiber aspect ratio, selecting PE fiber, and increasing fiber content will be the most effective

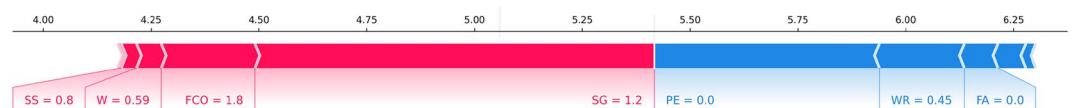




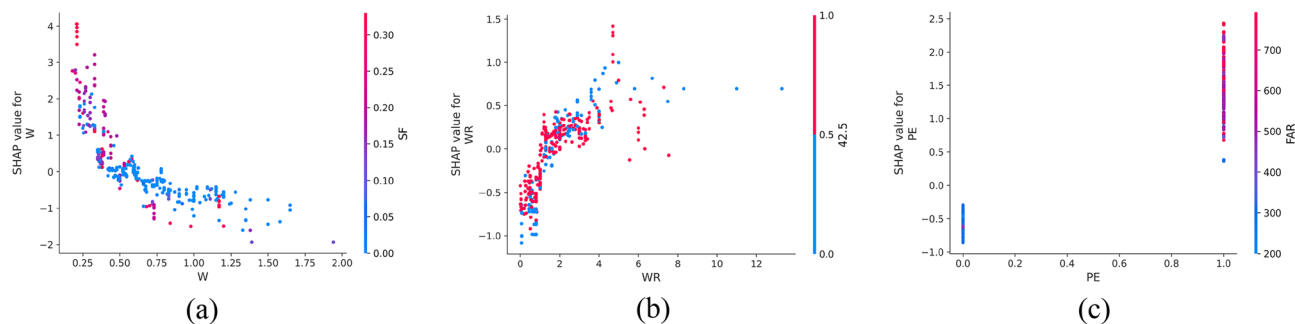
**Figure 18.** Feature importance of XGB modeling.



**Figure 19.** Global SHAP values of XGB modeling.



**Figure 20.** SHAP values of the individual sample.



**Figure 21.** SHAP partial dependency plots for the TS.

Model	Parameter	Definition	Value
KNN	n_neighbors	Number of nearest neighbors	3
RF	n_estimators	Number of decision trees	13
	max_depth	Maximum depth of each decision tree	9
	max_features	Maximum number of features	6
XGBoost	n_estimators	Number of decision trees	24
	min_child_weight	Minimum sum of instance weight	4
	gamma	Minimum amount of loss required for further branching	0.3

**Table 7.** The optimal parameter values of the model.

Model	Train set			Test set		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
LR	0.63	1.62	1.24	0.61	1.54	1.16
KNN	0.85	1.05	0.75	0.70	1.34	1.03
RF	0.92	0.76	0.56	0.72	1.28	0.94
XGBoost	0.95	0.57	0.42	0.80	1.10	0.84

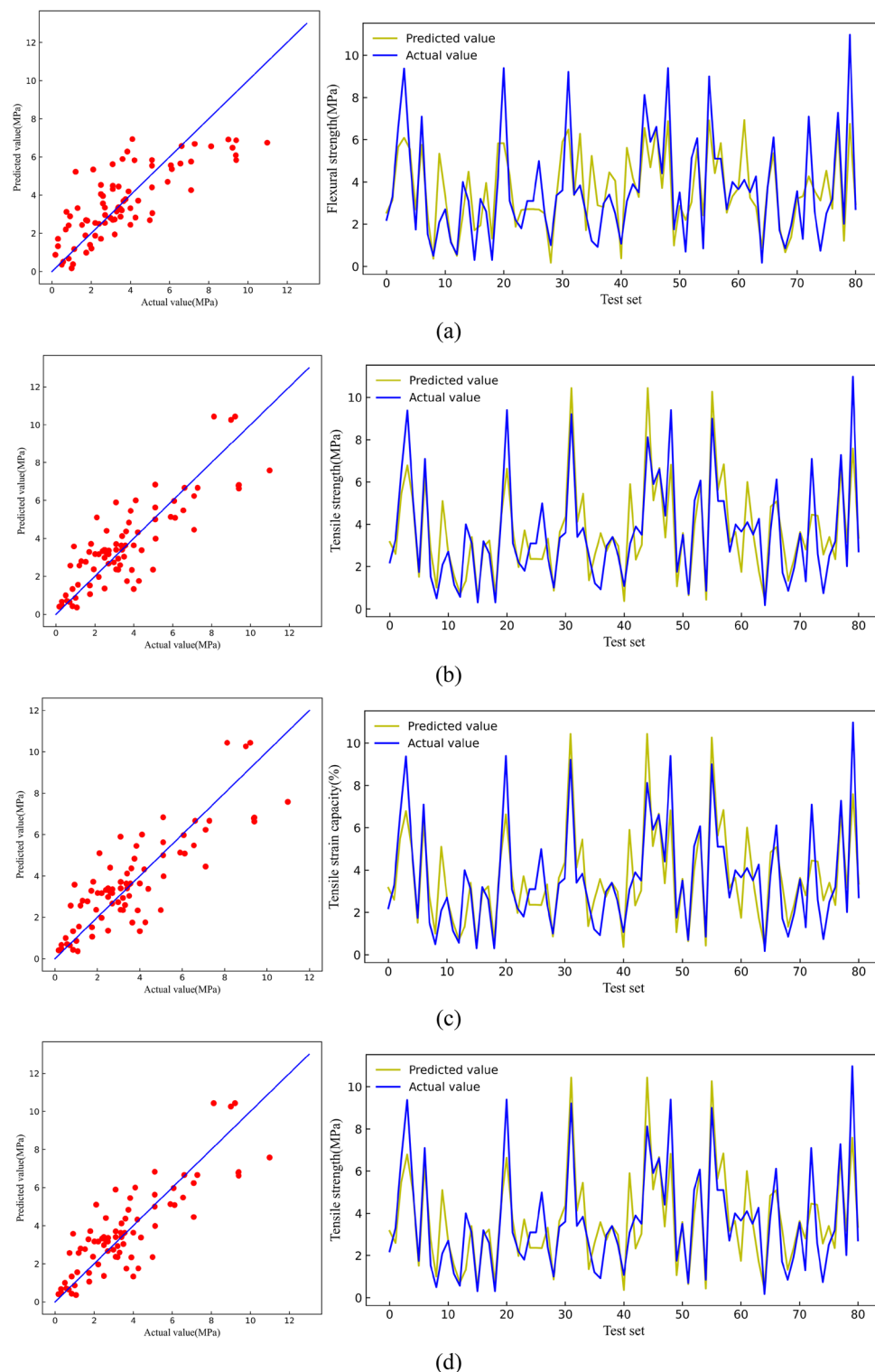
**Table 8.** Three evaluation indicators on the training set and the test set.

methods. The control methods of ECC mechanical properties are shown in Table 9. An observation can be made that there are effective methods to improve the four mechanical properties, such as the use of PE fibers and the reduction of the water-cement ratio, thereby providing effective guidance for the design of ECC.

## Conclusion

Based on the analysis using four machine-learning models, the four most critical mechanical properties of ECC were predicted, encompassing a wide range from ordinary to high compressive strength. Leveraging SHAP analysis technology, the feature importance of the optimal model was investigated and the sensitivity of feature parameters was thoroughly examined. Finally, based on the results of machine learning research, control suggestions for ECC mechanical properties were established. These findings are directly applicable to enhancing the design of ECC, facilitating its real-world application. The main conclusions are as follows:

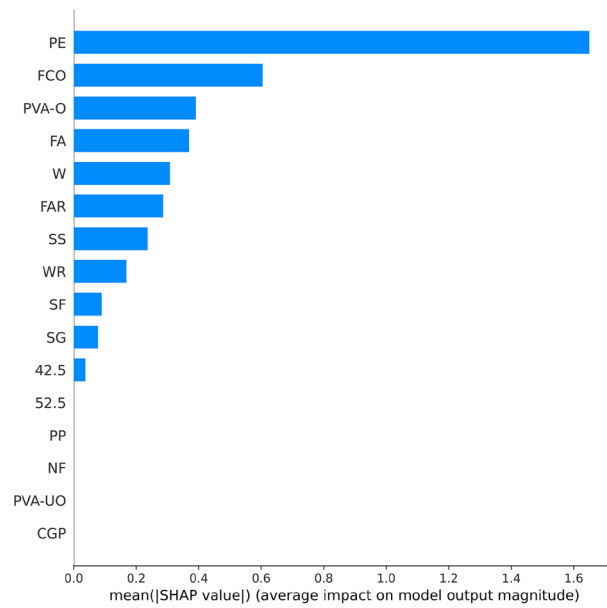
- The RF model performed best for compressive and flexural strength prediction, followed by the XGB model. For compressive strength, the R<sup>2</sup> of the RF for the training set was 0.99, and the R<sup>2</sup> and MAE for the test set were 0.92 and 5.40. For the flexural strength, the R<sup>2</sup> of the RF for the training set was 0.97, and the R<sup>2</sup> and MAE for the test set were 0.91 and 1.62. Therefore, RF is the most recommended model for predicting ECC's compressive and flexural strength.
- In predicting tensile strength and tensile strain capacity, the XGB model performed best, followed by the random forest model. For the tensile strength, the R<sup>2</sup> of the XGB model for the training set was 0.97, and the R<sup>2</sup> and MAE for the test set were 0.87 and 0.68. For the tensile strain capacity, the R<sup>2</sup> of the XGB model for the training set was 0.95, and the R<sup>2</sup> and MAE for the test set were 0.80 and 0.84. Therefore, the XGB model is the most recommended model for predicting the tensile strength and tensile strain capacity of ECC.
- Compressive strength was positively correlated with silica fume and water reducing agent and negatively correlated with water cement ratio. PE fiber positively affected flexural strength, and the water-cement ratio



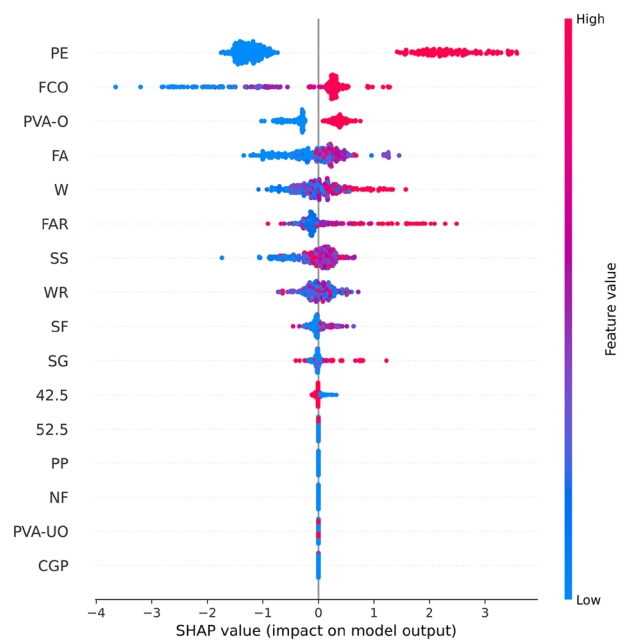
**Figure 22.** Comparison between predicted value and actual value of TSC [(a) LR, (b) KNN, (c) RF, (d) XGB].

had a negative effect on flexural strength. PE fiber positively affected tensile strength, while the water-cement ratio negatively affected tensile strength. PE fiber, fiber content, and oil film PVA fiber positively impacted tensile stress capacity.

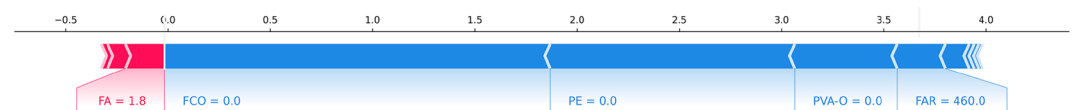
- Compressive strength and tensile strength were highly sensitive to the change of W and WR, and the flexural strength was highly sensitive to the change of W. The sensitivity between tensile strain capacity and input features was moderate and stable.



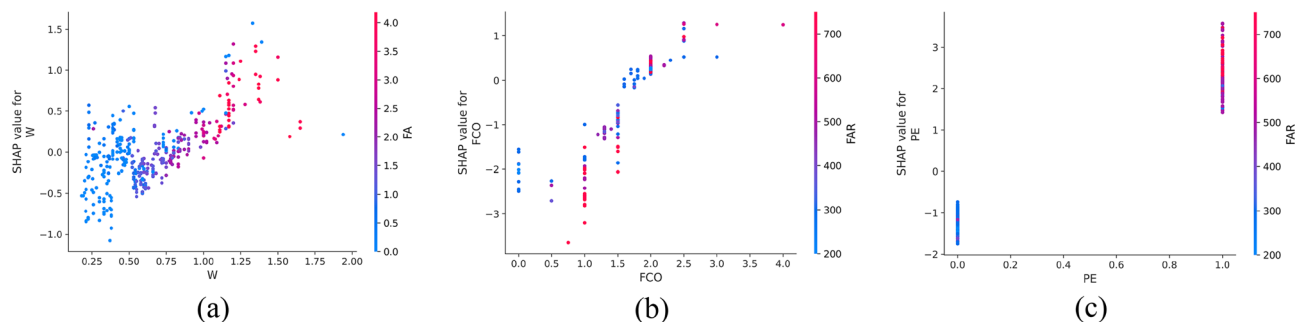
**Figure 23.** Feature importance of XGB modeling.



**Figure 24.** Global SHAP value of XGB modeling.



**Figure 25.** SHAP values of the individual sample.



**Figure 26.** SHAP partial dependency plots for the TSC.

Mechanical properties	Effective method
Compressive strength	(1) Reduce the water-cement ratio (2) Increase the content of silica fume (3) Apply PE fiber
Flexural strength	(1) Apply PE fiber (2) Reduce the water-cement ratio (3) Increase fiber aspect ratio
Tensile strength	(1) Apply PE fiber (2) Reduce the water-cement ratio (3) Increase fiber aspect ratio
Tensile strain capacity	(1) Increase fiber aspect ratio (2) Apply PE fiber (3) Increase fiber content

**Table 9.** Mechanical performance control methods of ECC.

From the perspective of the ECC mechanical index control strategy, the application of PE fiber is a very effective method to improve each mechanical index, and the second is to reduce the parameter of the cement ratio. In the design of ECC in practical engineering, these two characteristic parameters should be adjusted first to obtain the maximum effect with the least test. It avoids the repeated adjustment of each parameter in the mix ratio in the previous ECC design, which greatly reduces the test and time cost. At the same time, this paper only obtains the potential relationship between the characteristic parameters and the target values from many test data sets. In the future, we can consider the test data set and collect enough microstructure parameters of the fiber and matrix as the characteristic parameters to expect higher prediction accuracy.

### Data availability

The datasets used and/or analyzed during the current study are available from supplementary information files.

Received: 3 May 2024; Accepted: 27 June 2024

Published online: 03 July 2024

### References

- Li, V. C. & Leung, C. K. Y. Steady-state and multiple cracking of short random fiber composites. *J. Eng. Mech.* **118**(11), 2246–2264. [https://doi.org/10.1061/\(ASCE\)0733-9399\(1992\)118:11\(2246\)](https://doi.org/10.1061/(ASCE)0733-9399(1992)118:11(2246)) (1992).
- Li, V. C. From micromechanics to structural engineering—The design of cementitious composites for civil engineering applications. *Struct. Eng./Earthq. Eng.* **10**(2), 1–34. [https://doi.org/10.2208/jscej.1993.471\\_1](https://doi.org/10.2208/jscej.1993.471_1) (1994).
- Yu, J. *et al.* Tensile performance of sustainable strain-hardening cementitious composites with hybrid PVA and recycled PET fibers. *Cem. Concr. Res.* **107**, 110–123. <https://doi.org/10.1016/j.cemconres.2018.02.013> (2018).
- Ranade, R. *et al.* Micromechanics of high-strength, high-ductility concrete. *ACI Mater. J.* **110**(4), 375–384. <https://doi.org/10.14359/51685784> (2013).
- Cheung, K. F. *Development and Application of High Performance Cementitious Composites for Buildings* (Hong Kong University of Science and Technology, 2010).
- Kanda, T., Lin, Z. & Li, V. C. Application of pseudo strain hardening cementitious composites to shear resistant structural elements. *Eng. Mater. Sci.* **5**, 1477–1490 (1998).
- Suthiwarapirak, P., Matsumoto, T. & Kanda, T. Multiple cracking and fiber bridging characteristics of engineered cementitious composites under fatigue flexure. *J. Mater. Civ. Eng.* **16**(5), 433–443. [https://doi.org/10.1061/\(asce\)0899-1561\(2004\)16:5\(433\)](https://doi.org/10.1061/(asce)0899-1561(2004)16:5(433)) (2004).
- Maalej, M., Quek, S. T. & Zhang, J. Behavior of hybrid-fiber engineered cementitious composites subjected to dynamic tensile loading and projectile impact. *J. Mater. Civ. Eng.* **17**(2), 143–152. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2005\)17:2\(143\)](https://doi.org/10.1061/(ASCE)0899-1561(2005)17:2(143)) (2005).
- Niwa, J., Ishizeki, Y., Hirata, T., Matsumoto, K. & Zhang, R. Application of PP-ECC in beam-column joint connections of rigid-framed railway bridges to reduce transverse reinforcements. *Eng. Struct.* **86**, 146–156. <https://doi.org/10.1016/j.engstruct.2015.01.005> (2015).
- Lepech, M. D. & Li, V. C. Application of ECC for bridge deck link slabs. *Mater. Struct.* **42**(9), 1185–1195. <https://doi.org/10.1617/s11527-009-9544-5> (2009).

11. Deng, M.K., Gao, X.J. & Liang, X.W. Experimental investigation on ECC layer strengthening brick wall. In *International Conference on Electric Technology and Civil Engineering* (2012).
12. Deng, M. K. & Zhang, Y. X. Cyclic loading tests of RC columns strengthened with high ductile fiber reinforced concrete jacket. *Construct. Build. Mater.* **153**, 986–995. <https://doi.org/10.1016/j.conbuildmat.2017.07.175> (2017).
13. Li, V. C. *Engineered Cementitious Composites (ECC)* (Springer, 2019).
14. Kunieda, M. & Rokugo, K. Recent progress on HPFRCC in Japan required performance and applications. *J. Adv. Concr. Technol.* **4**(1), 19–33. <https://doi.org/10.3151/jact.4.19> (2006).
15. Gregor, F. Application of engineered cementitious composites (ECC) in prefabricated modular housing. *ACI Spec. Publ.* **268**, 17–28. <https://doi.org/10.14359/51663705> (2010).
16. Tetsushi, K. & Li, V. C. Practical design criteria for saturated pseudo strain hardening behavior in ECC. *ACT.* **4**(1), 59–72. <https://doi.org/10.3151/jact.4.59> (2006).
17. Tetsushi, K. & Li, V. C. Multiple cracking sequence and saturation in fiber reinforced cementitious composites. *Concr. Res. Technol. Jpn. Concr. Inst.* **9**(2), 19–33. [https://doi.org/10.3151/crt1990.9.2\\_19](https://doi.org/10.3151/crt1990.9.2_19) (1998).
18. Huang, Z. Y., Liang, T. T., Huang, B., Zhou, Y. W. & Ye, J. Q. Ultra-lightweight high ductility cement composite incorporated with low PE fiber and rubber powder. *Constr. Build. Mater.* **312**, 125430. <https://doi.org/10.1016/j.conbuildmat.2021.125430> (2021).
19. Zhou, Y. W., Xi, B., Sui, L. L., Zheng, S. Y. & Xing, F. Development of high strain-hardening lightweight engineered cementitious composites: Design and performance. *Cement Concr. Compos.* **104**, 103370. <https://doi.org/10.1016/j.cemconcomp.2019.103370> (2019).
20. Conce, A. *Bayesian Reasoning and Machine Learning* (Cambridge University Press, 2012).
21. Hu, T. Y., Zhang, H., Cheng, C., Li, H. X. & Zhou, J. T. Explainable machine learning: Compressive strength prediction of FRP-confined concrete column. *Mater. Today Commun.* **39**, 108883. <https://doi.org/10.1016/j.mtcomm.2024.108883> (2024).
22. Cheng, C., Taffese, W. Z. & Hu, T. Y. Accurate prediction of punching shear strength of steel fiber-reinforced concrete slabs: A machine learning approach with data augmentation and explainability. *Buildings.* **14**, 1223. <https://doi.org/10.3390/buildings14051223> (2024).
23. Jayasinghe, T. *et al.* Data-driven shear strength predictions of recycled aggregate concrete beams with/without shear reinforcement by applying machine learning approaches. *Construct. Build. Mater.* **387**, 131604. <https://doi.org/10.1016/j.conbuildmat.2023.131604> (2023).
24. You, X.M., Yan, G.X., Al-Masoudy, M.M., Kadimallah, M.A., Aikhalifah, T., Aitourise, F. & Ali, H.E. Application of novel hybrid machine learning approach for estimation of ultimate bond strength between ultra-high performance concrete and reinforced bar. *Adv. Eng. Softw.* **180**, 103442. <https://doi.org/10.1016/j.advengsoft.2023.103442> (2023).
25. Zhang, F. *et al.* Prediction of FRP-concrete interfacial bond strength based on machine learning. *Eng. Struct.* **274**, 115156. <https://doi.org/10.1016/j.engstruct.2022.115156> (2023).
26. Liu, Y. Z., Li, Y., Mu, J. L., Li, H. W. & Shen, J. L. Modeling and analysis of creep in concrete containing supplementary cementitious materials based on machine learning. *Constr. Build. Mater.* **392**, 131911. <https://doi.org/10.1016/j.conbuildmat.2023.131911> (2023).
27. Tran, V. Q., Mai, H. V. H., Nguyen, T. A. & Ly, H. B. Assessment of different machine learning techniques in predicting the compressive strength of self-compacting concrete. *Front. Struct. Civ. Eng.* **16**(7), 928–945. <https://doi.org/10.1007/s11709-022-0837-x> (2022).
28. Zhu, H. *et al.* Prediction of early compressive strength of ultrahigh-performance concrete using machine learning methods. *Int. J. Comput. Methods* <https://doi.org/10.1142/S0219876221410231> (2023).
29. Shanmugasundaram, N., Praveenkumar, S. & Gayathiri, D. K. Prediction on compressive strength of engineered cementitious composites using machine learning approach. *Construct. Build. Mater.* **342**, 127933. <https://doi.org/10.1016/j.conbuildmat.2022.127933> (2022).
30. Mahjoubi, S., Barhemat, R., Guo, P. W., Meng, W. N. & Bao, Y. Prediction and multi-objective optimization of mechanical, economical, and environmental properties for strain-hardening cementitious composites (SHCC) based on automated machine learning and metaheuristic algorithms. *J. Clean. Prod.* **329**, 129665. <https://doi.org/10.1016/j.jclepro.2021.129665> (2021).
31. Uddin, M. N., Li, L., Ahmed, A. & Almajhali, K. Y. M. Prediction of PVA fiber effect in engineered composite cement (ECC) by artificial neural network (ANN). *Mater. Today Proc.* **1**(65), 537–542. <https://doi.org/10.1016/j.matpr.2022.03.088> (2022).
32. Maalej, M. & Li, V. C. Flexural/tensile strength ratio in engineered cementitious composites. *J. Mater. Civ. Eng.* **6**(4), 513–528. [https://doi.org/10.1061/\(ASCE\)0899-1561\(1994\)6:4\(513\)](https://doi.org/10.1061/(ASCE)0899-1561(1994)6:4(513)) (1994).
33. Pan, Z. F., Wu, C., Liu, J. Z., Wang, W. & Liu, J. W. Study on mechanical properties of cost-effective polyvinyl alcohol engineered cementitious composites (PVA-ECC). *Construct. Build. Mater.* **78**, 397–404. <https://doi.org/10.1016/j.conbuildmat.2014.12.071> (2015).
34. Zhang, Z. G., Zhang, Q. & Li, V. C. Multiple-scale investigations on self-healing induced mechanical property recovery of ECC. *Cement Concr. Compos.* **103**, 293–302. <https://doi.org/10.1016/j.cemconcomp.2019.05.014> (2019).
35. Singh, M., Saini, B. & Chalak, H. D. Performance and composition analysis of engineered cementitious composite (ECC)—A review. *J. Build. Eng.* **26**, 100851. <https://doi.org/10.1016/j.jobbe.2019.100851> (2019).
36. Aslani, F. & Wang, L. *Development of Strain-Hardening Lightweight Engineered Cementitious Composites Using Hollow Glass Microspheres*. <https://doi.org/10.1002/suco.201900096> (Wiley, 2020).
37. Deng, H. W. Utilization of local ingredients for the production of high-early-strength engineered cementitious composites. *Adv. Mater. Sci. Eng.* <https://doi.org/10.1155/2018/8159869> (2018).
38. Ma, H., Zhang, Z. G., Ding, B. & Tu, X. Investigation on the adhesive characteristics of engineered cementitious composites (ECC) to steel bridge deck. *Construct. Build. Mater.* **191**, 679–691. <https://doi.org/10.1016/j.conbuildmat.2018.10.056> (2018).
39. Keskin, S. B. *et al.* The self-healing capability of large-scale engineered cementitious composites beams. *Compos. Part B Eng.* **101B**, 1–13. <https://doi.org/10.1016/j.compositesb.2016.06.073> (2016).
40. Li, V. C., Wang, S. X. & Wu, C. Tensile strain-hardening behavior or polyvinyl alcohol engineered cementitious composite (PVA-ECC). *ACI Mater. J.* **98**(6), 483–492. <https://doi.org/10.1089/apc.2006.20.829> (2001).
41. Wee, L.S., Lian, O.C., Zain, M.R.M. Evaluation of the design mix proportion on mechanical properties of engineered cementitious composites. *Key Eng. Mater.* <https://doi.org/10.4028/www.scientific.net/KEM.775.589> (2018).
42. Qiu, E. H. Effect of self-healing on fatigue of engineered cementitious composites (ECCs). *Cement Concr. Compos.* **94**, 145–152. <https://doi.org/10.1016/j.cemconcomp.2018.09.007> (2018).
43. Ma, H., Qian, S. Z. & Li, V. C. Tailoring engineered cementitious composite with emulsified asphalt for high damping. *Construct. Build. Mater.* **201**, 631–640. <https://doi.org/10.1016/j.conbuildmat.2018.12.222> (2019).
44. Lu, C. *Experimental and Theoretical Studies on the Cracking and Deformation Behavior of Strain Hardening Cementitious Composite (SHCC)* (Hong Kong University of Science and Technology, 2016).
45. Zhang, J., Gao, Y. & Wang, Z. Evaluation of shrinkage induced cracking performance of low shrinkage engineered cementitious composite by ring tests. *Compos. Part B Eng.* **52**(9), 21–29. <https://doi.org/10.1016/j.compositesb.2013.03.012> (2013).
46. Krishnaraja, A.R., Anandakumar, S., Jegan, M., Mukesh, T.S., Kumar, K.S. Study on impact of fiber hybridization in material properties of engineered cementitious composites. *Rev. Matér.* <https://doi.org/10.1590/S1517-707620190002.0662> (2019).

47. Ye, B.B., Zhang, Y.T., Han, J.G. & Pan, P. Effect of water to binder ratio and sand to binder ratio on shrinkage and mechanical properties of high-strength engineered cementitious composite. *Construct. Build. Mater.* **226**, 899–909 <https://doi.org/10.1016/j.conbuildmat.2019.07.303> (2019).
48. Guan, X. C. *et al.* An economical ultra-high ductile engineered cementitious composite with large amount of coarse river sand. *Construct. Build. Mater.* **201**, 461–472. <https://doi.org/10.1016/j.conbuildmat.2018.12.207> (2019).
49. Bai, L., Yu, J.P., Zhang, M. & Zhou, T.T. Experimental study on the bond behavior between H-shaped steel and engineered cementitious composites. *Construct. Build. Mater.* **196**, 214–232 <https://doi.org/10.1016/j.conbuildmat.2018.11.117> (2019).
50. Luo, J.J., Cai, Z.W., Yu, K.Q., Zhu, W.J. & Lu, Z.D. Temperature impact on the micro-structures and mechanical properties of high-strength engineered cementitious composites. *Construct. Build. Mater.* **226**, 686–698 <https://doi.org/10.1016/j.conbuildmat.2019.07.322> (2019).
51. Deng, M.K., Pan, J.J. & Sun, H.Z. Bond behavior of deformed bar embedded in engineered cementitious composites under cyclic loading. *Construct. Build. Mater.* **197**, 164–174 <https://doi.org/10.1016/j.conbuildmat.2018.11.200> (2019).
52. Jun, P. & Mechtcherine, V. Behaviour of strain-hardening cement-based composites (SHCC) under monotonic and cyclic tensile loading: Part 1—Experimental investigations. *Cement Concr. Compos.* **32**(10), 801–809. <https://doi.org/10.1016/j.cemconcomp.2010.07.019> (2010).
53. Said, S. H., Razak, H. A. & Othman, I. Flexural behavior of engineered cementitious composite (ECC) slabs with polyvinyl alcohol fibers. *Construct. Build. Mater.* **75**, 176–188. <https://doi.org/10.1016/j.conbuildmat.2014.10.036> (2015).
54. Ranade, R., Li, V. C., Stults, M. D., Heard, W. F. & Rushing, T. S. Composite properties of high-strength, high-ductility concrete. *ACI Mater. J.* **110**(4), 413–422. <https://doi.org/10.1016/j.optmat.2013.06.025> (2013).
55. Kamal, A., Kunieda, M. & Ueda, N. Evaluation of crack opening performance of a repair material with strain hardening behavior. *Cement Concr. Compos.* **30**, 863–871. <https://doi.org/10.1016/j.cemconcomp.2008.08.003> (2008).
56. Lei, D., Guo, L. P. & Chen, B. The connection between microscopic and macroscopic properties of ultra-high strength and ultra-high ductility cementitious composite (UHS-UHDCC). *Compos. Part B* **164**, 144–157. <https://doi.org/10.1016/j.compositesb.2018.11.062> (2019).
57. Chen, Y. X., Yu, J. & Leung, C. K. Y. Use of high strength strain-hardening cementitious composites for flexural repair of concrete structures with significant steel corrosion. *Construct. Build. Mater.* **167**, 325–337. <https://doi.org/10.1016/j.conbuildmat.2018.02.009> (2018).
58. Curosu, I., Liebscher, M., Mechtcherine, V., Bellmann, C. & Michel, S. Tensile behavior of high-strength strain-hardening cement-based composites (HS-SHCC) made with high-performance polyethylene, aramid and PBO fibers. *Cement Concr. Res.* **98**, 71–81. <https://doi.org/10.1016/j.cemconres.2017.04.004> (2017).
59. Ranade, R., Li, V. C. & Heard, W. Impact resistance of high strength-high ductility concrete. *Cement Concr. Res.* **98**, 24–35. <https://doi.org/10.1016/j.cemconres.2017.03.013> (2017).
60. Yu, K.Q., Yu, J.T., Dai, J.G., Lu, Z.D. & Shah, S.P. Development of ultra-high performance engineered cementitious composites using polyethylene (PE) fibers. *Construct. Build. Mater.* **158**, 217–227 <https://doi.org/10.1016/j.conbuildmat.2017.10.040> (2018).
61. Yu, K. Q., Dai, J. G., Lu, Z. D. & Leung, C. K. Y. Mechanical properties of engineered cementitious composites subjected to elevated temperatures. *J. Mater. Civ. Eng.* **27**(10), 1–11. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001241](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001241) (2015).
62. Li, L. Z., Cai, Z. W., Yu, K. Q., Zhang, Y. X. & Ding, Y. Performance-based design of all-grade strain hardening cementitious composites with compressive strengths from 40 MPa to 120 MPa. *Cement Concr. Compos.* **97**, 202–217. <https://doi.org/10.1016/j.cemconcomp.2019.01.001> (2019).
63. Shanmugasundaram, N. & Praveenkumar, S. Influence of supplementary cementitious materials, curing conditions and mixing ratios on fresh and mechanical properties of engineered cementitious composites—A review. *Construct. Build. Mater.* **309**, 125038 <https://doi.org/10.1016/j.conbuildmat.2021.125038> (2021).
64. Zhang, D., Ellis, B.R., Jaworska, B., Hu, W.H. & Li, V.C. Carbonation curing for precast engineered cementitious composites. *Construct. Build. Mater.* **313**, 125502 <https://doi.org/10.1016/j.conbuildmat.2021.125502>. (2021).
65. Tahmouresi, B., Nemati, P., Asadi, M.A., Saradar, A. & Moein, M. Mechanical strength and microstructure of engineered cementitious composites: A new configuration for direct tensile strength, experimental and numerical analysis. *Construct. Build. Mater.* **269**, 121361 <https://doi.org/10.1016/j.conbuildmat.2020.121361> (2021).
66. Razavi, S. M., Nazarpour, H. & Beygi, M. H. Evaluation of mechanical and microstructure properties of engineered cementitious composites made of zeolite. *Struct. Concr.* **22**(5), 2736–2747. <https://doi.org/10.1002/suco.202000594> (2021).
67. Wang, Q., Lai, M.H. & Zhang, J. Greener-engineered cementitious composite (ECC)—The use of pozzolanic fillers and uncoiled PVA fibers. *Construct. Build. Mater.* **247**, 118211 <https://doi.org/10.1016/j.conbuildmat.2020.118211> (2020).
68. Bang, J.W., Ganesh, P.G. & Yong, K.Y. Development of ecofriendly engineered cementitious composites using supplementary cementitious materials as a binder and bottom ash aggregate as fine aggregate. *Int. J. Polymer Sci.* **7**, 1–12 <https://doi.org/10.1155/2015/681051> (2015).
69. Breiman. Random forests. *Mach. Learn.* **45**(1), 5–32 <https://doi.org/10.1023/A:1010933404324> (2015).
70. Chen, T.Q. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* <https://doi.org/10.1145/2939672.2939785> (ACM, 2016).
71. Lundberg, S.M. & Lee, S.-I. (eds.) A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017).
72. Yang, M. Q., Zhu, J. J. & Mcgaughey, A. Predicting extraction selectivity of acetic acid in pervaporation by machine learning models with data leakage management. *Environ. Sci. Technol.* **57**(14), 5934–5946. <https://doi.org/10.1021/acs.est.2c06382> (2023).
73. Zhang, Z.G., Yang, F., Liu, J.C. & Wang, S.P. Eco-friendly high strength, high ductility engineered cementitious composites (ECC) with substitution of fly ash by rice husk ash. *Cement Concr. Res.* **137**, 106200 <https://doi.org/10.1016/j.cemconres.2020.106200> (2020).
74. Wang, S.X. & Li, V.C. *Tailoring of Pre-existing Flaws in ECC Matrix for Saturated Strain Hardening, Fracture Mechanics of Concrete Structures* (2011).
75. Lu, C., Li, V. C. & Leung, C. K. Y. Flaw characterization and correlation with cracking strength in engineered cementitious composites (ECC). *Cement Concr. Res.* **107**, 64–74. <https://doi.org/10.1016/j.cemconres.2018.02.024> (2018).
76. Li, V. C. & Wang, S. X. Microstructure variability and macroscopic composite properties of high performance fiber reinforced cementitious composites. *Probab. Eng. Mech.* **21**(3), 201–206. <https://doi.org/10.1016/j.probenmech.2005.10.008> (2006).

## Author contributions

Shuangquan Qing wrote the main manuscript text. Shuangquan Qing and Chuanxi Li reviewed the manuscript.

## Funding

The authors would like to acknowledge the financial support by Guangxi Science and Technology Major Program (AA23073017), Hunan Provincial Postgraduate Research Innovation Project (CX20210764).

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-66123-9>.

**Correspondence** and requests for materials should be addressed to S.Q.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2024