Scilight

Durability and Sustainability
https://www.sciltp.com/journals/ds

Article

Comparative and Interpretable Machine Learning for Fly Ash Geopolymer Strength: Prediction, Validation and Optimization

Xiaoshuang Shi [1,2,*], Haoxiang Hu [2], Zhenhua Duan [3] and Yanlin Li [4]

[1] MOE Key Laboratory of Deep Earth Science and Engineering, College of Architecture and Environment, Sichuan University, Chengdu 610065, China
[2] Failure Mechanics & Engineering Disaster Prevention and Mitigation, Key Laboratory of Sichuan Province, College of Architecture and Environment, Sichuan University, Chengdu 610065, China
[3] Department of Structural Engineering, Tongji University, Shanghai 200070, China
[4] Chengdu Railway Construction Management Co., Ltd., Chengdu 610041, China
* Correspondence: shixs@scu.edu.cn

How To Cite: Shi, X.; Hu, H.; Duan, Z.; et al. Comparative and Interpretable Machine Learning for Fly Ash Geopolymer Strength: Prediction, Validation and Optimization. *Durability and Sustainability* **2026**, *1*(1), 1.

Received: 27 April 2026
Revised: 9 June 2026
Accepted: 2 July 2026
Published: 29 July 2026

Abstract: This study develops and experimentally validates an interpretable machine learning framework integrating Extreme Gradient Boosting (XGBoost) with Shapley Additive Explanations (SHAP) to predict the 28-day compressive strength (CS) of fly ash-based geopolymer concrete (FA-GPC). A database of 355 experimental samples with 14 input parameters was compiled from literature. XGBoost outperformed support vector regression and decision tree models, achieving a test R^2 of 0.930 with robustness confirmed by 10-fold cross-validation. SHAP analysis identified fly ash content, curing temperature, Al_2O_3 content in fly ash, and NaOH concentration as the most significant positive contributors, while water content showed a strong negative influence. Optimal thresholds for key mixture ratios, including Na_2SiO_3/NaOH (~2.5) and alkaline activator-to-fly ash ratio, were revealed. Twelve new FA-GPC mixtures were cast and tested, yielding excellent agreement between predictions and experiments (mean absolute error = 2.41 MPa). A graphical user interface was developed to facilitate practical adoption without programming expertise.

Keywords: Fly ash geopolymer concrete (FA-GPC); compressive strength; machine learning; XGBoost; SHAP analysis; mix design optimization

1. Introduction

Cement, as one of the most critical building materials globally, results in a significant amount of carbon dioxide emissions during its production process. Indeed, cement production accounts for approximately 5–7% of global CO_2 emissions [1]. It is estimated that for every ton of cement produced, roughly 900 kg of CO_2 and 3 kg of NO_x are emitted, while consuming two tons of raw materials (primarily limestone and shale) [2]. While China is the largest producer of cement [3], and according to the World Resources Institute (WRI), China is also the world's largest contributor to annual CO_2 emissions from OPC production, exceeding 40% of the world total for the cement industry. In response to the growing concerns about environmental protection and climate change, it has become an urgent need to develop new binders that replace Ordinary Portland cement (OPC) and OPC concrete to reduce greenhouse gas emissions and natural resource use.

There is a growing global interest in geopolymers [4], and new inorganic gelling materials of geopolymers are considered as environmentally friendly materials for the construction industry [5]. As a potential alternative to OPC, geopolymer is an aluminosilicate zeolite material produced by the polymerization of silica-alumina raw materials and alkali activators [6]. These silicon-aluminum raw materials are classified into silica fume (SF), red mud (RM), fly ash (FA), and Ground Granulated Blast furnace Slag (GGBS) based on their different sources and physicochemical properties. These materials are all by-products or wastes obtained from the existing industrial production processes. Applying them in the preparation of building materials can effectively reduce CO_2 emissions and promote waste utilization. Fly ash (FA) is a byproduct of coal-fired power plants. The annual production of FA worldwide is about 800 million tons [7], and the annual production in China is more than 500 million tons per year [8]. The long-term accumulation of large amounts of fly ash not only causes problems related to land occupation but also precipitates toxic elements, which have adverse effects on human health [9]. Fly ash-based geopolymer concrete (FA-GPC) is an effective and promising alternative to OPC concrete, which can reduce CO_2 emissions by up to 25–45% [10] and has better early mechanical properties and durability [11,12]. Due to these characteristics, FA-GPC has become a viable option for the construction materials industry and is expected to achieve large-scale applications in the construction industry.

Compressive strength is an important mechanical property index for selecting construction materials, which is the core index for controlling construction quality and is also used as an important basis for structural design. However, the previous manual way of calculating the concrete mix ratio consumes lots of human labor, material resources and time to obtain the target compressive strength, which is unfavorable for projects with severe construction periods and environmental changes, and the accuracy is also closely related to the operation of the test personnel and weather conditions. In addition, most of the existing standards are only applicable to OPC concrete, and the theoretical mechanism of FA-GPC, as a relatively new material, is still unclear. Compared with OPC concrete, the factors affecting the compressive strength of GPC concrete are more complex, and it is challenging to determine the optimal mix ratio [13]. Therefore, how to design the mix ratio of concrete quickly and ensure accuracy is a popular research direction in the construction industry. Although different compressive strength prediction models for FA-GPC have been proposed in the literature [14–16], most of them generate fixed formulas based on minimal inputs. One of the most significant drawbacks of such use of constitutive and empirical models is that the expression of the function is solidified, hindering the development of prediction accuracy [17].

With the rapid development of technologies such as artificial intelligence in recent years, the construction industry has also begun to use Big Data (BD) and Machine Learning (ML) to predict material properties [18–20]. Cao et al. [21] used Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Extreme Gradient Boosting (XGBoost) techniques to examine the experimental and prediction results of the compressive strength of GPC, and the XGBoost model was regarded with high accuracy. Ghosh [22] used Linear Regression (LR), Decision Tree (DT), Random Forest (RF) and SVM algorithm to analyze the 28-day compressive strength data of GPC, and found that the relationship between the dependent and independent variable sets was predicted by a nonlinear model with the highest accuracy. Amin [23] used individual ML and ensemble ML algorithms to predict the compressive strength of GPC. The individual ML methods are DT and SVM, and the ensemble ML methods are boosting, RF and bagging. The results show that the ensemble ML is robust compared with the individual ML algorithm. Furthermore, it was found that the ensemble ML models can estimate the compressive strength of GPC with higher accuracy than individual ML models [24]. Peng [17] compared three algorithms, Back-propagation Neural Network (BPNN), SVM and Extreme Learning Machine (ELM), where the SVM model has the strongest generalization ability and the most stable prediction accuracy. Eftekhar Afzali [25] evaluated five models (ANN, SVM, DT, RF, GBM) in kaolin-based polymer concrete and found that GBM performed the best in predicting compressive strength. Among the different variables studied, SiO_2 content in FA has the greatest effect on compressive strength, followed by Al_2O_3 content and alkaline activator content or concentration. In summary, the application of the ML technique can reduce the workload of laboratory work and achieve high accuracy in predicting the compressive strength of GPC.

However, traditional ML models lack interpretability due to their complexity, provide little insight into the actual relationship between input and output features, and are considered black boxes [26], in the current papers on machine learning for predicting intensity, interpretability is an important component [27–29]. Shapley Additive Explanations (SHAP) is an after-the-fact approach that illuminates how ML models generate predictions by emphasizing the importance of features and the interdependencies between variables [30]. SHAP has been successfully applied in various fields of materials science, such as explaining component-performance relationships [31], the behavior of nanophotonic structures [32], the synthesis of advanced inorganic materials [33], biomedical engineering [34], etc. Meanwhile, Aparna Kamarthi et al. [35] focused on an automated machine learning framework for concrete compressive strength. Kamarthi's proposed interpretable AutoML pipeline integrates SHAP additive analysis, providing a more systematic and interpretable framework for predicting the compressive strength of sustainable concrete. Furthermore, Dai [27] combined a stacking ensemble model with SHAP analysis and NSGA-II optimization to predict the dynamic compressive strength of GPC (R^2 = 0.9756) and optimize mix designs for minimized cost and CO_2 emissions, with experimental validation. Researchers [36] also utilized interpretive techniques such as SHAP analysis to examine the key influencing factors in the cement-based environment, providing a reference for the interpretive analysis of fly ash geopolymers-based concrete.

Although machine learning has demonstrated considerable potential for predicting the compressive strength of geopolymer concrete, most existing studies focus primarily on improving predictive accuracy, while limited attention has been paid to model interpretability and the quantitative understanding of how mixture parameters influence strength development. Furthermore, experimental validation is often insufficient, which may reduce confidence in the practical applicability and generalization of the developed models. To address these limitations, this study integrates machine learning, SHAP-based interpretability analysis, and experimental verification within a unified framework. The proposed approach not only enables accurate prediction of the 28-day compressive strength of FA-GPC but also provides transparent insights into the relative importance and influence patterns of key variables. The findings contribute to a deeper understanding of FA-GPC behavior and offer useful guidance for data-driven mix design and practical engineering applications. The remainder of this paper presents the methodology, model evaluation, interpretability analysis, experimental validation, and concluding remarks.

2. Methods

2.1. Database Description and Analysis

The CS laboratory data of GPC were collected from 40 published pieces of literature, with a total of 355 experimental data (see Supplementary Materials). These data considered 15 parameters, 14 of which were input parameters, and the last parameter compressive strength (CS) was the output parameter [10]. These input parameters are the content of fly ash (FA, $kg \cdot m^{-3}$), the content of SiO_2 and Al_2O_3 in FA (%), coarse aggregate and fine aggregate ($kg \cdot m^{-3}$), the content of sodium hydroxide (NaOH, $kg \cdot m^{-3}$) and the concentration of sodium hydroxide (NaOH (M)), the content of sodium silicate (Na_2SiO_3, kg/m^3), the ratio of sodium hydroxide to sodium silicate (Na_2SiO_3/NaOH), the ratio of alkaline activator to fly ash (AA/FA), water (Water, $kg \cdot m^{-3}$) and polycarboxylic acid superplasticizer (SP, $kg \cdot m^{-3}$), curing temperature (Temperature, °C) and curing time (Duration, h). The units and statistical characteristics of these parameters are summed up in Table 1. Based on the statistical characteristics and Supplementary Materials, it can be observed that this dataset largely covers most of the valid parameter value ranges. For low variance features, such as Na_2SiO_3/NaOH, AA/FA, NaOH (M), the base values of these features are relatively small, and there are studied recommended range intervals in the fly ash-based polymer concrete system. This also results in their more concentrated distribution around the average value in the collected dataset, with a small variance. The research on the corresponding recommended range intervals further highlights the importance of these features.

Table 1. Statistical characteristics of input and output parameters.

Parameter	Unit	Mean	Std.	Min.	Max.
FA	$kg \cdot m^{-3}$	417.6	67.31	250	550
SiO_2	%	48.15	16.62	17.57	71.5
Al_2O_3	%	23.07	8.26	4.7	36.37
Coarse aggregate	$kg \cdot m^{-3}$	1102.13	227.71	279	1567
Fine aggregate	$kg \cdot m^{-3}$	618.39	121.39	320	1100
NaOH	$kg \cdot m^{-3}$	63.84	23.42	30.5	118
NaOH (M)	M	12.18	2.95	8	20
Na_2SiO_3	$kg \cdot m^{-3}$	134.96	49.55	75.3	293
Na_2SiO_3/NaOH	-	2.23	0.6	0.95	4.5
AA/FA	-	0.47	0.11	0.25	0.86
Water	$kg \cdot m^{-3}$	127.67	43.8	45.89	223.71
SP	$kg \cdot m^{-3}$	3.96	6.73	0	30
Temperature	(°C)	42.97	35.47	0	100
Duration	(h)	35.29	104.21	0	672
CS	MPa	32.53	12.68	7.6	72.1

The Pearson correlation coefficient is calculated by Equation (1), which is used to reflect the linear correlation between two normal continuous variables [37]. The correlation of all parameters in this data set is shown in Figure 1. As can be seen from the figure, the variables with large linear Pearson correlation coefficient with compressive strength are curing time, FA and Al_2O_3 content in FA, but the linear correlation is still weak. This shows that the relationship between the compressive strength and these input parameters is not a simple multivariate linear relationship, but a complex nonlinear mapping, so it is reasonable to use the machine learning model to predict the compressive strength of concrete in this paper.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\sum_{i=1}^{n}(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{1}$$

where, $\bar{x}$, $\bar{y}$ denote the mean values of input (x_i) and output (y_i), respectively. Equation (1) ranges from −1 to 1 and is invariant under linear transformations of the two variables.

Scatterplot is the use of a series of scatters to show the numerical distribution of variables in a rectangular coordinate system. The scatter plots of the effects of different ratio design variables on the compressive strength of FA-GPC are shown in Figure 2, and the results also show that there is no significant positive or negative correlation between any of the variables and the compressive strength.

Data normalization is essential to prevent features with larger numerical ranges from dominating the model. Because the raw variables exhibit marked scale differences, training directly on unscaled data can bias distance-based or gradient-based algorithms. Two common normalization methods are Min-Max scaling and Z-score standardization. Min-Max scaling maps values to a fixed interval, often [0, 1], but is sensitive to outliers [38]. Z-score standardization transforms each feature to have zero mean and unit variance, preserving the original distribution shape while being more robust to outliers [39]. Accordingly, this study adopts Z-score standardization, computing the mean and standard deviation exclusively from the training set and applying them consistently to both training and test sets. The calculation of the Z-score is via the formula:

$$Z = \frac{X - \mu}{\sigma} \tag{2}$$

where X is the data point, μ is the mean of the dataset, and σ is the standard deviation of the dataset.

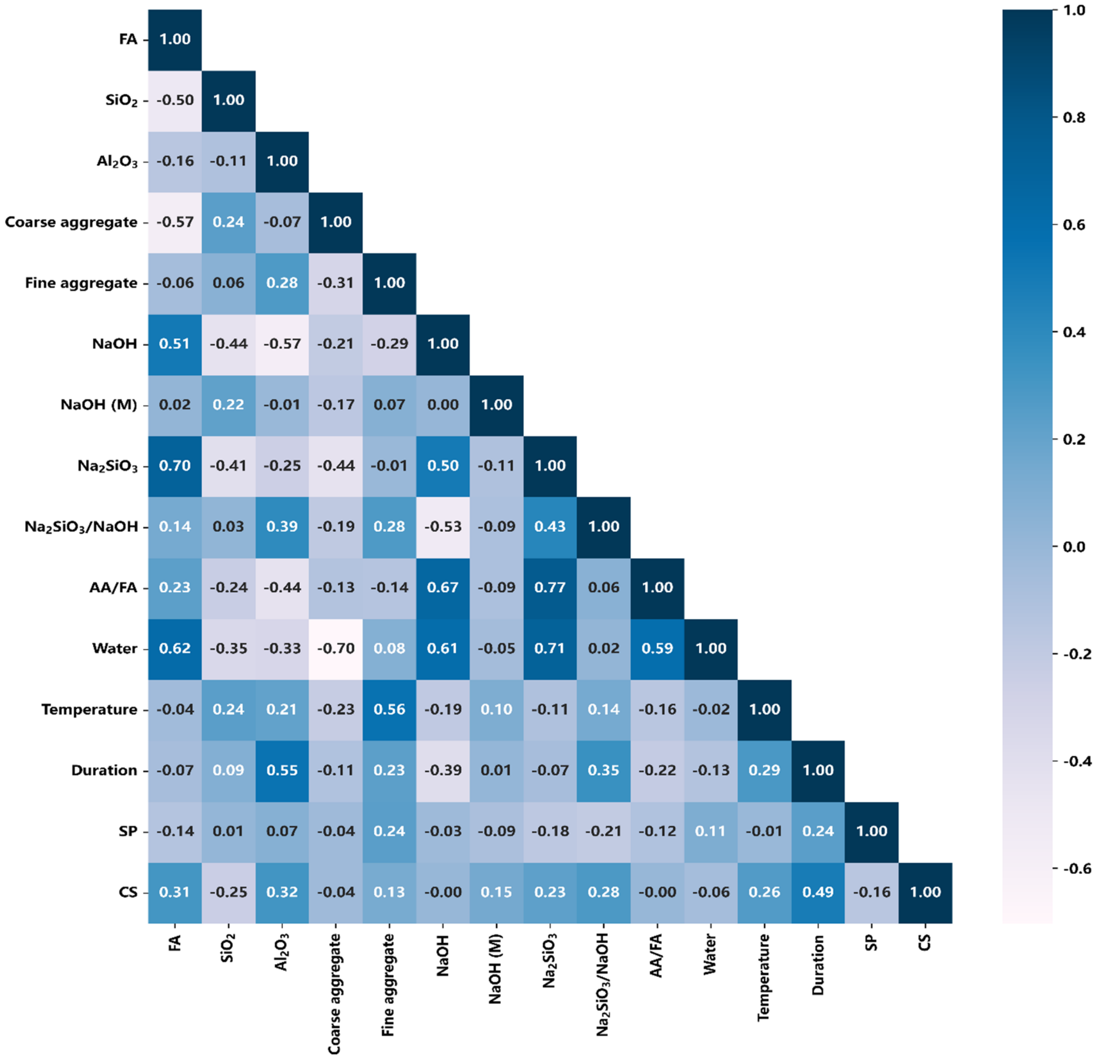


Figure 1. Pearson correlation coefficient between pairs of variables.

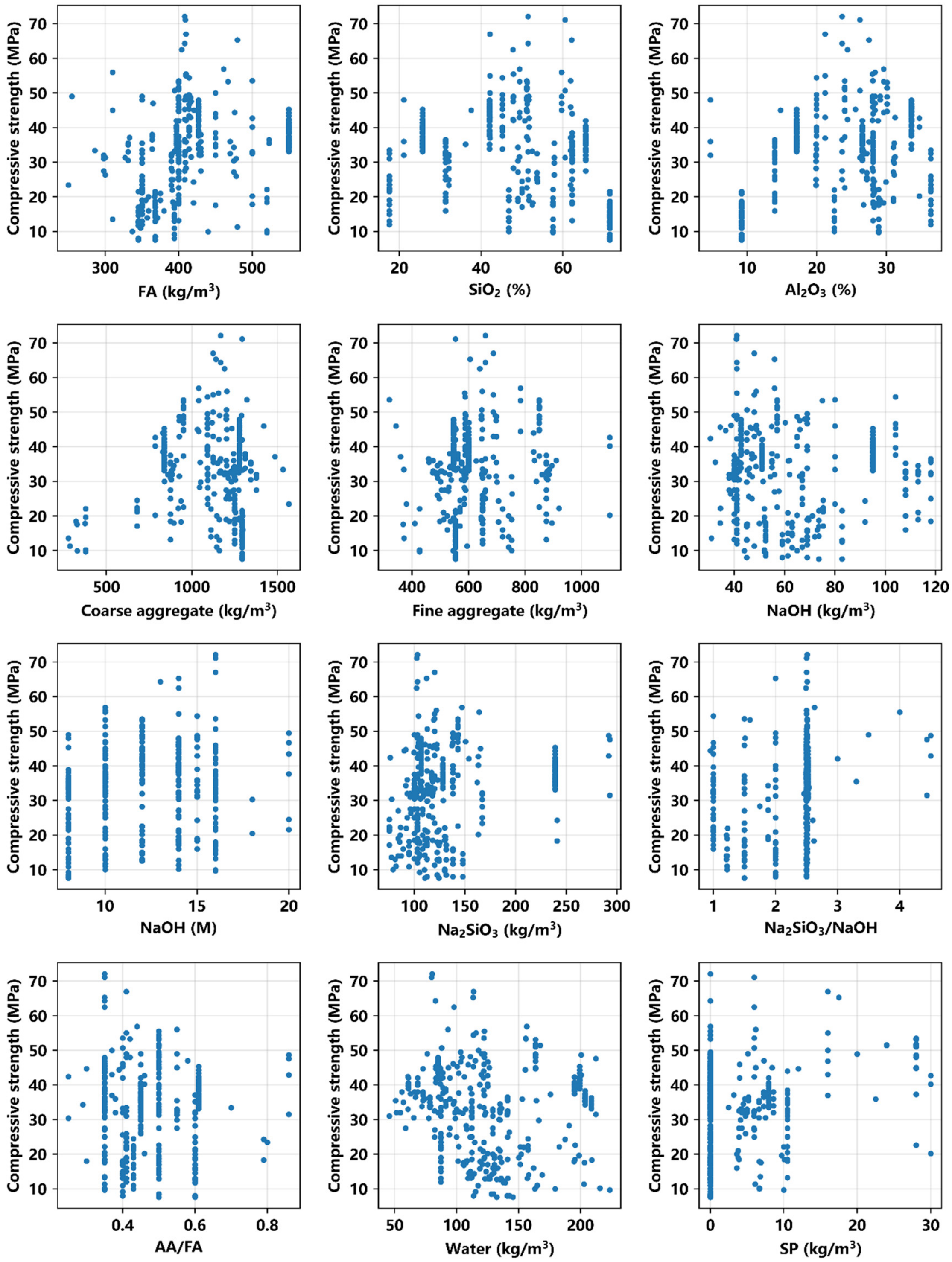


Figure 2. Correlation between mix design variables and compressive strength.

2.2. Machine Learning Algorithms

2.2.1. Support Vector Regression

Support vector machines (SVM) are well-known supervised machine learning techniques first developed by Cortes and Vapnik [40,41] that can be used for classification. The SVM model has the advantages of reasonable generalization performance, computational efficiency, high dimensional robustness, wide application range and simple use [42,43]. Support Vector Regression (SVR) is an important extension of SVM, which is applied to regression analysis [44]. Each data item is plotted as a point in n-dimensional space (n represents the number of variables), and then an appropriate hyperplane can be calculated via SVR to fit the data set. Specifically, given the training dataset $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where x_i is the input variable, and y_i is the response variable, the function $f(x)=\omega\cdot x+b$ is used as a solution to the problem in the linear case [45], the problem can be transformed into the following form as shown in Equation (3).

$$\begin{cases} Minimize\dfrac{1}{2}\|\omega\|^2 \\ y_i-(\omega\cdot x_i)-b\leq\varepsilon \\ (\omega\cdot x_i)+b-y_i\leq\varepsilon \end{cases} \tag{3}$$

where ω is the weight vector, and b is the bias.

To solve the unviable constraints of the optimization problem, two slack variables ξ_i and ξ^*_i and penalty parameter c ($c > 0$) were introduced.

$$\begin{cases} Minimize\dfrac{1}{2}\|\omega\|^2+c\sum_{i=1}^{n}(\xi_i+\xi_i^*) \\ y_i-(\omega\cdot x_i)-b\leq\varepsilon+\xi_i \\ (\omega\cdot x_i)+b-y_i\leq\varepsilon+\xi_i^* \\ \xi_i,\xi_i^*\geq 0 \end{cases} \tag{4}$$

For the nonlinear case, a kernel function $K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$ can be introduced to map the sample to a higher dimensional feature space, so that it can be linearly differentiable in this feature space. Linear, polynomial, and radial basis function (RBF) are the most popular kernel functions [46,47]. In this paper, the RBF kernel function is used, as shown in Equation (5).

$$K(x_i,x_j)=exp\left(-g\|x_i-x_j\|^2\right) \tag{5}$$

2.2.2. Decision Tree

DT is still one of the most widely used models in supervised learning because of its advantages of easy visualization, simultaneous application to classification and regression tasks, and large-scale data sets. A decision tree (DT) consists of a root node, several internal nodes and leaf nodes, where the leaf node corresponds to the decision result and other nodes correspond to the attribute judgment rules. DT is essentially a recursive judgment based on conditions layer by layer. DT can be applied to both classification tasks and regression tasks. Whether it is used for classification tasks or regression tasks mainly depends on the continuity of target variables, CART (Classification and Regression Tree) generates a classification decision tree if the result to be predicted is discrete data, otherwise it generates a regression decision tree [48,49].

2.2.3. Extreme Gradient Boosting

Gradient boosting machines (GBM) algorithm was originally proposed by Friedman [50]. Gradient Tree Boosting (GTB) is an ensemble tree model using the GBM algorithm. It is a boosting machine learning model, which uses the sequence of "weak" or "base" learners to create an arbitrary and accurate "strong" learner to minimize errors.

The Extreme gradient boosting (XGBoost), which is a highly scalable extension of GBM [51], is an extreme gradient boosted and tree-based machine learning model. It is worth noting that XGBoost is efficient and versatile, has been widely recognized for its machine learning and data mining challenges [51], and can be used in data augmentation and explainable ensemble learning [52]. Hao et al. applied XGBoost to predict the compressive strength, collapse, and carbon emissions of FA concrete, and verified the prediction accuracy and generalization ability of XGBoost in the field of concrete materials [53]. XGBoost creates the tree ensemble to improve the model accuracy by iteratively adding new CARTs, which can be represented by the k addition function, as described below.

$$\hat{y}_0=\sum_{k=1}^{K}f_k(x_i),f_k\in F \tag{6}$$

where K is the number of regression trees algorithm, f is a function in functional space (F), F is a set of all possible CARTs. Given the training dataset $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where x_{i} is the input variable, y_{i} is the response variable [54].

2.3. Shapley Additive Explanations

Most ML models have problems of high complexity and low interpretability. The previous method Feature Importance only tells which feature is important, but it is not clear how that feature affects the prediction result. Shapley Additive Explanations (SHAP) can be used to assess the importance of various features in a model and the impact of each feature on the predicted value [55]. SHAP, derived from the Shapley value [56], is a game theory-based approach where the output is a linear addition of the input variables, so it is clear whether the contribution of input variables to prediction is positive or negative. Given the training dataset $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where x_i is the input variable, y_i is the response variable. The explanation model G (x') with shortened input x' for an initial model $f(x)$ can be represented as Equation (7) [55].

$$F(x)=G(x')=\Phi_0+\sum_{i=1}^{k}(\Phi_i\times x_i') \tag{7}$$

where k is the number of input features. The mapping function $x=h_x(x')$ connects the inputs x' and x.

Kernel SHAP, Deep SHAP, and Tree SHAP are all methods for calculating SHAP values. Kernel SHAP, suitable for any model, was adopted to explain the predictions made by the XGBoost model. The SHAP model is a more comprehensive approach to model interpretability, which can be applied to both global interpretation and local interpretation. Moreover, this method has been applied to the sensitivity analysis of the bearing capacity and ductility of steel fiber reinforced concrete beams [57]. In order to conduct an in-depth analysis of its sensitivity, the global interpretation of SHAP Feature Importance (mean SHAP), SHAP summary Plot, and SHAP Dependence Plot is analyzed in this study.

2.3.1. SHAP Feature Importance

Some scholars have studied the Feature Importance of prediction models through tree-based models, such as using a random forest (RF) algorithm to obtain the variable importance [58] or using XGBoost to generate the feature importance of predictive tensile strength [54]. However, Feature Importance was based on the degradation of model performance, which resulted in feature importance ranking dependent on specific models. In contrast, the SHAP importance measure is based on the magnitude of feature attributes and changing the model does not change the importance of the feature, which indicates that it satisfies the 'consistency' requirement [59]. Take the average of the absolute value of each feature's SHAP value as the importance of that feature, and then rank these features from most important to least important to get a standard bar chart, which is the mean SHAP.

2.3.2. SHAP Summary Plot

SHAP summary Plot draws the Shapley value of each feature for each sample, showing which features are the most important and their influence range on the data set. The position on the y-axis is determined by the features, with each row representing one feature, and the position on the x-axis is determined by the per Shapley value. Color represents the characteristic value, and enables us to match how changes in feature values affect changes in risk. The overlap points jitter in the y-axis direction so we can understand the distribution of Shapley value for each feature and that these features are ranked according to their importance [60].

2.3.3. SHAP Dependence Plot

SHAP Partial dependence plot (PDP) is a global method that considers all instances and gives the global relationship between the features and the predicted results, showing the marginal effect of one or two features on the predicted results of a machine learning model. It can show whether the relationship between targets and features is linear, monotonous, or more complicated. To understand how a single feature affects the output of the model, the SHAP value for that feature can be compared to the characteristic value for all samples in the data-set. The PDP shows the change in the predicted value as the feature changes. The dispersion in the vertical direction of a single feature value represents interactions with other features, and to help reveal those interactions, the PDP automatically selects another feature for coloring. You can also set a specific second feature to color the scatter plot, showing the interaction effect of the second feature with the selected feature [60].

2.4. Prediction Performance Metrics

The performances of the models were evaluated by the following metric indexers: coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). R^2 ranges from 0 to 1, *MAE* and *RMSE* are used to evaluate the average error, and the most ideal prediction model is $R^2 = 1$, $RMSE = 0$, $MAE = 0$. The formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_p - y_e\right)^2}{\sum_{i=1}^{n}\left(y_e - \bar{y}\right)^2} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_e - y_p\right)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_e - y_p\right| \tag{10}$$

where y_e is the experimental value, y_p is the predicted value, and n is the number of samples.

2.5. Methodology Flow

The specific methodological process of this study is as Figure 3:

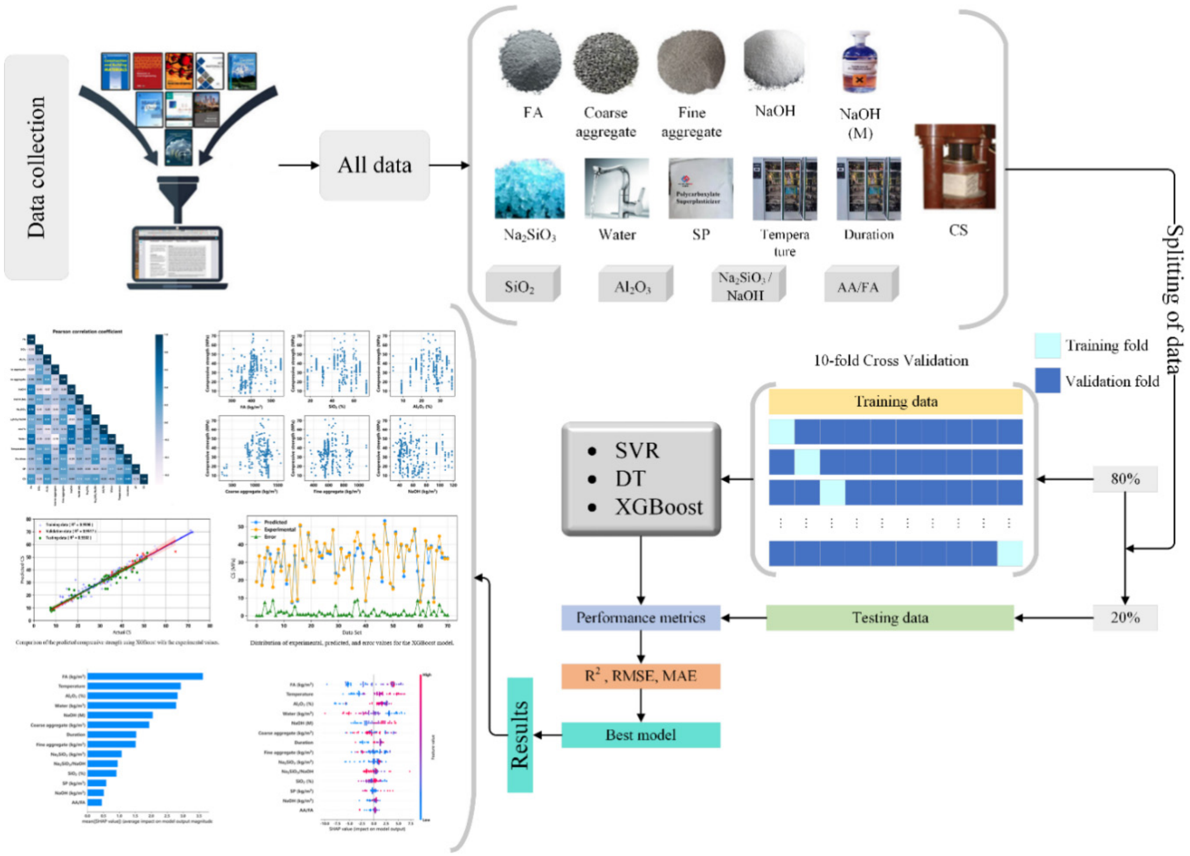


Figure 3. Methodology chart of the present work.

Firstly, 355 sets of experimental data on CS of fly ash (FA) were collected from 40 published literature, and the collected data were cleaned and normalized. The dataset is then randomly divided into 80% training set and 20% test set. Then, three machine learning models (SVR, DT and XGBoost) are trained using the training set, and the k-fold cross validation and statistical test methods are used to determine the optimal model. Finally, the Shap value was used to explain the XGBoost model, and the importance of each feature and its influence on the prediction results were evaluated. All algorithms are implemented based on Python 3.11 platform.

3. Results and Discussion

3.1. Assessment of the SVR Model

The results of the SVR model for predicting the CS of FA-GPC are shown in Figures 4 and 5 and Table 2. The correlation between the experimental results of the training set, validation set and testing set of the SVR model and the prediction results are shown in Figure 4. Figure 5 shows the dispersion of the experimental, predicted and error values for the testing set of the SVR model, where error values are the absolute value of the difference between the experimental value and the predicted value.

As can be seen from Figure 4 and Table 2, the R^2 score of the testing set of the SVR model is 0.927, which indicates that the prediction of the SVR model is relatively accurate and has less error compared with the experimental results. It can be clearly seen in Figure 4 that among the 71 pieces of testing data, there is one data with an error value of more than 10 MPa, and there are 5 data with an error value of 5~10 MPa. The error of all the remaining data is less than 5 MPa, and the average error of all the data is 1.821 MPa. Additionally, the A20 index, which measures the percentage of predictions within ±20% of the experimental value, is 92.96% for the testing set, 95.70% for the training set, and 89.29% for the validation set.

However, as can be seen from Figure 4, although the SVR model fits well in the training set (blue line) and the testing set (green line), the validation set (red line) deviates from the center line with an R^2 of 0.782. This phenomenon that the R^2 of the validation set is much lower than that of the training and the testing set indicates that the SVR, although it can predict the functional relationships of a given database, has poor generalization ability and overfitting. Other prediction performance metrics (RMSE, MAE) also verify this phenomenon.

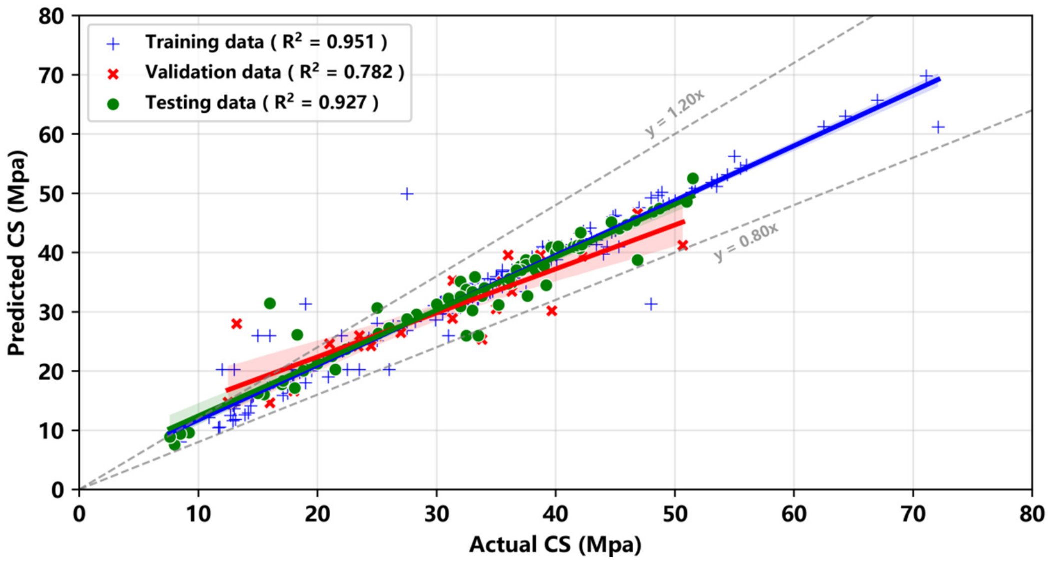


Figure 4. Comparison of the predicted compressive strength using SVR with the experimental values.

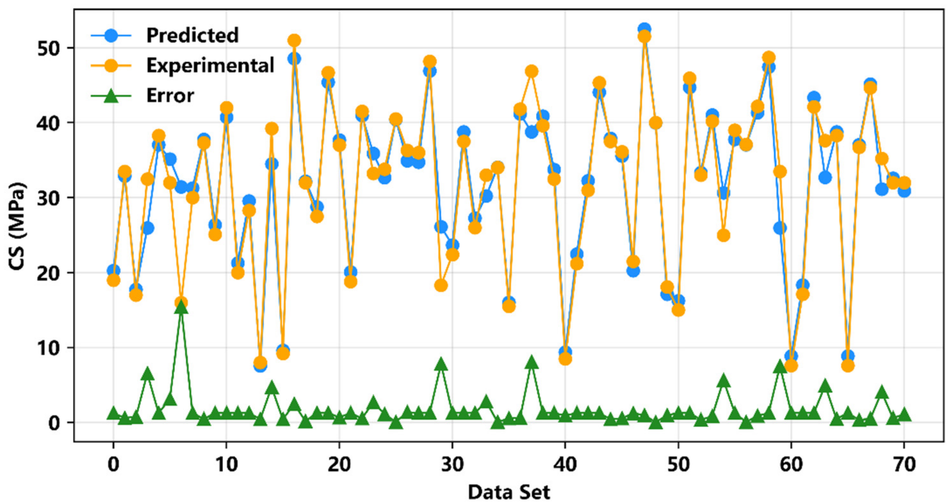


Figure 5. Distribution of experimental, predicted, and error values for the SVR model.

Table 2. Precision of three machine learning algorithms for the 28-day compressive strength of FA-GPC (MPa).

Model	Phase	MAE	RMSE	R^2	A20
SVR	Training set	1.5879	2.7955	0.9507	95.70%
	Validation set	2.9000	4.5270	0.7818	89.29%
	Testing set	1.8212	3.0243	0.9272	92.96%
DT	Training set	0.7773	2.0881	0.9704	96.47%
	Validation set	3.4519	4.3413	0.9030	89.66%
	Testing set	1.4605	2.8661	0.9346	90.14%
XGBoost	Training set	1.4467	2.4331	0.9598	95.29%
	Validation set	2.1014	3.0633	0.9517	96.55%
	Testing set	1.8080	2.9614	0.9302	91.55%

3.2. Assessment of the DT Model

The results of the DT model are shown in Figures 6 and 7 and Table 2. It can be seen from Figure 6 and Table 2 that the R^2 of the testing set for the DT model is 0.935, which is a higher prediction accuracy of the model compared to the SVR model (0.927). The error values of all the testing set data of the DT model are less than 10 MPa. There are 9 data with errors in the range of 5 to 10 MPa, while the errors of other data are less than 5 MPa, and the mean error is 1.460 MPa, which is 0.534 MPa lower than the mean error of the SVR model. The R^2 of the training set, validation set and testing set of the DT model are 0.970, 0.903 and 0.935 respectively, indicating that the training effect of the model is good, but the R^2 of the validation set still differs significantly from the training set and testing set, especially compared with the training set. The A20 index is 96.47% for the training set, 89.66% for the validation set, and 90.14% for the testing set. Compared to SVR, DT exhibits improved generalization and reduced overfitting.

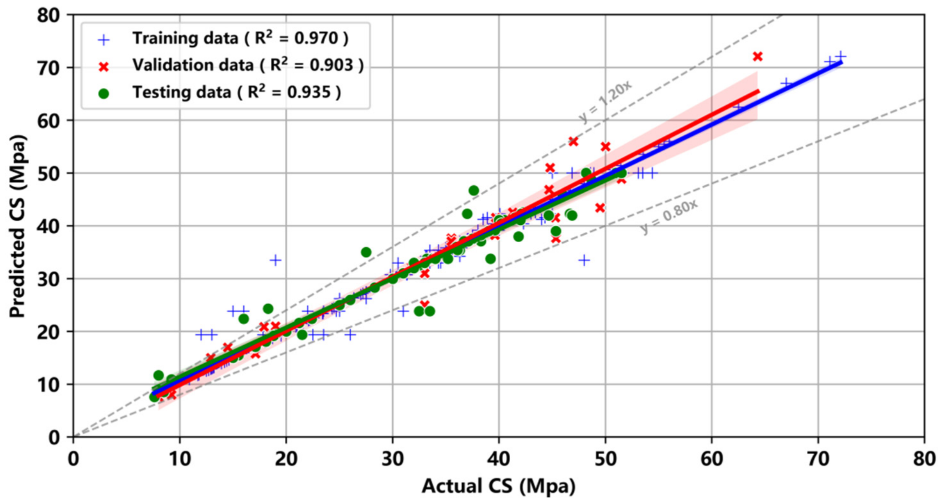


Figure 6. Comparison of the predicted compressive strength using DT with the experimental values.

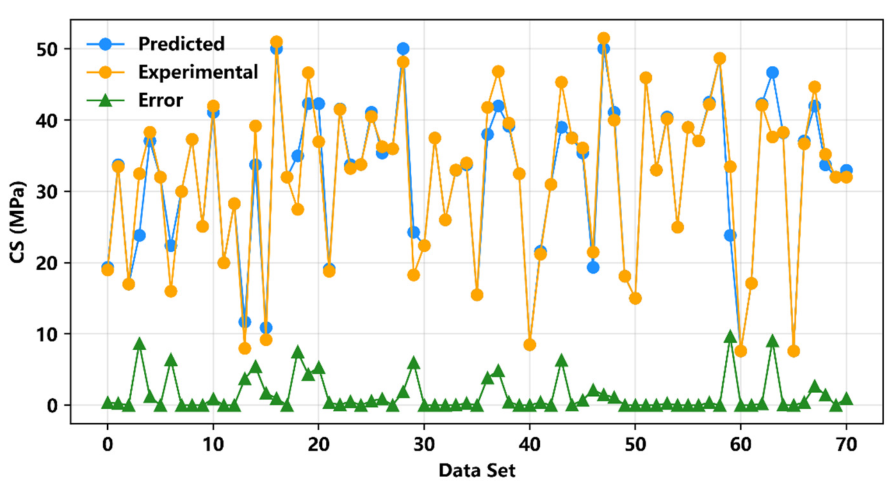


Figure 7. Distribution of experimental, predicted, and error values for the DT model.

3.3. Assessment of the XGBoost Model

As shown in Figures 8 and 9, R^2 values of the XGBoost model training set, validation set, and testing set are 0.960, 0.952, and approximately 0.930 or higher, and the difference between the three data was small. This indicates that the XGBoost model not only has a good prediction effect, but also has no overfitting phenomenon and strong generalization ability, which is suitable for CS prediction of FA-GPC. In the data of the testing set, there are 9 prediction errors within 5~10 MPa, the error of all the remaining prediction data is less than 5 MPa, and the average error of all is 1.808 MPa. The A20 index further confirms the model’ s accuracy and robustness, with 95.29% for training, 96.55% for validation, and 91.55% for testing, demonstrating excellent generalization ability.

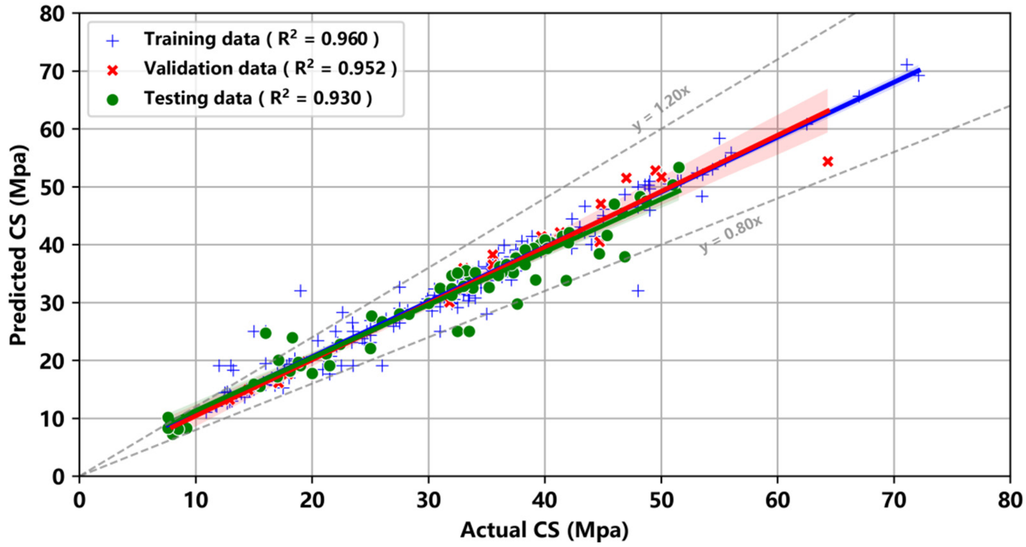


Figure 8. Comparison of the predicted compressive strength using XGBoost with the experimental values.

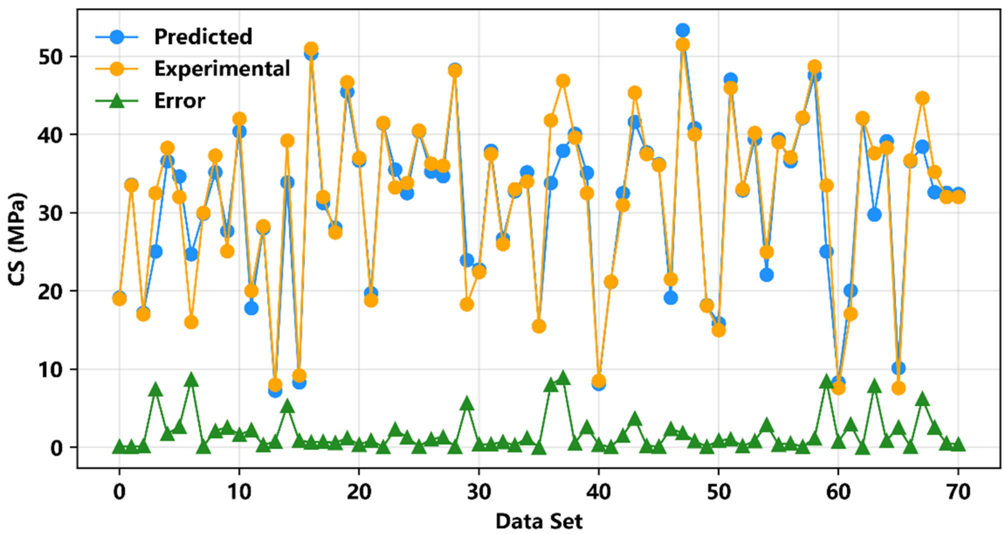


Figure 9. Distribution of experimental, predicted, and error values for the XGBoost model.

3.4. Models’ Validation

In order to minimize the random sampling bias of training, the K-fold cross-validation algorithm was used to evaluate the model performance. The steps of K-fold verification are as follows: first, all samples are divided into K sample subsets of equal size, and then these k subsets are iterated successively, each time taking the current subset as the validation set and all the remaining samples as the training set for model training and evaluation, and finally the average value of K evaluation indicators is taken as the final evaluation index. The results of Kohavi’s study showed that the ten-fold validation test has reliable variance and optimal computation time [61,62]. Therefore, this study applied 10-fold cross-validation for model evaluation.

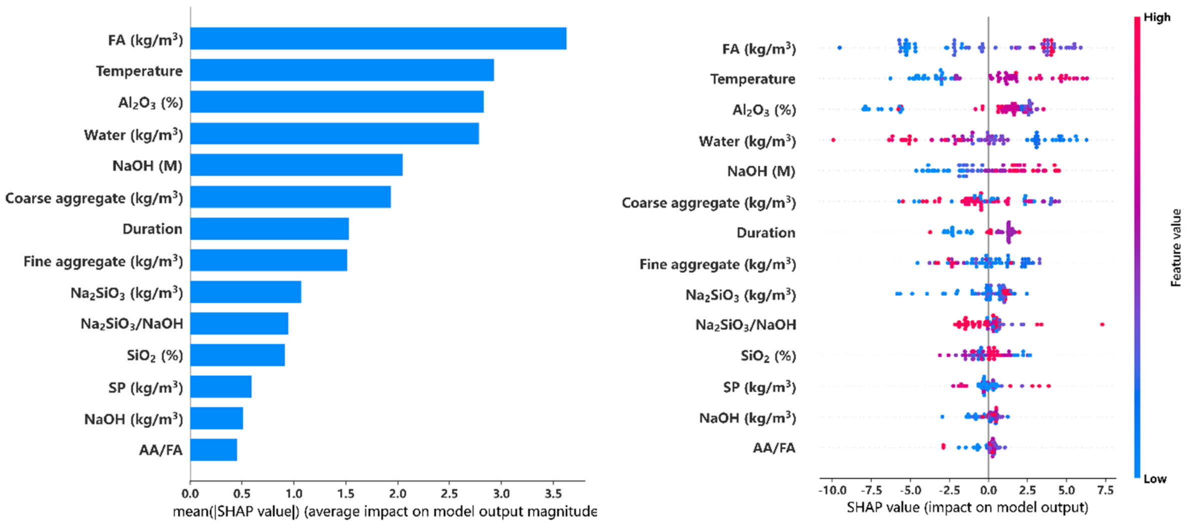
The R^2, RMSE, and MAE were calculated to evaluate the validity of the k-fold validation, as shown in Table 3. the MAE value of SVR varies from 1.821 to 2.184 MPa with an average of 2.004 MPa, for DT, the MAE value range is 1.461~3.191 MPa, and the average value is 2.279 MPa. In addition, the MAE of XGBoost technology varies between 1.808 and 2.475 MPa with an average of 1.992 MPa. While the average RMSE value of SVR, DT, and XGBoost models are 3.270, 4.991, and 3.439 MPa, respectively. According to Table 3, it can be seen that the average R^2 of SVR, DT, and XGBoost are 0.915, 0.787, and 0.903, respectively, which indicates that the CS of GPC is more accurately predicted by SVR and XGB models. According to the analysis of SVR and XGBoost models in the previous section, XGBoost not only has high prediction accuracy, but also has strong generalization ability, so the XGBoost model is suitable for predicting CS of FA-GPC.

Table 3. Results of the k-fold process.

K-Fold	SVR			DT			XGBoost		
	MAE	**RMSE**	**R^2**	**MAE**	**RMSE**	**R^2**	**MAE**	**RMSE**	**R^2**
1	2.177	3.479	0.904	2.801	5.238	0.782	1.867	2.598	0.946
2	1.994	3.248	0.916	1.461	2.866	0.935	1.808	2.961	0.930
3	2.000	3.281	0.914	2.364	5.458	0.763	2.089	3.567	0.899
4	2.184	3.607	0.896	1.761	3.702	0.891	1.922	3.338	0.911
5	1.821	3.024	0.927	1.959	4.986	0.802	1.898	3.133	0.922
6	1.923	3.120	0.923	3.006	6.995	0.610	2.068	3.540	0.900
7	2.043	3.225	0.917	2.486	5.714	0.740	1.843	3.148	0.921
8	1.959	3.284	0.914	3.191	7.003	0.610	2.475	4.978	0.803
9	1.984	3.265	0.915	2.273	4.778	0.818	1.833	3.201	0.918
10	1.954	3.172	0.920	1.485	3.171	0.920	2.114	3.921	0.878
Average	2.004	3.271	0.915	2.279	4.991	0.787	1.992	3.439	0.903

4. Shapley Additive Explanation (SHAP) Analysis

Black-box models with highly complex and non-linear architectures like XGBoost are not easily interpreted. Therefore, we used the SHAP value agnostic tool to interpret the prediction model. Figure 10a shows the SHAP value results for all feature variables, which are the average of the absolute Shapley values for each input variable feature in the entire sample data. The results show that the SHAP value of FA content is the largest when predicting the CS of FA-GPC. The feature "curing temperature" has the second maximum SHAP value. Overall, FA content, curing temperature, Al_2O_3 content in FA, water and NaOH(M) are the most important features for predicting the CS using XGBoost.



(**a**) (**b**)

Figure 10. The SHAP analysis utilizing (**a**) mean-SHAP values, (**b**) SHAP summary Plot.

Whether the input variable has a positive or negative effect on the output result is not known from Figure 10a, but can be obtained from the SHAP summary plot in Figure 10b. Figure 10b shows a summary plot of the SHAP values for each feature using the XGBoost model to predict the CS of FA-GPC, for each feature, the color indicates the feature value and its corresponding x-axis (SHAP) value indicates the contribution to the output attributes, if the SHAP values of the feature are centered around zero, it indicates that they are of low importance to the model output and have no significant effect on the CS in the data set.

Based on the conclusion form the data set, we can analyze each input parameter in combination with the influence patterns of micro or macro effects:

4.1. Positive Driving Factor

According to the Figure 10b, it is easy to explain the influence of the feature "FA content", the red points on the right and the blue points on the left are the high and low values of FA content, respectively, which indicates that FA content has a positive impact on the prediction result of CS, and the predicted value of CS increases when the FA content is increased. The reactivity of FA largely depends on its chemical composition, because the content of Al_2O_3 and SiO_2 in FA affects the Pozzolanic reaction, they are activated by portlandite formed during the hydration process of cement, and more hydrated gel is generated to fill the pores in concrete [63]. Therefore, the CS of FA-GPC is greatly affected by the content of Si and Al in FA and the amorphous Si/Al ratio [64,65].

The CS of FA-GPC increases with increasing curing temperature and curing time due to the higher curing temperature and the longer curing time leading to the formation of more reaction products [66]. However, long-term curing at high temperature will destroy the granular structure of the geopolymer, leading to dehydration and excessive shrinkage [67], so there are some data that the CS decreases when the curing time increases. The optimal curing temperature of fly ash geopolymer is 60–80 °C. While the optimal curing time is affected by the completeness of the hydration process of the geopolymers and the consumption degree of Na^+ ion and OH^- ion. Under an environment with a sufficient concentration of NaOH of a certain modulus, the complete strength acquisition cycle can be completed within a relatively short curing period of only 3 days [68,69].

The concentration of the chemical activator has a significant effect on the mechanical properties of the geopolymer, the dissolution of Si and Al during the synthesis of the geopolymer largely depends on the concentration of NaOH [70]. With the increase of NaOH concentration, the more the breakage of the glassy chain at the surface of the FA, the better the dissolution effect of FA particles, the better the polymerization effect, and the higher the compressive strength [69,71].

4.2. Negative Inhibitory Factor

The negative effect of water on the CS is because when the amount of liquid in the system increases, the porosity increases, which reduces the compactness of the material and hinders the polymerization process [72]. Therefore, reducing the water content in the synthesis water phase has a positive effect on the polycondensation process of the oligomers and the hardening of the system.

4.3. Coupling Influencing Factors

The influence mechanisms of 'Na_2SiO_3/NaOH', 'SiO_2/Al_2O_3' and 'AA/FA' on CS are similar. They are all a pair of coupled variables interacting and influencing each other deeply. They also have similar influence laws, that is, when the feature value is high, the effect on the CS is partly positive and partly negative. As can be seen from the figure, the addition of soluble sodium silicate to the geopolymer system increases the mechanical properties of geopolymer material, because soluble sodium silicate increases the SiO_2/Na_2O ratio in the system, leading to a gradual shift of the chemical system from the monosilicate, chains and cyclic trimers to complex structures and finally to a three-dimensional polymeric framework [70]. However, when Na_2SiO_3/NaOH ratio exceeds a certain threshold value (generally 2.5), the presence of high content of cyclic silicate species hinders the polycondensation process and will have a negative impact on the CS. Moreover, the deficiency of NaOH in the mixture prevents the dissolution of reactive silica and alumina to complete the condensation process, which will also reduce the strength of the concrete [73].

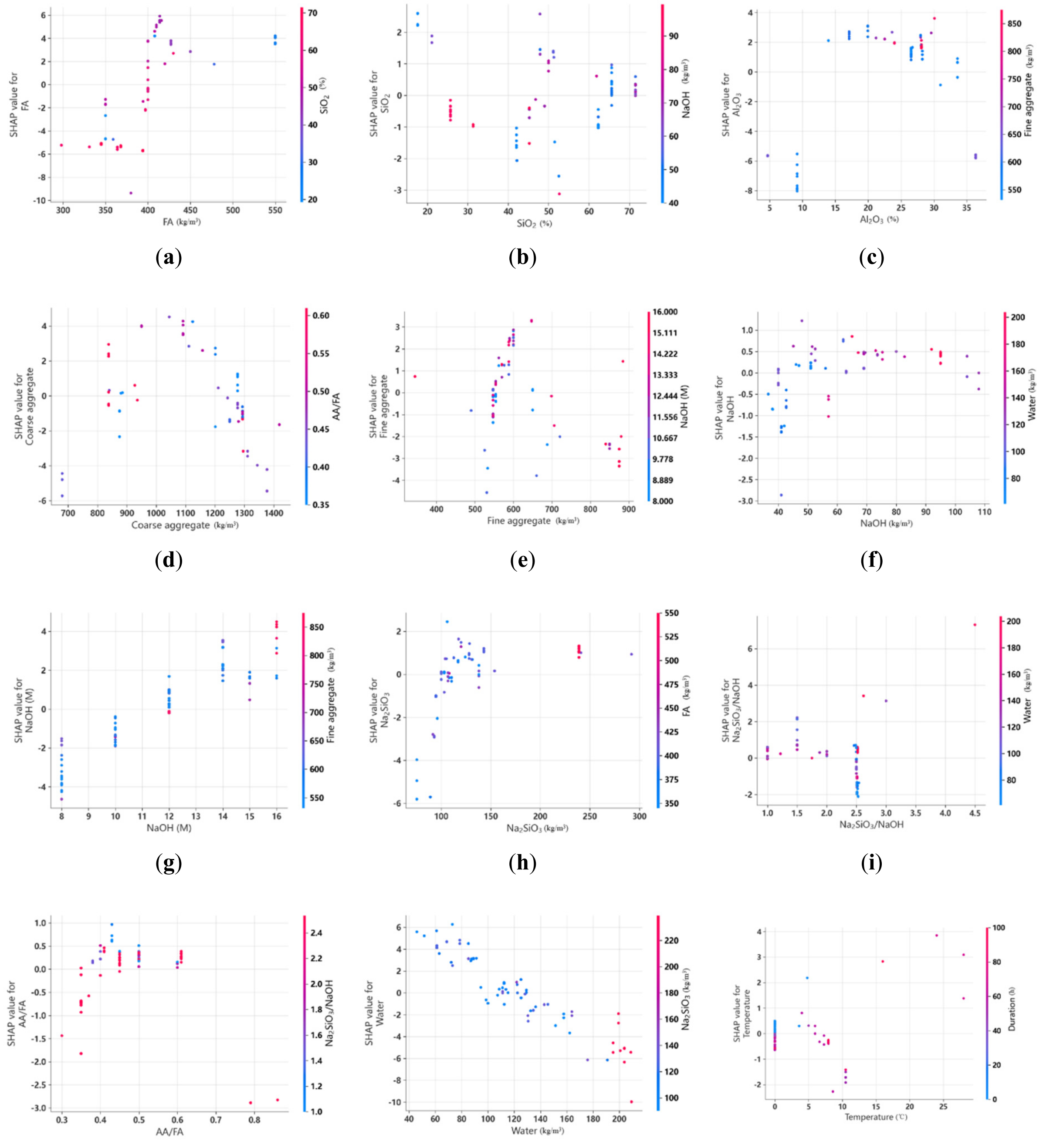
A similar phenomenon is observed for AA/FA, there is an optimum AA/FA ratio for the CS of FA mortars. When the AA/FA ratio is lower than the optimal value, the amount of alkaline activator required to dissolve the FA particles is insufficient, when the amount of AA/FA is higher than the optimal value, the excess alkaline activator inhibits the polymerization process, both of which will lead to a reduction in the CS of the geopolymer [74].

Al_2O_3/SiO_2 is even more special. As the main components of FA, they form a mutual influence mechanism of waxing and waning. Therefore, according to Figure 11b, it can be found that the influence of Al_2O_3 (%) alone on the SHAP value of CS is roughly that it first increases and then decreases with the increase of the proportion of Al_2O_3, while for SiO_2, it first decreases and then increases. Due to the trade-off between the two, that is, when the ratio of the two is within a certain appropriate range, the best effect on CS will be achieved. This point is also consistent with the conclusion of the Si/Al ratio in FA mentioned earlier [64,65], and the optimal Si/Al is about 2.5 [75–77].

4.4. Specific Conditional Dependent Factors

Although SP should theoretically have an important influence on CS, in fact, its role depends greatly on the specific environment. the addition of SP may lead to a decrease in CS due to their instability in highly alkaline media [78], it has also been shown that some kinds of SP are stable in alkaline media and do not interfere with the polymerization process, enhancing workability and reducing the total porosity thus making the microstructure of the geopolymer denser and instead improving the compressive strength [79]. In addition to this, Puertas et al. [80] reported that the addition of vinyl copolymers and polyacrylate copolymer-based SP to FA-based geopolymer paste and mortar did not produce any significant change in the compressive strength of the geopolymers. However, excessive SP will have adverse effects on CS [81]. The reason is that it will accelerate bleeding and segregation, resulting in a decrease in the CS value [82].

The above assessment is based on the database suggested by the current study and more data may yield more accurate results.

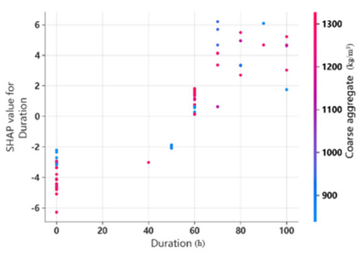


(**a**) (**b**) (**c**)

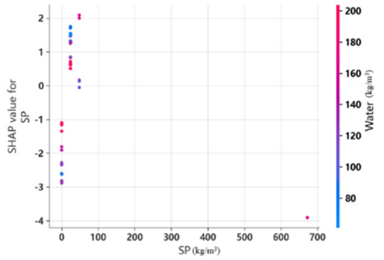
(**d**) (**e**) (**f**)

(**g**) (**h**) (**i**)

(**j**) (**k**) (**l**)



(m)



(n)

Figure 11. SHAP dependency plots for the compressive strength of fly ash-based geopolymer concrete. (**a**) The interaction between FA and SiO_2; (**b**) The interaction between SiO_2 and NaOH; (**c**) The interaction between Al_2O_3 and Fine aggregate; (**d**) The interaction between Coarse aggregate and AA/FA; (**e**) The interaction between Fine aggregate and NaOH(M); (**f**) The interaction between NaOH and Water; (**g**) The interaction between NaOH(M) and Fine aggregate; (**h**) The interaction between Na_2SiO_3 and FA; (**i**) The interaction between Na_2SiO_3/NaOH and Water; (**j**) The interaction between AA/FA and Na_2SiO_3/NaOH; (**k**) The interaction between Water and Na_2SiO_3; (**l**) The interaction between Temperature and Duration; (**m**) The interaction between Duration and Coarse aggregate; (**n**) The interaction between SP and Water;.

4.5. SHAP Partial Dependence Plot (PDP)

The SHAP Partial dependence plot (PDP) technique can be used to understand the effect of a certain feature on the predictive performance of the model and its relationship with other variables. Figure 11 shows the SHAP PDP for all input variables. Taking Figure 11a,d,k as an example, they have obvious scatter distribution patterns. The interaction between FA and SiO_2 is shown in Figure 11a, which shows that the increase in FA leads to an increase of CS. In contrast, Figure 11k shows that an increase in water content has a negative effect on CS. In addition, the CS value of FA-GPC increases and then gradually decreases with the increase of coarse aggregate content as shown in Figure 11d, which is consistent with the SHAP summary plot in Figure 10b. This is because the CS of FA mortar depends on the bond between the binder and the aggregate, the surface area, the surface texture and the angles [83], there is an optimal balance between the aggregate and the binder content of the FA-GPC materials, if the aggregate content is too high, the amount of binder is not sufficient to coat the entire surface of the aggregate. Without a uniform consolidation material [84], the CS of concrete cannot be guaranteed. In addition, it can be seen from Figure 11a that when the FA content is greater than 400 kg/m^3 and SiO_2 in FA is less than 50%, it will have a positive impact on the CS of FA-GPC. In Figure 11k, the optimal content of coarse aggregate is 1000–1100 kg/m^3, and the CS decreases when the content exceeds the optimal value. when the content of water exceeds 100 kg/m^3, CS will be negatively affected regardless of the value taken for Na_2SiO_3, as shown in Figure 11d.

5. Experimental Verification

In order to verify the prediction value of the strength model of fly ash geopolymer concrete based on XGBoost, the mix ratio was designed and the actual strength of fly ash base polymer concrete was tested by pouring.

5.1. Raw Material Performance and Mix Ratio Design

Low calcium fly ash (ASTM grade F) was used, and the Si: Al ratio and specific gravity of the fly ash used were both 1.9. And 90% of the fly ash could pass through a 45 µm sieve. Table 4 lists the chemical composition of fly ash determined by X-ray fluorescence analysis.

Table 4. Chemical Composition of fly ash (%).

SiO_2	Al_2O_3	$Fe_2O_3 + Fe_3O_4$	CaO	Na_2O	MgO	MnO_2	TiO_2	SO_3	Others	Loss of Burn
59.70	28.36	4.57	2.10	0.04	0.83	0.04	1.82	0.40	2.14	1.06

The alkali solution used is a mixture of NaOH and Na_2SiO_3 solutions. NaOH solution at the desired molar concentration of 10 mol/L was prepared with 98% purity NaOH solid. The content of SiO_2, Na_2O and water in

Na_2SiO_3 solution is 34.64%, 16.27% and 49.09%, respectively. The NaOH and Na_2SiO_3 solutions were fully mixed 24 h in advance to prepare the alkali solution, and the temperature was lowered to ambient temperature before use. Crushed granite was used for coarse aggregate and natural river sand was used for fine aggregate. The nominal size of coarse aggregate is 20 mm, and the nominal size of fine aggregate is 4.5 mm. The relative densities of coarse aggregate and fine aggregate are 2.72 and 2.64, respectively, and the fineness modulus of fine aggregate is 2.36.

In order to verify the effect of aggregate content, the ratio of sodium silicate to sodium hydroxide, and the ratio of alkali to fly ash on the 28-day compressive strength of fly ash base polymer concrete shown in shap analysis, three groups of 12 mixing ratios were designed, as shown in Table 5. Among them, the units of temperature and time are degrees Celsius (C) and hours (h), respectively, and the units of other non-ratio parameters are all $kg{\cdot}m^{-3}$. M stands for Mix ratio group, A stands for Aggregate content, S stands for the ratio of Na_2SiO_3/NaOH, L stands for Alkaline solution which represents the ratio of alkali activator solution to fly ash. For example, the mixture ratio of the number M1A60, where M1 is the mix ratio of the M1 group, with the aggregate content as the variable parameter, A60 represents the aggregate content of 60% of the concrete volume, similarly M2S20 is the mix ratio of sodium silicate and sodium hydroxide as the variable parameter of the M2 group, and the ratio of sodium silicate and sodium hydroxide is 2.0.

Table 5. Mix ratio of fly ash geopolymer concrete.

Group Name	FA ($kg{\cdot}m^{-3}$)	Coarse Aggregate ($kg{\cdot}m^{-3}$)	Fine Aggregate ($kg{\cdot}m^{-3}$)	NaOH ($kg{\cdot}m^{-3}$)	Na_2SiO_3 ($kg{\cdot}m^{-3}$)	Na_2SiO_3/NaOH	AA/FA	Water ($kg{\cdot}m^{-3}$)	SP ($kg{\cdot}m^{-3}$)	Curing Temperature (°C)	Curing Time (h)
M1A60	421	1032	556	66	165	2.5	0.55	113.27	8	100	24
M1A65	365	1118	602	60	151	2.5	0.55	115.29	7	100	24
M1A70	310	1204	648	49	122	2.5	0.55	97.50	6	100	24
M1A75	255	1290	695	40	100	2.5	0.55	79.53	5	100	24
M2S20	319	1204	648	48	96	2.0	0.45	96.05	6	100	24
M2S25	318	1204	648	41	102	2.5	0.45	95.20	6	100	24
M2S30	316	1204	648	36	107	3.0	0.45	96.22	6	100	24
M2S35	315	1204	648	32	110	3.5	0.45	95.95	6	100	24
M3L35	314	1204	648	31	78	2.5	0.35	101.56	6	100	24
M3L45	305	1204	648	39	98	2.5	0.45	102.86	6	100	24
M3L55	298	1204	648	47	117	2.5	0.55	104.68	6	100	24
M3L65	290	1204	648	54	135	2.5	0.65	103.32	6	100	24

5.2. Model Feedback Optimization

According to Shap additive analysis, six important factors affecting the compressive strength of FA-GPC are obtained: FA, SiO_2, Al_2O_3, Water, curing time, curing temperature. The optimization modification direction of XGBoost model is proposed, which is to increase the influence weight of important factors on the model prediction results, so as to ensure the prediction accuracy of the model. By creating polynomial features, the important features are extended polynomially, while the other features remain unchanged. Due to the limited amount of data in the database, in order to ensure the robustness of the model simulation prediction and avoid over-fitting, only the important features are extended by quadratic polynomial, and the two features of curing time and curing temperature are not extended by quadratic polynomial.

The R^2, MAE and RMSE of the optimized model corresponding to the training set are 0.976, 1.1194 and 2.0261, respectively, and the R^2, MAE and RMSE of the test set are 0.814, 3.0294 and 4.8350, respectively. The comparison between the predicted compressive strength of the optimized model and the experimental value is shown in Figure 12. The R^2 gap between the training set and the test set is large, and there is a tendency of overfitting. It is judged that the expansion of the quadratic term leads to the increase of the total number of features, which affects the robustness of the model simulation.

5.3. Test and Prediction Model Results

Table 6 shows the comparison of the test intensity values obtained for all mixture trials with the intensity values predicted by the XGBoost model and the optimized XGBoost model, where the errors are presented as relative percentages in Table 6.

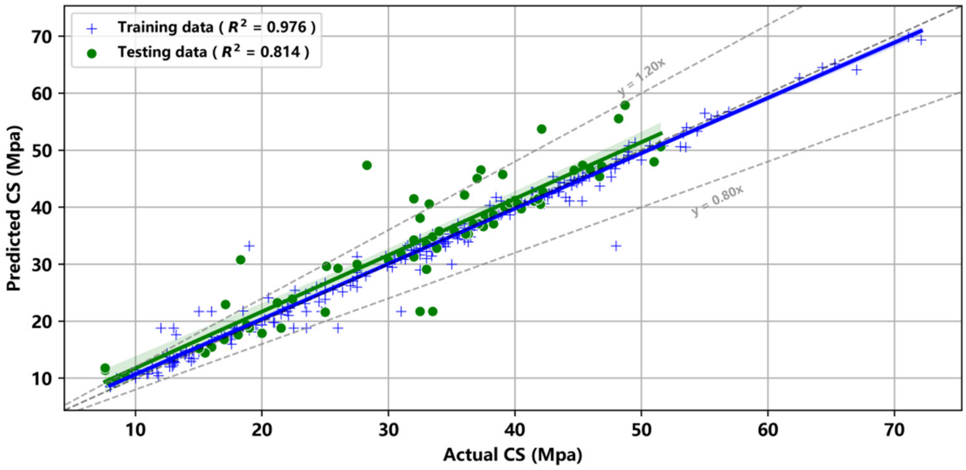


Figure 12. Comparison of the optimized XGBoost predicted compressive strength with the experimental value.

Table 6. Comparison of the test strength value and the predicted strength value of the mixture ratio (Relative error = (predicted value − test value)/test value).

Group Name	Strength of Experiment (Mpa)	Strength of Prediction (Mpa)	Prediction Strength of the Optimized Model (Mpa)	Absolute Error	Absolute Error of Model Prediction after Optimization	Relative Error	Relative Error of Model Prediction after Optimization
M1A60	45	45.66	45.60	0.66	0.6	1.47%	1.33%
M1A65	47	49.43	47.18	2.43	0.18	5.17%	0.38%
M1A70	56	52.50	53.57	3.5	2.43	6.25%	4.34%
M1A75	49	49.83	49.37	0.83	0.37	1.69%	0.76%
M2S20	40	46.42	48.02	6.42	8.02	16.05%	20.05%
M2S25	47	49.17	48.08	2.17	1.08	4.62%	2.30%
M2S30	43	44.05	40.96	1.05	2.04	2.44%	4.74%
M2S35	38	45.31	44.40	7.31	6.4	19.24%	16.84%
M3L35	35	34.31	35.76	0.69	−0.76	1.97%	2.17%
M3L45	41	40.94	47.60	0.06	−6.6	0.15%	16.10%
M3L55	52	52.93	51.53	0.93	0.47	1.79%	0.90%
M3L65	42	53.31	50.95	11.31	−8.95	26.93%	21.31%

Table 6 shows that the average absolute error of all mixtures is 2.41 MPa, which is lower than 3 MPa, indicating high accuracy. The absolute errors of M1A60, M1A75, M3L35, M3L45 and M3L55 are all less than 1 MPa. The average absolute error of the optimized model is 2.34, and the absolute error is reduced. Although the average error on these 12 mixtures is slightly lower, the model is not preferred due to overfitting. The absolute errors of M1A60, M1A65, M1A75, M3L35 and M3L55 are all less than 1MPa. According to Table 6, the influence law of each parameter on the compressive strength of FA-GPC can be observed. The test results of the first group of mix ratios show that the compressive strength of fly ash geopolymer concrete gradually increases with the increase of the total aggregate content, and the strength decreases when the aggregate content reaches 70%. The results of the second mix ratio show that the compressive strength of fly ash geopolymer concrete increases with the increase of sodium silicate and sodium hydroxide content, and decreases gradually after reaching 2.5. The increase of compressive strength is mainly due to the increase of sodium silicate dosage and the denser microstructure of geopolymer. On the other hand, the reduction of compressive strength is due to the fact that under the high ratio of sodium silicate to sodium hydroxide, in the process of geopolymer formation, the lack of sodium hydroxide in the mixture hinders the condensation process of active silica and alumina dissolution in the binder, which will reduce the strength of concrete [73]. When the ratio of alkali to fly ash in group 3 reaches 0.55, the compressive strength gradually increases, and when it exceeds 0.55, the compressive strength decreases, because excessive alkali will hinder the condensation of silicate [85].

Figure 13 compares the relative errors of the original XGBoost model and the polynomial-expanded model for the 12 additional experimental mixtures. Although the polynomial-expanded model exhibits slightly smaller and more stable relative errors for some mixtures, this result should be interpreted cautiously. The introduction of quadratic terms increased the dimensionality and complexity of the feature space and resulted in a substantial difference between the training and test performance, with (R^2) decreasing from 0.976 to 0.814. This performance gap indicates that the expanded model was affected by overfitting and that its apparent improvement on the 12 mixtures does not provide sufficient evidence of superior generalization. Therefore, the polynomial-expanded model is not considered the preferred general-purpose model. The original XGBoost model remains the recommended model because it achieves a more balanced performance across the training, validation, test, and cross-validation results. Future work should increase the dataset size, strengthen regularization, limit model complexity, and retain only physically meaningful interaction or polynomial terms to improve generalizability and reduce the risk of overfitting.

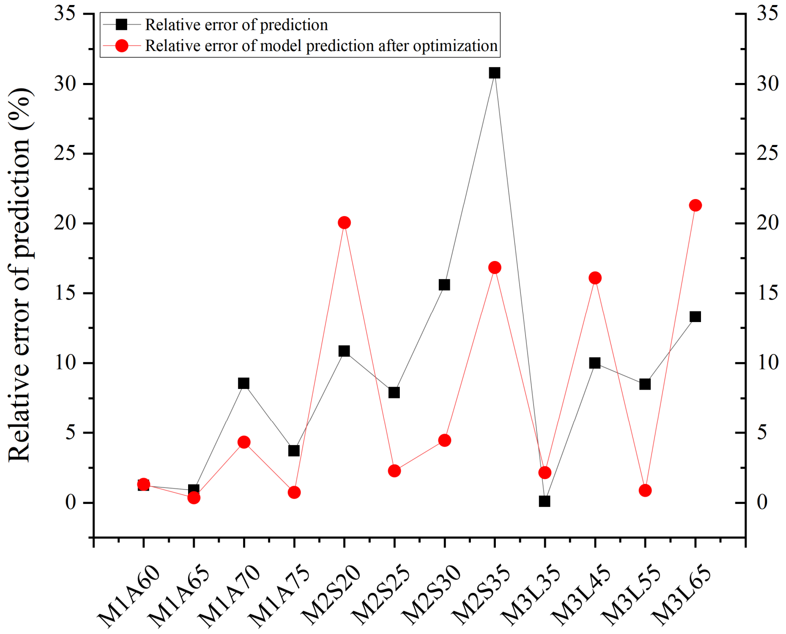


Figure 13. Comparison of the relative prediction error of the model before and after optimization.

5.4. Graphic User Interface (GUI) Design

The results show that XGBoost model has high accuracy and good performance in predicting the 28d compressive strength of FA-GPC. A Graphical User Interface (GUI) enables materials engineers and researchers to utilize the powerful functions of machine learning models without requiring extensive programming or technical knowledge. Therefore, in order to enhance the usability and accessibility, a GUI based on the XGBoost algorithm for predicting the 28d compressive strength of FA-GPC was developed for the user to seamlessly interact with the model. As shown in Figure 14, the computer software encourages users to manually input the required parameters, and directly displays the predicted compressive strength value by clicking the predict button. This development not only simplifies the application of the model in real scenes, but also helps to optimize the decision-making process in the field of mix design.

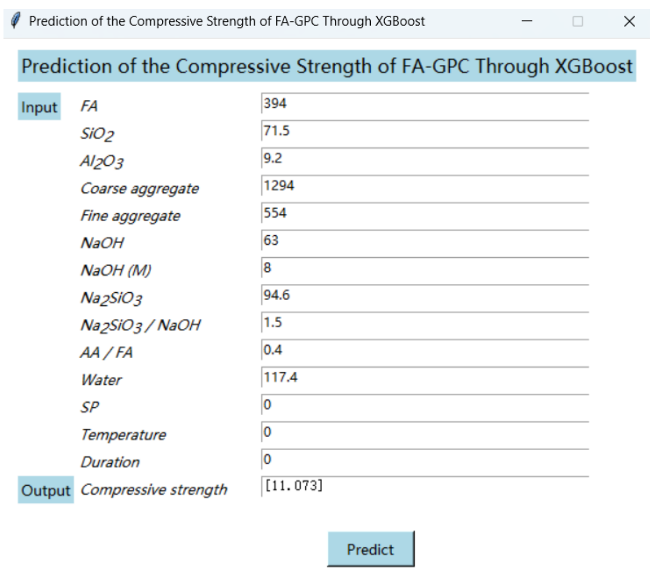


Figure 14. GUI of prediction of 28d compressive strength of FA-GPC through XGBoost.

6. Conclusions

Through a detailed literature search, 355 sets of mix ratio data were collected. This study aims to predict the 28d compressive strength (CS) of fly ash-based geopolymer concrete (FA-GPC) using individual and ensemble machine learning (ML) methods. Two individual ML methods, namely Support Vector Regression (SVR) and Decision Tree (DT), and an ensemble ML method Extreme Gradient Boosting (XGBoost) were used to predict the CS of FA-GPC. The collected data included 14 different input variables. R^2, RMSE and MAE were used to evaluate the performance of each model. The performance of the models was verified by K-fold cross-validation, and then, Shapley Additive Explanations (SHAP) was used to quantify the influence and significance of the considered inputs. Finally, according to the key factors obtained by Shapley Additive Explanations, the binomial expansion of the key features of the model was proposed to optimize the model, and the prediction effect of the model was verified by experiments. The following conclusions can be drawn from the obtained results:

SVR, DT and XGBoost all performed well in terms of testing set prediction, the R^2 are 0.927, 0.935, and 0.930 respectively. The performance of the adopted models was verified using k-fold cross-validation, and the results showed that the average R^2 of SVR, DT, and XGBoost was 0.915, 0.787, and 0.903, respectively, and the accuracy of SVR and XGBoost models is superior in predicting CS of FA-GPC. However, SVR has an overfitting phenomenon and weak generalization ability. Therefore, among these three models, the XGBoost model performs best.

Based on the results of SHAP analysis, FA content, curing temperature, content of Al_2O_3 in FA, water and NaOH concentration were the most important features. FA content, NaOH concentration, curing temperature and curing time had a strong positive correlation with CS, Al_2O_3 and SiO_2 of FA had a greater effect on CS, while water showed a significant negative effect on CS. For this dataset, there was an optimal value for Na_2SiO_3/NaOH, Al_2O_3/SiO_2, and AA/FA. When the optimal value was exceeded, CS gradually decreased. For SP, it has shown different influence effects on CS in different environments, and its specific mechanism still needs to be further explored.

The results show that using the XGBoost model to predict the strength of the fly ash base polymer concrete test mix can obtain high compressive strength prediction accuracy, and the average absolute error of all mix ratios is 2.41MPa. The binomial expansion of the key features of the model has a certain effect, but it is necessary to further adjust the parameters and expand the data set to improve the robustness of the model and avoid overfitting

It is shown that using ML to predict the mechanical properties of FA-GPC can achieve high accuracy in predicting the compressive strength of GPC. These initiatives can accelerate the development and application of geopolymer concrete.

Although the proposed XGBoost model demonstrated satisfactory predictive accuracy and interpretability, its performance remains dependent on the characteristics of the training dataset. The model may be sensitive to changes in data distribution and may exhibit reduced prediction accuracy when applied to FA-GPC mixtures outside the range represented in the current database. In addition, the experimental validation was performed using a limited number of mix proportions, which may not fully capture the variability of practical engineering applications. Therefore, future work should focus on expanding the dataset with more diverse materials and curing conditions, as well as conducting large-scale external validation to further improve model robustness, transferability, and generalization capability.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/2607221602237891/DS-26040200-SI.zip.

Author Contributions

X.S.: Conceptualization, Project administration, Funding acquisition; H.H.: Formal analysis, Investigation, Writing—review & editing; Z.D.: Resources, Supervision; Y.L.: Methodology, Software, Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding

We would like to thank the financial support from National Natural Science Foundation of China General Program (52578313), Science and Technology Department of Sichuan Province (24QYCX0386), the Sichuan Provincial Engineering Research Center of City Solid Waste Energy and Building Materials Conversion and Utilization Technology (GF2024ZD01).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The raw data and processed data required to reproduce these findings are available on reasonable request to shixs@scu.edu.cn.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

Prime Novelty Statement

We confirm that this manuscript has not been published previously by any of the authors and is not under consideration for publication in another journal.

References

1. Benhelal, E.; Zahedi, G.; Shamsaei, E.; et al. Global strategies and potentials to curb CO_2 emissions in cement industry. *J. Clean. Prod.* **2013**, *51*, 142–161. https://doi.org/10.1016/j.jclepro.2012.10.049.
2. Almutairi, A.L.; Tayeh, B.A.; Adesina, A.; et al. Potential applications of geopolymer concrete in construction: A review. *Case Stud. Constr. Mater.* **2021**, *15*, e00733. https://doi.org/10.1016/j.cscm.2021.e00733.
3. Xu, J.-H.; Yi, B.-W.; Fan, Y. A bottom-up optimization model for long-term CO2 emissions reduction pathway in the cement industry: A case study of China. *Int. J. Greenh. Gas Control* **2016**, *44*, 199–216. https://doi.org/10.1016/j.ijggc.2015.11.028.
4. Meshram, R.B.; Kumar, S. Comparative life cycle assessment (LCA) of geopolymer cement manufacturing with Portland cement in Indian context. *Int. J. Environ. Sci. Technol.* **2022**, *19*, 4791–4802. https://doi.org/10.1007/s13762-021-03336-9.
5. Meyer, C. The greening of the concrete industry. *Cem. Concr. Compos.* **2009**, *31*, 601–605. https://doi.org/10.1016/j.cemconcomp.2008.12.010.
6. Van Deventer, J.S.J.; Provis, J.L.; Duxson, P. Technical and commercial progress in the adoption of geopolymer cement. *Miner. Eng.* **2012**, *29*, 89–104. https://doi.org/10.1016/j.mineng.2011.09.009.
7. Woszuk, A.; Bandura, L.; Franus, W. Fly ash as low cost and environmentally friendly filler and its effect on the properties of mix asphalt. *J. Clean. Prod.* **2019**, *235*, 493–502. https://doi.org/10.1016/j.jclepro.2019.06.353.
8. Shang, J.; Dai, J.-G.; Zhao, T.-J.; et al. Alternation of traditional cement mortars using fly ash-based geopolymer mortars modified by slag. *J. Clean. Prod.* **2018**, *203*, 746–756. https://doi.org/10.1016/j.jclepro.2018.08.255.
9. Sushil, S.; Batra, V.S. Analysis of fly ash heavy metal content and disposal in three thermal power plants in India. *Fuel* **2006**, *85*, 2676–2679. https://doi.org/10.1016/j.fuel.2006.04.031.
10. Turner, L.K.; Collins, F.G. Carbon dioxide equivalent (CO_2-e) emissions: A comparison between geopolymer and OPC cement concrete. *Constr. Build. Mater.* **2013**, *43*, 125–130. https://doi.org/10.1016/j.conbuildmat.2013.01.023.
11. Zhang, P.; Gao, Z.; Wang, J.; et al. Properties of fresh and hardened fly ash/slag based geopolymer concrete: A review. *J. Clean. Prod.* **2020**, *270*, 122389. https://doi.org/10.1016/j.jclepro.2020.122389.
12. Amran, M.; Debbarma, S.; Ozbakkaloglu, T. Fly ash-based eco-friendly geopolymer concrete: A critical review of the long-term durability properties. *Constr. Build. Mater.* **2021**, *270*, 121857. https://doi.org/10.1016/j.conbuildmat.2020.121857.
13. Shahmansouri, A.A.; Akbarzadeh Bengar, H.; Ghanbari, S. Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *J. Build. Eng.* **2020**, *31*, 101326. https://doi.org/10.1016/j.jobe.2020.101326.
14. Cong, X.; Zhou, W.; Elchalakani, M. Experimental study on the engineering properties of alkali-activated GGBFS/FA concrete and constitutive models for performance prediction. *Constr. Build. Mater.* **2020**, *240*, 117977. https://doi.org/10.1016/j.conbuildmat.2019.117977.
15. Le, H.-B.; Bui, Q.-B.; Tang, L. Geopolymer Recycled Aggregate Concrete: From Experiments to Empirical Models. *Materials* **2021**, *14*, 1180. https://doi.org/10.3390/ma14051180.

16. Zhang, H.; Li, L.; Yuan, C.; et al. Deterioration of ambient-cured and heat-cured fly ash geopolymer concrete by high temperature exposure and prediction of its residual compressive strength. *Constr. Build. Mater.* **2020**, *262*, 120924. https://doi.org/10.1016/j.conbuildmat.2020.120924.
17. Peng, Y.; Unluer, C. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Constr. Build. Mater.* **2022**, *316*, 125785. https://doi.org/10.1016/j.conbuildmat.2021.125785.
18. Getahun, M.A.; Shitote, S.M.; Abiero Gariy, Z.C. Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes. *Constr. Build. Mater.* **2018**, *190*, 517–525. https://doi.org/10.1016/j.conbuildmat.2018.09.097.
19. Naderpour, H.; Rafiean, A.H.; Fakharian, P. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *J. Build. Eng.* **2018**, *16*, 213–219. https://doi.org/10.1016/j.jobe.2018.01.007.
20. Nguyen, T.; Kashani, A.; Ngo, T.; et al. Deep neural network with high-order neuron for the prediction of foamed concrete strength. *Comput. -Aided Civ. Infrastruct. Eng.* **2019**, *34*, 316–332. https://doi.org/10.1111/mice.12422.
21. Cao, R.; Fang, Z.; Jin, M.; et al. Application of Machine Learning Approaches to Predict the Strength Property of Geopolymer Concrete. *Materials* **2022**, *15*, 2400. https://doi.org/10.3390/ma15072400.
22. Ghosh, A.; Ransinchung, G.D. Application of machine learning algorithm to assess the efficacy of varying industrial wastes and curing methods on strength development of geopolymer concrete. *Constr. Build. Mater.* **2022**, *341*, 127828. https://doi.org/10.1016/j.conbuildmat.2022.127828.
23. Amin, M.N.; Khan, K.; Javed, M.F.; et al. Prediction of Mechanical Properties of Fly-Ash/Slag-Based Geopolymer Concrete Using Ensemble and Non-Ensemble Machine-Learning Techniques. *Materials* **2022**, *15*, 3478. https://doi.org/10.3390/ma15103478.
24. Amin, M.N.; Khan, K.; Ahmad, W.; et al. Compressive Strength Estimation of Geopolymer Composites through Novel Computational Approaches. *Polymers* **2022**, *14*, 2128. https://doi.org/10.3390/polym14102128.
25. Eftekhar Afzali, S.A.; Shayanfar, M.A.; Ghanooni-Bagha, M.; et al. The use of machine learning techniques to investigate the properties of metakaolin-based geopolymer concrete. *J. Clean. Prod.* **2024**, *446*, 141305. https://doi.org/10.1016/j.jclepro.2024.141305.
26. Lyngdoh, G.A.; Zaki, M.; Krishnan, N.M.A.; et al. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning. *Cem. Concr. Compos.* **2022**, *128*, 104414. https://doi.org/10.1016/j.cemconcomp.2022.104414.
27. Dai, F.; Gai, W.; Yang, S.; et al. Machine learning-driven multi-objective optimization and dynamic performance prediction of geopolymer concrete. *J. Clean. Prod.* **2026**, *540*, 147501. https://doi.org/10.1016/j.jclepro.2026.147501.
28. Li, Y.; Lu, C.; Deng, W.; et al. Predicting compressive strength of red mud-based geopolymers via machine learning: Insights from experimental validation. *J. Build. Eng.* **2026**, *118*, 115126. https://doi.org/10.1016/j.jobe.2025.115126.
29. Ekanayake, I.U.; Meddage, D.P.P.; Rathnayake, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Stud. Constr. Mater.* **2022**, *16*, e01059. https://doi.org/10.1016/j.cscm.2022.e01059.
30. Roscher, R.; Bohn, B.; Duarte, M.F.; et al. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* **2020**, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199.
31. Lundberg, S.M.; Nair, B.; Vavilala, M.S.; et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2*, 749–760. https://doi.org/10.1038/s41551-018-0304-0.
32. Yeung, C.; Tsai, J.-M.; King, B.; et al. Elucidating the Behavior of Nanophotonic Structures through Explainable Machine Learning Algorithms. *ACS Photonics* **2020**, *7*, 2309–2318. https://doi.org/10.1021/acsphotonics.0c01067.
33. Tang, B.; Lu, Y.; Zhou, J.; et al. Machine learning-guided synthesis of advanced inorganic materials. *Mater. Today* **2020**, *41*, 72–80. https://doi.org/10.1016/j.mattod.2020.06.010.
34. Johnsen, P.V.; Riemer-Sørensen, S.; DeWan, A.T.; et al. A new method for exploring gene–gene and gene–environment interactions in GWAS with tree ensemble methods and SHAP values. *BMC Bioinform.* **2021**, *22*, 230. https://doi.org/10.1186/s12859-021-04041-7.
35. Kamarthi, A.; Kaliyamoorthy, B. An interpretable automated optimized machine learning for predicting concrete compressive strength. *Expert Syst. Appl.* **2026**, *298*, 129656. https://doi.org/10.1016/j.eswa.2025.129656.
36. Golafshani, E.; Khodadadi, N.; Ngo, T.; et al. Modelling the compressive strength of geopolymer recycled aggregate concrete using ensemble machine learning. *Adv. Eng. Softw.* **2024**, *191*, 103611. https://doi.org/10.1016/j.advengsoft.2024.103611.
37. Zhou, H.; Deng, Z.; Xia, Y.; et al. A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing* **2016**, *216*, 208–215. https://doi.org/10.1016/j.neucom.2016.07.036.
38. Kakasor Ismael Jaf, D.; Ismael Abdulrahman, P.; Salih Mohammed, A.; et al. Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO_2) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete. *Constr. Build. Mater.* **2023**, *400*, 132604. https://doi.org/10.1016/j.conbuildmat.2023.132604.
39. Sichani, M.N.; Dehkordi, O.M.; Khorshidi, M.; et al. Machine learning based optimization of fly ash content for improving geopolymer concrete compressive strength. *Sci. Rep.* **2025**, *15*, 44785. https://doi.org/10.1038/s41598-025-29088-x.

40. Drucker, H.; Burges, C.J.C.; Kaufman, L.; et al. Support vector regression machines. In Proceedings of the Proceedings of the 10th International Conference on Neural Information Processing Systems, Denver, Colorado, 1996; pp. 155–161.
41. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. https://doi.org/10.1007/BF00994018.
42. Chen, B.-T.; Chang, T.-P.; Shih, J.-Y.; et al. Estimation of exposed temperature for fire-damaged concrete using support vector machine. *Comput. Mater. Sci.* **2009**, *44*, 913–920. https://doi.org/10.1016/j.commatsci.2008.06.017.
43. Entezami, A.; Sarmadi, H. Machine learning-aided prediction of windstorm-induced vibration responses of long-span suspension bridges. *Comput. Aided Civ. Infrastruct. Eng.* **2024**, *40*, 1043–1060. https://doi.org/10.1111/mice.13387.
44. Zhou, Y.; Zhang, Y.; Pang, R.; et al. Seismic fragility analysis of high concrete faced rockfill dams based on plastic failure with support vector machine. *Soil Dyn. Earthq. Eng.* **2021**, *144*, 106587. https://doi.org/10.1016/j.soildyn.2021.106587.
45. Ghorbani, B.; Arulrajah, A.; Narsilio, G.; et al. Experimental investigation and modelling the deformation properties of demolition wastes subjected to freeze–thaw cycles using ANN and SVR. *Constr. Build. Mater.* **2020**, *258*, 119688. https://doi.org/10.1016/j.conbuildmat.2020.119688.
46. Zhang, W.; Khan, A.; Huyan, J.; et al. Predicting Marshall parameters of flexible pavement using support vector machine and genetic programming. *Constr. Build. Mater.* **2021**, *306*, 124924. https://doi.org/10.1016/j.conbuildmat.2021.124924.
47. Wu, D.; Jennings, C.; Terpenny, J.; et al. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J. Manuf. Sci. Eng.* **2017**, *139*, 071018. https://doi.org/10.1115/1.4036350.
48. Erdal, H.I. Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1689–1697. https://doi.org/10.1016/j.engappai.2013.03.014.
49. Song, H.; Ahmad, A.; Farooq, F.; et al. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Constr. Build. Mater.* **2021**, *308*, 125021. https://doi.org/10.1016/j.conbuildmat.2021.125021.
50. Jerome, H.F. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. https://doi.org/10.1214/aos/1013203451.
51. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
52. Zhang, Z.; Huang, J.; Su, Q.; et al. High embankment slope stability prediction using data augmentation and explainable ensemble learning. *Comput. -Aided Civ. Infrastruct. Eng.* **2025**, *40*, 2833–2858. https://doi.org/10.1111/mice.13478.
53. Hao, M.; Li, Y.; Chen, X.; et al. Optimization design method of C30/C40 fly ash concrete based on machine learning and elite retention genetic algorithm. *Adv. Eng. Softw.* **2025**, *210*, 104019. https://doi.org/10.1016/j.advengsoft.2025.104019.
54. Nguyen, H.; Vu, T.; Vo, T.P.; et al. Efficient machine learning models for prediction of concrete strengths. *Constr. Build. Mater.* **2021**, *266*, 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950.
55. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017; pp. 4768–4777.
56. Roth, A.E. Introduction to the Shapley value. In *The Shapley Value: Essays in Honor of Lloyd S. Shapley*; Roth, A.E., Ed.; Cambridge University Press: Cambridge, 1988; pp. 1–28.
57. Shafighfard, T.; Kazemi, F.; Bagherzadeh, F.; et al. Chained machine learning model for predicting load capacity and ductility of steel fiber–reinforced concrete beams. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 3573–3594. https://doi.org/10.1111/mice.13164.
58. Zhang, J.; Li, D.; Wang, Y. Toward intelligent construction: Prediction of mechanical properties of manufactured-sand concrete using tree-based models. *J. Clean. Prod.* **2020**, *258*, 120665. https://doi.org/10.1016/j.jclepro.2020.120665.
59. Lundberg, S.M.; Erion, G.; Chen, H.; et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9.
60. Lundberg, S.; Erion, G.; Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1802.03888.
61. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Volume 2, pp. 1137–1143.
62. Farooq, F.; Ahmed, W.; Akbar, A.; et al. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *J. Clean. Prod.* **2021**, *292*, 126032. https://doi.org/10.1016/j.jclepro.2021.126032.
63. Cao, C.; Sun, W.; Qin, H. The analysis on strength and fly ash effect of roller-compacted concrete with high volume fly ash. *Cem. Concr. Res.* **2000**, *30*, 71–75. https://doi.org/10.1016/S0008-8846(99)00203-3.
64. Lee, B.; Kim, G.; Kim, R.; et al. Strength development properties of geopolymer paste and mortar with respect to amorphous Si/Al ratio of fly ash. *Constr. Build. Mater.* **2017**, *151*, 512–519. https://doi.org/10.1016/j.conbuildmat.2017.06.078.
65. Cho, Y.K.; Jung, S.H.; Choi, Y.C. Effects of chemical composition of fly ash on compressive strength of fly ash cement mortar. *Constr. Build. Mater.* **2019**, *204*, 255–264. https://doi.org/10.1016/j.conbuildmat.2019.01.208.

66. Noushini, A.; Castel, A.; Aldred, J.; et al. Chloride diffusion resistance and chloride binding capacity of fly ash-based geopolymer concrete. *Cem. Concr. Compos.* **2020**, *105*, 103290. https://doi.org/10.1016/j.cemconcomp.2019.04.006.
67. Palomo, A.; Blanco-Varela, M.T.; Granizo, M.L.; et al. Chemical stability of cementitious materials based on metakaolin. *Cem. Concr. Res.* **1999**, *29*, 997–1004. https://doi.org/10.1016/S0008-8846(99)00074-5.
68. Wang, J.; Xie, J.; Wang, C.; et al. Study on the optimum initial curing condition for fly ash and GGBS based geopolymer recycled aggregate concrete. *Constr. Build. Mater.* **2020**, *247*, 118540. https://doi.org/10.1016/j.conbuildmat.2020.118540.
69. Sajan, P.; Jiang, T.; Lau, C.; et al. Combined effect of curing temperature, curing period and alkaline concentration on the mechanical properties of fly ash-based geopolymer. *Clean. Mater.* **2021**, *1*, 100002. https://doi.org/10.1016/j.clema.2021.100002.
70. Panias, D.; Giannopoulou, I.P.; Perraki, T. Effect of synthesis parameters on the mechanical properties of fly ash-based geopolymers. *Colloids Surf. A Physicochem. Eng. Asp.* **2007**, *301*, 246–254. https://doi.org/10.1016/j.colsurfa.2006.12.064.
71. Görhan, G.; Kürklü, G. The influence of the NaOH solution on the properties of the fly ash-based geopolymer mortar cured at different temperatures. *Compos. Part B Eng.* **2014**, *58*, 371–377. https://doi.org/10.1016/j.compositesb.2013.10.082.
72. Ruiz-Santaquiteria, C.; Skibsted, J.; Fernández-Jiménez, A.; et al. Alkaline solution/binder ratio as a determining factor in the alkaline activation of aluminosilicates. *Cem. Concr. Res.* **2012**, *42*, 1242–1251. https://doi.org/10.1016/j.cemconres.2012.05.019.
73. Xie, T.; Visintin, P.; Zhao, X.; et al. Mix design and mechanical properties of geopolymer and alkali activated concrete: Review of the state-of-the-art and the development of a new unified approach. *Constr. Build. Mater.* **2020**, *256*, 119380. https://doi.org/10.1016/j.conbuildmat.2020.119380.
74. Hadi, M.N.S.; Al-Azzawi, M.; Yu, T. Effects of fly ash characteristics and alkaline activator components on compressive strength of fly ash-based geopolymer mortar. *Constr. Build. Mater.* **2018**, *175*, 41–54. https://doi.org/10.1016/j.conbuildmat.2018.04.092.
75. Wang, Y.; Liu, X.; Zhang, W.; et al. Effects of Si/Al ratio on the efflorescence and properties of fly ash based geopolymer. *J. Clean. Prod.* **2020**, *244*, 118852. https://doi.org/10.1016/j.jclepro.2019.118852.
76. Li, J.; Tao, Y.; Zhuang, E.; et al. Optimal amorphous oxide ratios and multifactor models for binary geopolymers from metakaolin blended with substantial sugarcane bagasse ash. *J. Clean. Prod.* **2022**, *377*, 134215. https://doi.org/10.1016/j.jclepro.2022.134215.
77. Liu, J.; Doh, J.-H.; Dinh, H.L.; et al. Effect of Si/Al molar ratio on the strength behavior of geopolymer derived from various industrial waste: A current state of the art review. *Constr. Build. Mater.* **2022**, *329*, 127134. https://doi.org/10.1016/j.conbuildmat.2022.127134.
78. Nematollahi, B.; Sanjayan, J. Effect of different superplasticizers and activator combinations on workability and strength of fly ash based geopolymer. *Mater. Des.* **2014**, *57*, 667–672. https://doi.org/10.1016/j.matdes.2014.01.064.
79. Carabba, L.; Manzi, S.; Bignozzi, M.C. Superplasticizer Addition to Carbon Fly Ash Geopolymers Activated at Room Temperature. *Materials* **2016**, *9*, 586. https://doi.org/10.3390/ma9070586.
80. Puertas, F.; Palomo, A.; Fernández-Jiménez, A.; et al. Effect of superplasticisers on the behaviour and properties of alkaline cements. *Adv. Cem. Res.* **2003**, *15*, 23–28. https://doi.org/10.1680/adcr.2003.15.1.23.
81. Liang, W.; Yin, W.; Zhong, Y.; et al. Mixed artificial intelligence models for compressive strength prediction and analysis of fly ash concrete. *Adv. Eng. Softw.* **2023**, *185*, 103532. https://doi.org/10.1016/j.advengsoft.2023.103532.
82. Tavana Amlashi, A.; Emadaldin, M.G.; Abolfazl, E.S.; et al. Estimation of the compressive strength of green concretes containing rice husk ash: A comparison of different machine learning approaches. *Eur. J. Environ. Civ. Eng.* **2023**, *27*, 961–983. https://doi.org/10.1080/19648189.2022.2068657.
83. Mermerdaş, K.; Manguri, S.; Nassani, D.E.; et al. Effect of aggregate properties on the mechanical and absorption characteristics of geopolymer mortar. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 1642–1652. https://doi.org/10.1016/j.jestch.2017.11.009.
84. Tenn, N.; Allou, F.; Petit, C.; et al. Formulation of new materials based on geopolymer binders and different road aggregates. *Ceram. Int.* **2015**, *41*, 5812–5820. https://doi.org/10.1016/j.ceramint.2015.01.010.
85. Santhanam, K.; Reshma, E.K.; George, C.; et al. Exploring the potential of alkali-activated cassava peel ash for early-age geopolymer concrete strength development. *Constr. Build. Mater.* **2025**, *498*, 144062. https://doi.org/10.1016/j.conbuildmat.2025.144062.