

Research Article

Qingqing Tian, Hang Gao, Lei Guo*, Zexuan Li, and Qiongyao Wang

CSG compressive strength prediction based on LSTM and interpretable machine learning

<https://doi.org/10.1515/rams-2023-0133>

received April 24, 2023; accepted October 09, 2023

Abstract: As a new type of environmentally friendly building material, cemented sand and gravel (CSG) has advantages distinct from those of concrete. Compressive strength is an important mechanical property of CSG. However, his method of testing is mainly by doing experiments. For this reason, a deep learning algorithm, long short-term memory (LSTM) model, was proposed to predict the compressive strength of CSG by using four input variables, namely cement content, sand rate, water-binder ratio, and fly ash content, with a total of 114 sample data. Three metrics – coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) – were used to evaluate the model's performance, and the predicted results were compared with the traditional machine learning algorithm, namely the random forest (RF) model. Finally, SHapley Additive exPlanations can be combined to explain the contribution degree of each input feature in the machine learning inquiry model to the prediction results. The results show that the prediction accuracy and reliability of LSTM are higher. The LSTM model has $R^2 = 0.9940$, RMSE = 0.1248, and MAE = 0.0960, while the RF model has $R^2 = 0.9147$, RMSE = 0.4809, and MAE = 0.4397. The LSTM model can accurately predict CSG compressive strength. Cement and sand rate contribute more to the predicted results than other input characteristics.

Keywords: long short-term memory, random forest, cemented sand and gravel, compressive strength, deep learning, machine learning

1 Introduction

In response to the world development trend and the national call for “carbon neutrality,” the construction of environmentally friendly water conservancy projects has become the mainstream trend [1]. Compared with a roller-compacted concrete dam, cement consumption is less, aggregate preparation and mixing facilities are greatly simplified, temperature control measures can be cancelled, the construction speed is significantly accelerated, and the project cost is significantly reduced. Its material, cemented sand and gravel (CSG), is an economical, safe, green, and low-carbon new building material formed by river bed gravel or local waste materials after mixing, rolling, and vibrating with cementing materials and water [2,3]. As an ultra-poor cementing material, it is similar to concrete but has many differences and advantages. Compared with concrete, the reduction in cement content significantly reduces its hydration heat. The selection of raw materials is very simple, and the aggregate with large particle sizes removed does not need to be sieved. To avoid the destruction of land vegetation to the greatest extent [4], it is necessary to study the properties of CSG to apply and popularize it more efficiently.

For the various properties of CSG, compressive strength is directly related to the safety of the structure and is a necessary condition for evaluating the performance of the structure in the whole life cycle [5]. Therefore, compressive strength is one of the most important properties of CSG. However, CSG is a heterogeneous mixture of complex materials, and each component is randomly distributed in the CSG mix ratio, and factors such as cement content, water-binder ratio, and waste composition will affect the compressive strength [6]. Therefore, it is very difficult to accurately predict the compressive strength of CSG with a complex matrix. Scholars have conducted relevant studies on the compressive strength performance of CSG, including test methods, and relevant scholars have taken some tests to assess the compressive strength of CSG, statistical rule analysis, and influence analysis. Chen *et al.* [7] established a dataset of CSG mix ratio and 28-day compressive strength and analyzed the distribution law of CSG compressive strength

* **Corresponding author: Lei Guo**, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; Henan Water Conservancy Investment Group Co., Ltd, Zhengzhou 450002, China; Henan Key Laboratory of Water Environment Simulation and Treatment, Zhengzhou 450002, China, e-mail: guoleiboss@126.com

Qingqing Tian: North China University of Water Resources and Electric Power, Zhengzhou 450046, China; China Institute of Water Resources and Hydropower Research, Beijing 100044, China

Hang Gao, Zexuan Li, Qiongyao Wang: North China University of Water Resources and Electric Power, Zhengzhou 450046, China

by using skewness kurtosis and single-sample Kolmogorov–Smirnov test. Li *et al.* [8] conducted an experimental study on the effects of sand rate, water-binder ratio, fly ash, and other parameters on the properties of CSG materials in different mix ratios. Chai *et al.* [9] studied the influence of fly ash content on the compressive strength of CSG. However, taking the test method generally requires a lot of time and economic cost, so it is necessary to seek a low-cost and high-prediction accuracy method to predict its performance.

With the development of artificial intelligence, intelligent algorithms have been applied in many fields. Zhou *et al.* [10] proposed a fire prediction model based on the CatBoost algorithm to predict fire points. Liu *et al.* [11] proposed a prediction model based on the XGBoost algorithm for pipeline safety assessment. Ilić *et al.* [12] realized water-quality prediction in five different regions through the Naïve Bayes algorithm. Wang *et al.* [13] developed an improved back propagation (BP) neural network to predict surface runoff coefficients in different rainfall conditions. Intelligent algorithms are also used to predict the strength of civil engineering materials, mainly concrete. For example, Wu and Zhou [14,15] employed an optimized support vector regression (SVR) model to predict the splitting tensile and compressive strength of sustainable concrete, and the results showed that the optimized model can achieve an accurate prediction of concrete mechanical properties. Latif [16] used boosted decision tree regression (BDTR) to predict the compressive strength of concrete and compared it with support vector machine (SVM). The results show that BDTR has better prediction accuracy with an R^2 of 0.86, which can accurately predict the compressive strength of concrete, but this may depend on the input adequacy of the data set. Ahmad *et al.* [17] adopted decision tree, bagging regressor, and AdaBoost regressor to predict geopolymer concrete compressive strength, and the results showed that the bagging model had the best prediction accuracy. But it is also possible to compare the accuracy of predictions with other machine learning models. Yuan *et al.* [18] used machine learning methods such as gradient boosting and random forest (RF) to predict the compressive strength and flexural strength of recycled aggregate concrete. The results show that RF has better prediction accuracy than gradient boosting, and it is suggested that environmental characteristics can be further added as input variables. Mozumder *et al.* [19] tried using SVR to predict the uniaxial compressive strength of FRP-confined concrete. The results show that SVR can be used as an alternative physical tool to predict the strength of fiber reinforced polymer (FRP)-confined concrete. However, the research on the properties of materials is mainly focused on concrete, but there is a lack of research on the properties of CSG, especially when it comes to the application of intelligent algorithms in strength.

In addition, most of the above intelligent algorithms are traditional machine learning algorithms, whose prediction ability is limited. Compared with traditional machine learning algorithms, it may be a better choice to explore deep learning models with better prediction performance. For example, Liu *et al.* [20] proposed a convolutional neural network (CNN)-based foreign exchange rate prediction model, and the results show that its long-term prediction accuracy is better than artificial neural networks, SVR, and other models. Salinas *et al.* [21] adopted an autoregressive recurrent neural network (RNN) (DeepAR) to produce probabilistic predictions and showed an accuracy improvement of about 15% compared to the latest methods. Wu *et al.* [22] combined a deep learning gate recurrent unit network with wavelet packet decomposition for the automatic diagnosis of internal defect signals of concrete structures, and the results showed that the accuracy rate of the model reached 90.76% and the prediction performance was good. Zhang and Ci [23] used the deep belief network (DBN), and their results suggest that DBN has excellent performance in forecasting and direction. However, selecting too many hidden layers in pursuit of prediction accuracy may cause the problem of gradient disappearance [24]. One of the characteristics of long short-term memory (LSTM) model is that it can learn long-term dependencies to avoid the problem of disappearing gradients [25]. Qiu *et al.* [26] predicted river water temperature by LSTM and achieved a good prediction effect. Latif [27] predicted the compressive strength of concrete through LSTM and proved the superiority of the LSTM model.

In view of the multiple advantages of CSG over concrete, the rare application of intelligent algorithms in CSG material properties, the advantages of deep learning, and the advantages and disadvantages of different algorithms, this article uses 114 sets of compressive strength test data to predict the compressive strength of CSG for the first time through deep learning LSTM. Compared with the traditional machine learning RF, the effectiveness of the LSTM model and the advantages of deep learning compared with traditional machine learning are verified, which provides a theoretical basis for the practical application of CSG and promotes the practical application of CSG.

2 Experimental design and method

2.1 Experimental raw materials and mix ratio

The purpose of this experiment was to measure the compressive strength of CSG under different mix ratios. According to

Table 1: Mix ratio of CSG

Experimental variables	Variable level			
Cement ($\text{kg}\cdot\text{m}^{-3}$)	40	50	60	70
Sand rate (%)	10	20	30	40
Water-binder ratio	1.0	1.2	1.4	—
Fly ash ($\text{kg}\cdot\text{m}^{-3}$)	20	30	40	50

the “Technical Guidelines for Cemented Granular Material damming” (SL678-2014), the amount of cementing material should not have been less than $80 \text{ kg}\cdot\text{m}^{-3}$, and the amount of cement should not have been less than $32 \text{ kg}\cdot\text{m}^{-3}$. The sum of cement content and fly ash content in this test was between 80 and $110 \text{ kg}\cdot\text{m}^{-3}$, and the mix ratio is shown in Table 1.

The cement used in the experiment was 425# ordinary Portland cement, produced in Henan Yodongda Cement Co., Ltd, with physical and mechanical properties as shown in Table 2; the fly ash was Class FII dry discharge fly ash from Zhengzhou Thermal Power Plant, with properties as shown in Table 3. The mixing water was tap water, and the coarse aggregate and fine aggregate came from the North Ruhe material Yard in Ruzhou City.

2.2 Experimental process

2.2.1 Preparation methods

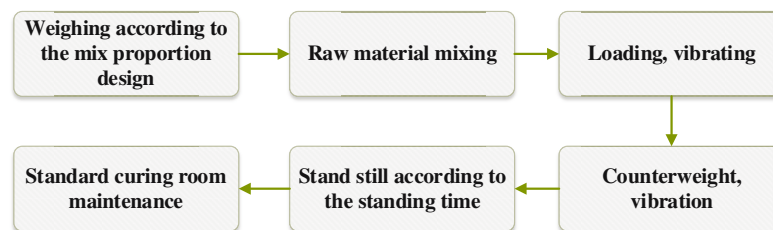
In view of the fact that CSG material properties are between roller-compacted concrete and earth-stone materials, the forming and maintenance of CSG specimens in this test were carried out according to the “Technical Guidelines for Dam Construction with Cemented Particle Material” (SL678-2014) and “Test Procedure for Hydraulic Rolled Concrete” (DL/T5433-2009). The preparation process of the CSG specimen in this test is shown in Figure 1. First, the aggregate was screened and then stored in silos with different particle size standards. The mixture was designed according to the mix ratio, and the mixture was mixed at the end. In order to improve the mixing uniformity, mechanical mixing and manual mixing were combined, and the mixing machine used was the single horizontal shaft concrete SJD-60 mixer. After the mixing, loading, vibrating, and forming were carried out. During the loading, the cast iron test mold pre-painted with oil was used. First, manual vibration was carried out for no less than 25 times and then moved to the shaking table. The weight block was put on, and the hand was righted. The vibration time was strictly controlled to

Table 2: Physical and mechanical properties of cement

Project	Specific surface area ($\text{m}^2\cdot\text{kg}^{-1}$)	Density ($\text{kg}\cdot\text{m}^{-3}$)	Setting time (min)		Compressive strength (MPa)		Flexural strength (MPa)	
			Initial set	Final set	3 day	28 day	3 day	28 day
Result	382.00	3035.00	143.00	191.00	30.40	49.60	6.30	8.60

Table 3: Performance of fly ash

Ignition loss (%)	Specific surface area ($\text{m}^2\cdot\text{kg}^{-1}$)	Density ($\text{g}\cdot\text{cm}^{-3}$)	Chemical composition (%)				
			Fe_2O_3	CaO	MgO	Al_2O_3	SiO_2
2.48	420	2.42	3.87	2.27	0.81	29.09	53.36

**Figure 1:** Preparation process of a CSG specimen.

vibrate the shaking table. After the vibration was completed, the specimen was moved down, i.e., the specimen was formed. After the specimen was covered and left for 48 h, the mold was removed, and finally the specimen was put into the standard curing room for curing until the test age.

2.2.2 Experimental methods and results

The compressive strength test was carried out according to the “Standard for Test Methods of Physical and Mechanical Properties of Concrete” (GB/T 50081-2019). The pressure testing machine adopted WAW-1000 electro-hydraulic servo universal testing machine, as shown in Figure 2. The compressive strength test of the CSG specimen was carried out by using the computer to automatically control the test process. The specimen was placed in the middle of the pressure plate under the testing machine after inspection, and the bearing surface was perpendicular to the top surface when forming. Under continuous and uniform loading, the specimen approached failure and began to rapidly deform until it was completely destroyed, and the failure load was recorded. The CSG specimen is a secondary standard cube specimen of $150 \times 150 \times 150$ mm. There are 114 groups of tests, with three specimens in each group of tests. The compressive strength was tested according to the above method standards, and the representative value of CSG compressive strength was determined according to the “Concrete Strength Inspection and Evaluation Standard” (GB50107). The 114 sets of data obtained after the final selection are shown in Figure 3.

3 Methodology

3.1 LSTM networks

LSTM networks are a special type of RNN. It was proposed by Hochreiter and Schmidhuber in 1997 [28]. RNN is a loop



Figure 2: Compressive strength test of a 150 mm cube CSG specimen.

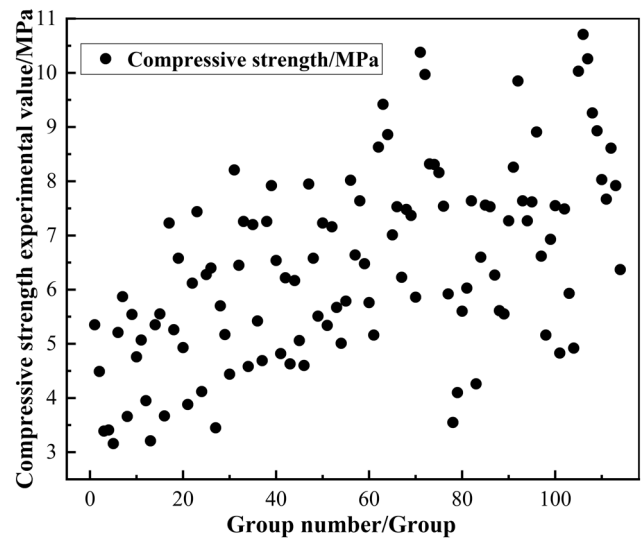


Figure 3: Compressive strength experimental value of CSG.

network in which information is transmitted from the current loop to the next loop. This chain structure indicates that RNN is a normal neural network structure that can be used. However, regular RNN has the problem of long-term dependencies, which means that as the distance between loops increases, the link of information in the RNN can break. However, LSTM can solve the problems of gradient disappearance and gradient explosion during long sequence training in the machine learning field. The LSTM is able to learn long-term dependencies due to the presence of special properties in the model repetition module.

LSTM uses storage units and gates to control long-term information stored or retained in the network. As a powerful recursive neural network model, LSTM can extract long- and short-term correlations in time series, enabling the model to extract data features effectively [29–31]. LSTM includes a forget gate, an input gate, an update gate, and an output gate in the main structure. The main equations of the LSTM structure are as follows:

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f), \quad (1)$$

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i), \quad (2)$$

$$g_t = \tanh(W_g(h_{t-1}, x_t) + b_g), \quad (3)$$

$$c_t = f_t c_{t-1} + i_t g_t, \quad (4)$$

$$o_t = \sigma(W_o(h_{t-1}, x_t) + b_o), \quad (5)$$

$$h_t = o_t \tanh(c_t), \quad (6)$$

where f_t , i_t , g_t , and o_t are the output values of forgetting gate, input gate, update gate, and output gate, respectively; W_f , W_i , W_g , and W_o are weight vectors; b_f , b_i , b_g , and b_o are

bias vectors; c_t and σ are memory unit and sigmoid activation functions, respectively.

3.2 RF

The RF algorithm was proposed by Breiman in 2001 [32]. The core idea of RF is ensemble learning, which consists of multiple decision classification trees, each of which is built from a bootstrap sample of application data. In the process of tree construction, variables are randomly selected as the candidate variable set at each split, and the results are collected by randomly selecting the features of each classification tree. Finally, the results are stably and accurately predicted by majority voting or average according to each specific problem. Assuming a set of input data sets is $\{H(x, \theta_i), i = 1, 2, \dots, k\}$ and the prediction value of a single decision tree is $\{H(x, \theta_i)\}$, the final prediction result of the RF model is the average of the prediction results of all decision trees:

$$\bar{H}(x) = \frac{1}{k} \sum_{i=1}^k H(x, \theta_i), \quad (7)$$

where $\bar{H}(x)$ is the predicted value of the RF model, θ_i is a random variable of a single decision tree, x is the characteristic variable, and k is the number of decision trees.

3.3 Interpretable machine learning method

Interpretability shortage is one of the restrictions of using a machine learning model, and SHapley Additive exPlanations (SHAP) belongs to the method of model post-interpretation. Its core idea is to calculate the marginal contribution of features to the model output and then explain the “black box model” from the global and local levels [33]. SHAP is an interpretable machine learning method based on game theory that involves constructing a combination of different input variables to compare the average change of the model output and then quantifying the specific contribution of each feature to the model results. The SHAP value of each input variable is the weighted average of the marginal contribution of the variable, which is calculated as follows:

$$\Phi_i = \sum_{S \subseteq N(i)} \frac{|S|!(n - |S| - 1)!}{n!} [F(S \cup \{i\}) - F(S)], \quad (8)$$

where Φ_i is the SHAP value of input variable i . A positive (negative) SHAP value indicates that variable i contributes to the prediction result; n is the number of input variables; N is the complete set of input variables; S is the set

excluding variable i and is a subset of N ; and $F(S)$ is the prediction based on the input S .

4 Database establishment

4.1 Sample data

The sample data used in this study came from the experimental data of CSG mechanical properties obtained in the previous experiment, with a total of 114 sample data. As an additional data set independent of the training set and the test set, the validation set can be used to evaluate the model to better determine whether the model has good generalization ability. In view of the necessity of the verification set and to further increase the reliability of the model, the sample data were randomly divided into three parts: 77 training set data, 20 verification set data, and 17 test set data, and the modeling was carried out on the basis of these data. Input variables in this article were completely consistent with the test mix ratio and material dosage. Input variables included cement content, sand rate, water-binder ratio, and fly ash content, which involved all raw materials (cement, fly ash, sand, water, and sand gravel) in the test. Although increasing the number of input features in some machine learning algorithms could improve the model's performance to a certain extent, it usually requires more data support [34]. In view of the limited amount of data in this article, it was appropriate to select four input variables, and the model in this article could achieve very good prediction accuracy, whereas too many input features might have easily led to overfitting problems. At the same time, the computational complexity of the model will be increased [35]. In addition, the output variable was set to compressive strength. The description of the overall data in the model is shown in Table 4, which provides the mean value, median value, standard deviation, sample variance, range, minimum value, maximum value, sum, and number of sample data of the corresponding data. In order to ensure that all models obtain the best results from the parameters, it was crucial to identify these parameters.

4.2 Data preprocessing and evaluation index

Due to the differences in the units of input variables in this study, if they were directly substituted into the model for learning, the final result would be affected to some extent, so it is necessary to normalize the data. The normalization process retains all the features, converts the parameter

Table 4: Statistical description of the overall data

Category	Cement (kg·m ⁻³)	Sand rate	Water- binder ratio	Fly ash (kg·m ⁻³)
Mean	57.54	0.28	1.19	36.40
Median	60	0.3	1.2	40
Standard deviation	9.78	0.097	0.16	10.44
Sample variance	95.72	0.0096	0.027	109.00
Range	30	0.30	0.40	30
Minimum	40	0.10	1	20
Maximum	70	0.40	1.40	50
Sum	6,560	32.1	118.8	4,150
Count	114	114	114	114

values into data between 0 and 1, and converts dimensionless expressions into dimensionless expressions, so that the data are numerically comparable and the model has higher accuracy. In this study, min–max standardization was used as a normalization method, and the expression equation is as follows:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (9)$$

where y_i represents the normalized data, $x_{\min}$ represents the minimum feature data, $x_{\max}$ represents the maximum value of the feature, and x_i represents the data before normalization.

In order to test the performance of the model, this article evaluated its performance through coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Each index has its own method to infer the performance of the model. R^2 is used to check the linear correlation between the experimental value and the predicted value. When R^2 is >0.8 and <1 , the model is considered valid [36]. RMSE is used to evaluate the difference between the experimental value and the predicted value and is the most commonly used index to measure the regression quality of regression trees. MAE is used to assess the average error between the experimental value and the predicted value. For these two indicators, the lower the value, the higher the model performance. R^2 , RMSE, and MAE formulas are as follows:

$$R^2 = 1 - \frac{\sum_{z=1}^N (y_z - f_z)^2}{\sum_{z=1}^N (y_z - \bar{y}_z)^2}; \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - f_i)^2}{N}}; \quad (11)$$

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - f_i|}{N}, \quad (12)$$

where N is the number of sample data, y_i is the experimental value (MPa), $\bar{y}_i$ is the average of the experimental value (MPa), f_i is the predicted value of model regression (MPa), and $\bar{f}_i$ is the average of the predicted value from the model regression (MPa).

5 Model prediction and result analysis

The purpose of this study was to evaluate the performance of the LSTM model and the RF model for CSG compressive strength prediction. In order to analyze and compare the predictive performance of the two models, the experimental value and the predicted value were plotted in relation to each other. The input data set included cement content, sand rate, water-binder ratio, and fly ash content, and the output variable was compressive strength. R^2 , RMSE, and MAE were used as the performance indexes for model evaluation, and the predictive performance of the two models was compared and analyzed. The effectiveness of deep learning algorithm and machine learning algorithm in CSG compressive strength prediction was compared. The deep learning algorithm chosen in this article was LSTM, and the machine learning algorithm was RF.

Both the LSTM model and the RF model successfully predicted CSG compressive strength. Grid search is widely used as a machine learning tuning technology. It evaluates the model effect of each different parameter combination for random combinations of parameters within a given range through cross-validation to find a set of optimal parameters. This method has good applicability in both machine learning and deep learning and has strong robustness during operation. It can exhaust all possible parameter combinations to ensure that the global optimal solution can be found, especially in problems with small parameter space [37,38]. In view of the advantages of this method and the small parameter space of the proposed model, the grid search method was used to optimize the parameters of the LSTM and RF models.

For the parameter adjustment of the LSTM model, grid search was used to find the optimal parameter structure, and cross-validation was carried out under the conditions of a given range of relevant parameters. The optimal parameter structure was obtained as follows: the activation function was “relu,” the step size was 2, the number of

hidden layers was 256, the number of batch selections was 32, the learning rate was 0.001, the output scale was 1, and the output scale corresponded to an output variable (compressive strength). In addition, the epoch number of iterations was 250, which determined the fitting and convergence of model training to a large extent. The model required several iterations to fit convergence. As the number of epochs increased, the number of weight update iterations increased, and the curve changed from the initial state of unfitting to the state of optimal fitting and finally into the state of overfitting. The iteration of the LSTM model in this article is shown in Figure 4. It could be seen from the figure that both the training set and the verification set began to converge when the number of iterations reached 200 and then entered the optimization

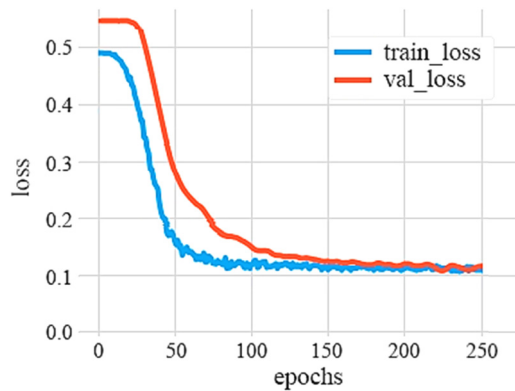


Figure 4: LSTM model iteration process.

fitting state. Therefore, the epoch number of iterations of the LSTM model in this article was 250.

For the parameter adjustment of the RF model, the tuning of the RF model mainly involves two parameters: the number of decision trees and the number of leaf nodes [39]. The number of decision trees is often referred to as “the number of trees in a RF,” and the number of leaf nodes controls the depth of each decision tree. These two parameters affect the overall performance and running speed of the RF, and a few other parameters will also have a certain impact on the model, such as “min_samples_split” and “max_feature.” Given the range of relevant parameters, the number of main parameter decision trees was 0–200 and the number of leaf nodes was 0–10. After cross-verification, the optimal parameter structure was obtained as follows: the number of decision trees was 39, the number of leaf nodes was 1, “min_samples_split” was 2, “max_feature” was “auto”, where “min_samples_split” and “max_feature” were default values, and the model would have obtained the best performance under this parameter structure.

The LSTM and RF models were divided into the same training set, validation set, and test set to facilitate comparative analysis, and training and learning were carried out under the optimal parameter structure. The results after successful prediction were analyzed. Figures 5 and 6, respectively, show the errors between the predicted value and the experimental value of the LSTM and the RF models under different data sets. By observing the bar chart, it could be seen that the absolute error of the LSTM model’s prediction results under the training set and the test set was

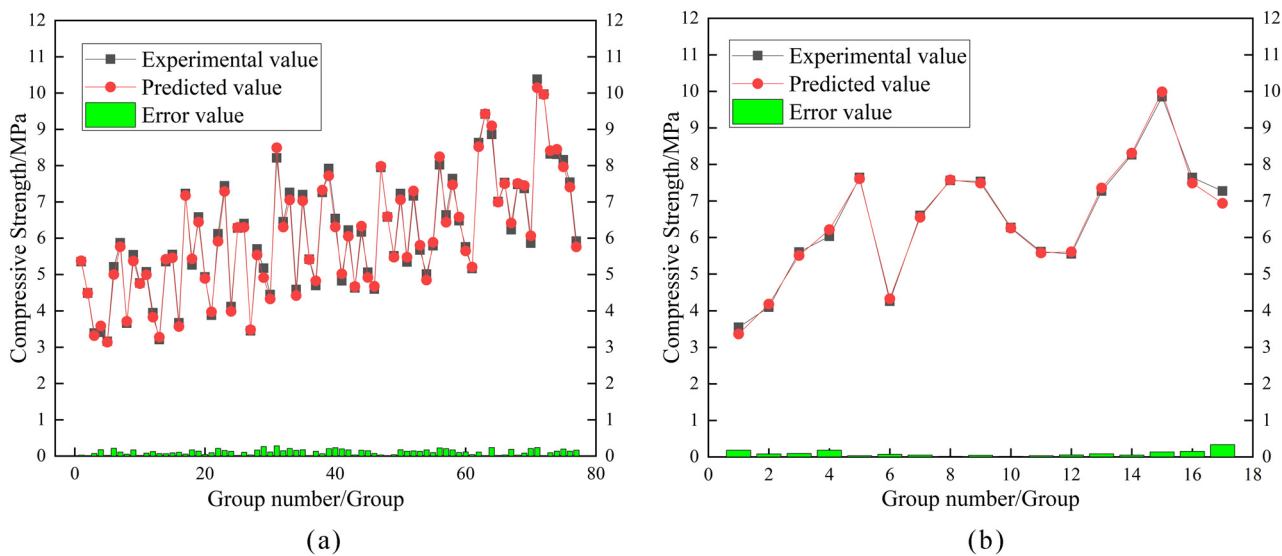


Figure 5: Error plots of predicted and experimental values of the LSTM model in different data sets: (a) train set and (b) test set.

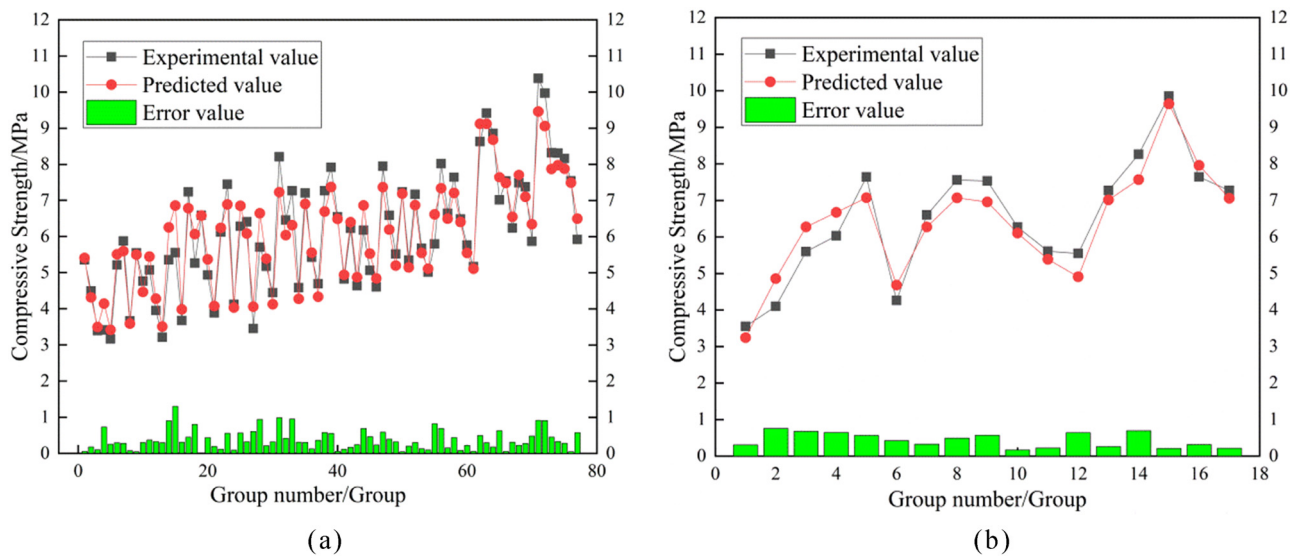


Figure 6: Error plots of predicted and experimental values of the RF model in different data sets: (a) train set and (b) test set.

mainly concentrated in $[0-0.3]$, while the absolute error of the RF model's prediction results under the training set and the test set was mainly concentrated in $[0.2-1.0]$, and the error of the LSTM model's prediction results was much smaller than that of RF model. It can be seen from the dot plot in the Figures 5 and 6 that the predicted value curves of the LSTM model in both the training set and the test set were more in line with the test value curves than those of the RF model, and the predicted value was more in line with the test value, which further reflected that the prediction accuracy of the LSTM model is much higher than that of the RF model.

In order to more intuitively represent the prediction accuracy of the model, the scatterplot of the prediction effect of each sample of the LSTM model and RF model is shown in Figure 7, and the fitting of the prediction results is also shown. Compared with the RF model, the scatter point of the LSTM model was closer to the diagonal line, indicating that the model had better fitting performance under both the training set and the test set. The predicted value of the LSTM model had better consistency with the compressive strength test value of the CSG model, and it had higher prediction accuracy than the RF model. This can also be seen in the coefficient of determination

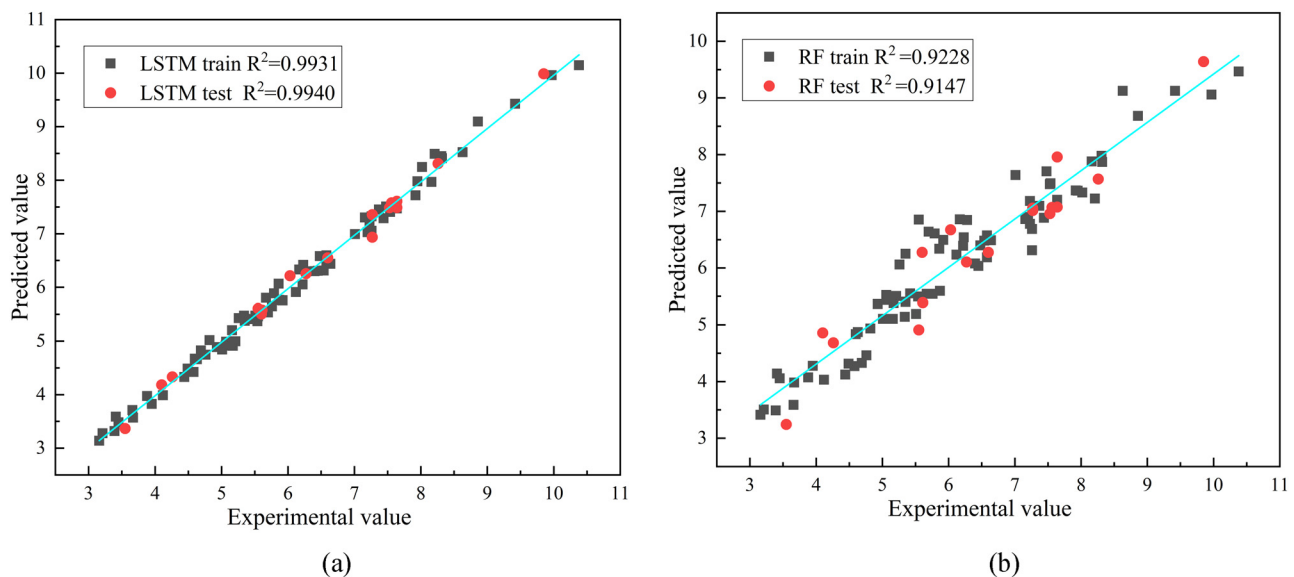


Figure 7: Fitting curves of different models: (a) LSTM model and (b) RF model.

between the predicted value and the experimental value. In addition, the R^2 of the LSTM model and the RF model were close to each other in the training set and the test set, indicating that the models were not overfitted, which verified the effectiveness of the models.

Statistical analysis was carried out on the LSTM model and RF model, and the evaluation indicators are shown in Table 5. It could be seen that the R^2 value of the LSTM model was 0.9940, which was closer to 1 than that of the RF model, the RMSE value was 0.1248, and the MAE value was 0.0960, all of which were much lower than those of RF model, indicating that the prediction accuracy of the LSTM model was higher. The results of this study were similar to those of previous literature. Gao [40] realized the deep learning CNN could predict the compressive strength of recycled concrete and compared it with the traditional machine learning BP neural network and SVM. The results showed that CNN had higher prediction accuracy, and the training error of 28-day compressive strength was 0.25%. The test error was 0.66%. In addition, Latif [27] predicted the compressive strength of concrete through deep learning LSTM and compared it with the traditional machine learning SVM. The results showed that the accuracy of the LSTM model was better than that of SVM, and the R^2 values were 0.98 and 0.78, respectively. Similarly, Chen *et al.* [41] used LSTM to predict the compressive strength of high-strength concrete and compared it with the traditional SVR. The results showed that the LSTM model had higher accuracy and reliability. Therefore, it could be concluded that the results of this study were similar to those of other studies in the field, and LSTM could be used as a reliable prediction model for CSG.

6 Interpretability analysis

The above research showed that for CSG with a given mix ratio, the LSTM model can accurately predict its compressive strength. For CSG, if the predicted compressive strength value does not meet expectations, the content of each component of CSG needs to be adjusted continuously. However, in a situation where the influence and contribution of each

Table 5: Model performance comparison

Model	Dataset	R^2	RMSE	MAE
LSTM	Train set	0.9931	0.1377	0.1179
	Test set	0.9940	0.1248	0.0960
RF	Train set	0.9228	0.4655	0.3758
	Test set	0.9147	0.4809	0.4397

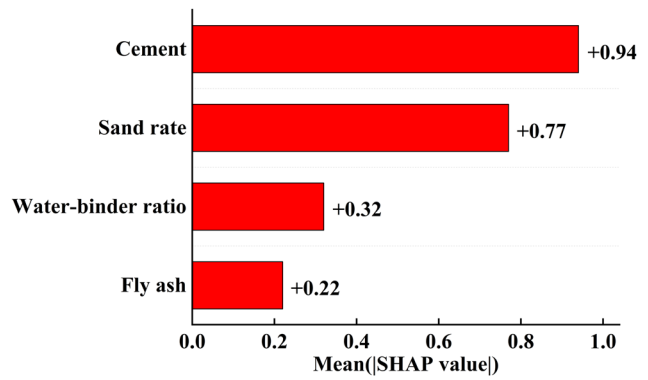


Figure 8: Global importance of the input variables.

input variable to the output result are unknown, these attempts are blind, and there will be a lot of trial and error. Based on this, in order to enhance the interpretability of the model, this article put forward a displayed SHAP, which can explain the machine learning method to study the importance of each input variable to the output and contribution to the size of the positive and negative.

As shown in Figure 8, the average SHAP value on the X axis indicates how important the input variable was to the output result. It could be found that in this study, the input variable that had the greatest influence on the compressive strength of CSG was cement content, followed by sand rate, water-binder ratio, and fly ash. In addition, the influence of global features that illuminated the input features is illustrated in Figure 9, where each point represents the features and SHAP value observed separately in the data set. The X axis represents the SHAP value of each input variable, i.e., the impact of each input variable against pressure strength, and the Y axis represents the importance ranking of the four input variables. The high eigenvalue of each sample in the figure indicates that this input variable had a positive and negative effect on the output result. It could be clearly seen from Figure 9 that the influence of cement content and fly ash content on compressive strength was positive, i.e., the compressive strength increased with the increase in dosage. On the contrary, the influence of sand

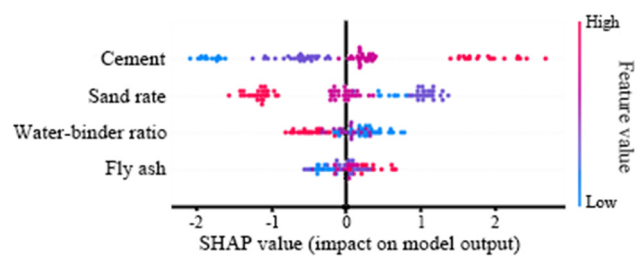


Figure 9: Global feature influences of the input features.

ratio and water-binder ratio on compressive strength was negative, and the increase in the content of these two components would have led to a decrease in the compressive strength of CSG, which was consistent with the actual law in a certain range, and the interpretation result had a certain credibility.

7 Conclusion

In this article, deep learning LSTM is used to predict the compressive strength of CSG, and the prediction results of the LSTM model are compared with machine learning RF. The main conclusions are summarized as follows:

- 1) The LSTM model can well deal with the complex non-linear relationship between variables, and the coefficient of determination R^2 exceeds 0.99 in both the training set and the test set.
- 2) Compared with the traditional machine learning RF, deep learning LSTM has higher prediction accuracy and better performance. It can be used as a method to predict the compressive strength of CSG. The predicted compressive strength of CSG can be obtained through LSTM model prediction before the laboratory compression test, which will greatly reduce the time and material costs of the laboratory compression test. Good for environmental protection.
- 3) Among the four input variables in this article, cement content and sand rate are the two variables that have the greatest influence on compressive strength.
- 4) The influence of cement content and fly ash content on compressive strength is positive, and the compressive strength increases with the increase in the content, while the influence of sand rate and water-binder ratio on compressive strength is negative, and their increase will lead to a decrease in the compressive strength of CSG.

In this article, LSTM and RF models are proposed to predict the compressive strength of CSG, both of which can be predicted successfully. LSTM has better accuracy and generalization ability, but there are some limitations. Since LSTM is a recursive neural network, it requires a lot of computational resources and time, and it is recommended to develop a more simplified deep learning model in future studies. The performance of RF models is often limited when processing high-dimensional data, and it is suggested that this problem should be studied in future studies. In addition, it is very important to evaluate the generalization ability of the model, which is also one of the challenges faced by the current research field. In future studies, more data sets can be used to evaluate the generalization ability

of the model in this article so as to better adapt to the needs of practical applications. In addition to expanding the data set, it is necessary to integrate the prediction model into the existing system or software before practical application and consider the strategy of data update and model retraining to ensure the effectiveness and sustainability of the model. Furthermore, better methods can be adopted for model interpretation and analysis to determine the limitations and risks of the model.

Acknowledgments: The authors would like to thank the National Key R & D Plan Project (2021YFC3001000) and the Natural Science Foundation of China (52109154). The authors also express their sincere gratitude to the editor and reviewers for their valuable comments, which have greatly improved this article.

Funding information: This work was funded by the National Key R & D Plan Project (2021YFC3001000) and the Natural Science Foundation of China (52109154).

Author contributions: Hang Gao and Zexuan Li were responsible for the experiments and data collation in this article. Lei Guo, Qingqing Tian, and Qiongyao Wang were responsible for the overall direction and advice of the article. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

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

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

References

- [1] Jia, J., M. Lino, F. Jin, and C. Zheng. The cemented material dam: a new, environmentally friendly type of dam. *Engineering*, Vol. 2, No. 4, 2016, pp. 490–497.
- [2] Huang, H., K. Huang, X. C. Zhang, and L. W. Han. Hysteresis and damping effect of cemented sand and gravel material under cyclic loading. *Journal of Building Materials*, Vol. 21, No. 5, 2018, pp. 739–748.
- [3] Jiang, M., X. Cai, X. Guo, Q. Liu, and T. Zhang. Adiabatic temperature rise test of cemented sand and gravel (CSG) and its application to temperature stress prediction of CSG dam. *Advances in Materials Science and Engineering*, Vol. 2020, 2020, pp. 1–12.
- [4] Sun, M. Q., L. Guo, S. F. Yang, Q. H. Chai, S. K. Chen, L. W. Han, et al. *Study on mechanical properties, durability and dam type of cement-sand-gravel materials*. Beijing: China Water Conservancy and Hydropower Publishing House; 2016.

- [5] Du, W. T., Q. C. Wang, J. P. Dai, B. Zhang, R. S. Bi, and H. Cao. Study on compressive strength of concrete under the action of multiple factors. *Concrete*, Vol. 396, No. 10, 2022, pp. 43–4651.
- [6] Yan, L., J. X. He, and H. H. Yang. Research on the anti-compression and anti-freeze properties of the cemented sand and gravel material. *Journal of Water Resources and Water Engineering*, Vol. 30, No. 1, 2019, pp. 197–202.
- [7] Duan, R., M. R. Boland, Z. Liu, Y. Liu, H. H. Chang, H. Xu, et al. Statistical law and predictive analysis of compressive strength of cemented sand and gravel. *Science and Engineering of Composite Materials*, Vol. 27, No. 1, 2020, pp. 291–298.
- [8] Li, X., S. Zhang, H. Wu, and D. Tian. Study on the material properties of cement-sand-gravel. *Conference Series: Earth and Environmental Science*. IOP Publishing: Vol. 513, No. 1, 2020, id. 012036.
- [9] Chai, Q., F. Wan, L. Xiao, and F. Wu. The influence of fly ash content on the compressive strength of cemented sand and gravel material. *Crystals*, Vol. 11, No. 11, 2021, pp. 1426–1434.
- [10] Ma, J. X., M. Ye, K. Ma, K. Zhou, Y. Zhang, X. Wang, et al. Fire prediction based on CatBoost algorithm. *Mathematical Problems in Engineering*, Vol. 2021, 2021, pp. 1–9.
- [11] Liu, W., Z. X. Chen, and Y. Hu. XGBoost algorithm-based prediction of safety assessment for pipelines. *International Journal of Pressure Vessels and Piping*, Vol. 197, 2022, pp. 104655–104668.
- [12] Ilić, M., Z. Srdjević, and B. Srdjević. Water quality prediction based on Naïve Bayes algorithm. *Water Science and Technology*, Vol. 85, No. 4, 2022, pp. 1027–1039.
- [13] Wang, S., W. Yuan, J. Zhou. Analysis of runoff coefficient prediction based on LM-BP neural network. *Journal of Physics: Conference Series*, Vol. 2333, No. 1, 2022, id. 012020, IOP Publishing.
- [14] Wu, Y. Q. and Y. S. Zhou. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Construction and Building Materials*, Vol. 330, 2022, pp. 127298–127308.
- [15] Wu, Y. Q. and Y. S. Zhou. Splitting tensile strength prediction of sustainable high-performance concrete using machine learning techniques. *Environmental Science and Pollution Research*, Vol. 29, No. 59, 2022, pp. 89198–89209.
- [16] Latif, S. D. Developing a boosted decision tree regression prediction model as a sustainable tool for compressive strength of environmentally friendly concrete. *Environmental Science and Pollution Research*, Vol. 28, No. 46, 2021, pp. 65935–65944.
- [17] Halim, F., A. Ahmad, M. Adil, A. Khan, M. Ghareeb, M. Alzara, et al. Compressive strength prediction of fly ash-based geopolymers concrete via advanced machine learning techniques. *Case Studies in Construction Materials*, Vol. 16, 2022, id. e00840.
- [18] Yuan, X., Y. Tian, W. Ahmad, A. Ahmad, K. I. Usanova, A. M. Mohamed, et al. Machine learning prediction models to evaluate the strength of recycled aggregate concrete. *Materials*, Vol. 15, No. 8, 2022, pp. 2823–2847.
- [19] Mozumder, R. A., B. Roy, and A. I. Laskar. Support vector regression approach to predict the strength of FRP confined concrete. *Arabian Journal for Science and Engineering*, Vol. 42, 2017, pp. 1129–1146.
- [20] Liu, C., W. Y. Hou, and D. Liu. Foreign exchange rates forecasting with convolutional neural network. *Neural Processing Letters*, Vol. 46, 2017, pp. 1095–1119.
- [21] Salinas, D., V. Flunkert, J. Gasthaus, and T. Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, Vol. 36, No. 3, 2020, pp. 1181–1191.
- [22] Wu, Y., J. Zhang, C. Gao, and J. Xu. Internal defect detection quantification and three-dimensional localization based on impact echo and classification learning model. *Measurement*, Vol. 218, 2023, pp. 113153–113167.
- [23] Zhang, P. Y. and B. C. Ci. Deep belief network for gold price forecasting. *Resources Policy*, Vol. 69, 2020, pp. 101806–101816.
- [24] Bilal, A., A. Jourabloo, M. Ye, X. Liu, and L. Ren. Do convolutional neural networks learn class hierarchy? *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 1, 2017, pp. 152–162.
- [25] Landi, F., L. Baraldi, M. Cornia, and R. Cucchiara. Working memory connections for LSTM. *Neural Networks*, Vol. 144, 2021, pp. 334–341.
- [26] Qiu, R., Y. Wang, B. Rhoads, D. Wang, W. Qiu, Y. Tao, et al. River water temperature forecasting using a deep learning method. *Journal of Hydrology*, Vol. 595, 2021, pp. 126016–126027.
- [27] Latif, S. D. Concrete compressive strength prediction modeling utilizing deep learning long short-term memory algorithm for a sustainable environment. *Environmental Science and Pollution Research*, Vol. 28, No. 23, 2021, pp. 30294–30302.
- [28] Hochreiter, S. and J. Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735–1780.
- [29] Greff, K., R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, No. 10, 2016, pp. 2222–2232.
- [30] Shrestha, A. and A. Mahmood. Review of deep learning algorithms and architectures. *IEEE Access*, Vol. 7, 2019, pp. 53040–53065.
- [31] Wang, K. J., X. X. Qi, and H. D. Liu. Photovoltaic power forecasting based LSTM-convolutional network. *Energy*, Vol. 189, 2019, pp. 116225–116235.
- [32] Breiman, L. *Random forests*. *Machine Learning*, Vol. 45, No. 1, 2001, pp. 5–32.
- [33] Lee, Y. G., J. Y. Oh, D. Kim, and G. Kim. SHAP value-based feature importance analysis for short-term load forecasting. *Journal of Electrical Engineering & Technology*, Vol. 18, No. 1, 2023, pp. 579–588.
- [34] Domingos, P. A few useful things to know about machine learning. *Communications of the ACM*, Vol. 55, No. 10, 2012, pp. 78–87.
- [35] Yang, X., Y. Wang, R. Byrne, G. Schneider, and S. Yang. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, Vol. 119, 2019, pp. 3–11.
- [36] Gandomi, A. H., S. K. Babanajad, A. H. Alavi, and Y. Farnam. Novel approach to strength modeling of concrete under triaxial compression. *Journal of Materials in Civil Engineering*, Vol. 24, 2012, pp. 1132–1143.
- [37] Kandasamy, K., W. Neiswanger, J. Schneider, B. Poczos, and E. P. Xing. Neural architecture search with Bayesian optimisation and optimal transport. *Advances in Neural Information Processing Systems*, Vol. 31, 2018, pp. 1–10.
- [38] Ilievski, I., T. Akhtar, J. Feng, and C. Shoemaker. Efficient hyperparameter optimization for deep learning algorithms using deterministic RBF surrogates. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, No. 1, 2017, pp. 822–829.
- [39] Mai, H. V. T., T. A. Nguyen, H. B. Ly, and V. Q. Tran. Prediction compressive strength of concrete containing GGBFS using random forest model. *Advances in Civil Engineering*, Vol. 2021, 2021, pp. 1–12.
- [40] Gao, W. Influencing factors and deep learning prediction model of compression strength of recycled concrete. *Concrete*, Vol. 11, 2018, pp. 58–61 + 70.
- [41] Chen, H., X. Li, Y. Wu, L. Zuo, M. Lu, and Y. Zhou. Compressive strength prediction of high-strength concrete using long short-term memory and machine learning algorithms. *Buildings*, Vol. 12, No. 3, 2022, pp. 302–313.