

## Research Article

# Prediction Compressive Strength of Concrete Containing GGBFS using Random Forest Model

Hai-Van Thi Mai , Thuy-Anh Nguyen , Hai-Bang Ly , and Van Quan Tran 

*University of Transport Technology, Hanoi 100000, Vietnam*

Correspondence should be addressed to Van Quan Tran; quantv@utt.edu.vn

Received 28 December 2020; Revised 13 April 2021; Accepted 30 April 2021; Published 10 May 2021

Academic Editor: Junfei Zhang

Copyright © 2021 Hai-Van Thi Mai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Improvement of compressive strength prediction accuracy for concrete is crucial and is considered a challenging task to reduce costly experiments and time. Particularly, the determination of compressive strength of concrete using ground granulated blast furnace slag (GGBFS) is more difficult due to the complexity of the composition mix design. In this paper, an approach using random forest (RF), which is one of the powerful machine learning algorithms, is proposed for predicting the compressive strength of concrete using GGBFS. The RF model is first evaluated to determine the best architecture, which constitutes 500 growth trees and leaf size of 1. In the next step, the evaluation of the model is conducted over 500 simulations considering the effect of random sampling. Finally, the best prediction results are given in function of statistical measures such as the correlation coefficient ( $R$ ), root mean square error (RMSE), and mean absolute error (MAE), respectively, which are 0.9729, 4.9585, and 3.9423 for the testing dataset. The results show that the RF algorithm is an excellent predictor and practically used for engineers to reduce experimental cost.

## 1. Introduction

Nowadays, ground granulated blast furnace slag (GGBFS) has been used as supplementary cementitious material in Portland concretes. GGBFS is a product of the glassy granular material formed when molten blast furnace slag is quickly cooled by water. GGBFS can replace 35–65% Portland cement in concrete. Using GGBFS as a partial replacement of Portland cement enhances concrete strength and durability through a denser matrix formation. It could also increase the performance of concrete structures. Moreover, GGBFS as a partial replacement requires approximately only 25% of the energy needed to produce Portland cement [1–4]. Thanks to these numerous advantages, the determination of GGBFS content in the design phase of concrete is essential and meaningful, especially in improving the concrete compressive strength.

Numerous investigations were performed to calculate GGBFS concrete mix design, including the experimental and statistical methods. Some experimental investigations have

been carried out to estimate the compressive strength of GGBFS concrete [5–9]. However, the experimental methods are, in general, time consuming and relatively costly. Besides, several mathematical models and empirical equations have been proposed to estimate the compressive strength of concrete. In fact, the mathematical equations are regression equations based on the experimental results. As an example, with four cement dosages, including 175, 210, 245, and 280 kg/m<sup>3</sup>, Oner and Akyuz [5] have proposed two equations and obtained a correlation coefficient of 0.99. However, this value is validated only in the range of cement dosage of the authors. The accuracy of regression equations is strongly dependent on the number of experimental tests and the range of cement content. Therefore, a new approach needs to be developed for reducing the time consumed and experimental cost due to a high number of experimental tests. Also, a universal prediction approach with high prediction accuracy needs to be constructed.

In recent years, artificial intelligence (AI) or machine learning (ML) is gradually becoming popular and applied in

numerous scientific fields [10–13]. The random forest (RF) is one of the most powerful algorithms of ML for data science, which has been widely used in the construction field [14]. The RF model is successfully applied to solve numerous technical issues of civil engineering [15–18], geotechnical engineering [19–22], earth sciences [23–25], and environmental protection [26, 27]. For example, Mohana [17] has used the RF model and 268 experimental data to predict the compressive strength of concrete containing GGBFS. The author has also mentioned that the prediction accuracy is evaluated through the correlation coefficient ( $R$ ), equal to 0.94, for the best RF architecture. Such a value is considered low due to the number of simulations performed by the author. Moreover, the number of experimental data is limited, which reduces the prediction accuracy of concrete compressive strength using the RF model.

Therefore, the primary purpose of this study is to propose an efficient RF model to increase the compressive strength prediction accuracy of concrete containing GGBFS, thanks to the higher data samples collected from the literature. Moreover, efficient RF architecture will be performed by performing numerous simulations for increasing the RF model's reliability. Precisely, the performance of ML model is strongly affected by the parameter or architecture selections of the corresponding ML algorithms. Therefore, this study performs firstly the determination of RF architecture for better predicting the compressive strength of concrete containing GGBFS. To acquire the purpose, numerous experimental samples from the literature are gathered and randomly split up into two parts, namely, the training part (70% of data) and the testing part (30% of data). The best RF architecture is obtained and used to predict the compressive strength of concrete containing GGBFS with the evaluation of three statistical measurements, consisting of the correlation coefficient ( $R$ ), mean absolute error (MAE), and root mean square error (RMSE). The sensitivity of concrete compressive strength on different input parameters is finally performed.

## 2. Significance of the Research Study

Accurate prediction of the concrete compressive strength using supplementary cementitious materials, such as GGBFS, is crucial thanks to many further advantages and contributions to construction design. Although many machine learning models have been proposed to predict the compressive strength of concrete in the available literature (i.e., [28–32]), the reliability assessment of the models still needs to be quantified. Therefore, the present investigation is proposed in order to

- (1) introduce the variability in the sampling process to construct the training and testing datasets
- (2) assess the prediction reliability of the RF model using Monte Carlo simulations
- (3) finely tune the hyperparameters to obtain the best RF model

- (4) show that the performance of the best model is compared with 7 investigations published in the literature, confirming its simplicity and effectiveness
- (5) show a reliable variable importance analysis by taking the average results of 500 simulations

## 3. Database Construction

The experimental database used in this study is collected from published articles [5, 6, 31, 33, 34] (Table 1). There are 453 samples, divided into two parts, 70% training data (317 samples) and 30% testing data (136 samples). Two shapes of samples include 327 cubic samples and 36 cylindrical samples, about 8% of all samples; therefore, the shape variable is not considered in this investigation. The total databases include 8 input variables from  $X_1$  to  $X_8$ : cement content,  $\text{kg}/\text{m}^3$  ( $X_1$ ); water content,  $\text{kg}/\text{m}^3$  ( $X_2$ ); coarse aggregate,  $\text{kg}/\text{m}^3$  ( $X_3$ ); fine aggregate or sand,  $\text{kg}/\text{m}^3$  ( $X_4$ ); GGFBS content,  $\text{kg}/\text{m}^3$  ( $X_5$ ); hyperplasticizing,  $\text{kg}/\text{m}^3$  ( $X_6$ ); superplasticizer, % ( $X_7$ ); and age of samples, day ( $X_8$ ). The output variable of the present study is the compressive strength, MPa (denoted as  $Y$ ). The corresponding correlation analysis of data is shown in Figure 1.

The input variables from  $X_1$  to  $X_5$  are distributed in a wide range, while the variables  $X_6$  to  $X_8$  are in a narrow range. Precisely, the cement content ( $X_1$ ) ranges from 70 to 360 ( $\text{kg}/\text{m}^3$ ), but it is mainly in the range of 180 to 270 ( $\text{kg}/\text{m}^3$ ). The highest sample number is about 79, which corresponds to 180  $\text{kg}/\text{m}^3$  of cement content. Similarly, the water content ( $X_2$ ) ranges from 70 to 295  $\text{kg}/\text{m}^3$ . As shown in Figure 1, the coarse aggregate content ( $X_3$ ) is varied from about 400 to 1200 ( $\text{kg}/\text{m}^3$ ), but no sample has coarse aggregate content in the range of 500 to 700 ( $\text{kg}/\text{m}^3$ ). The fine aggregate or sand content ( $X_4$ ) is mainly in two ranges from 500 to 950 ( $\text{kg}/\text{m}^3$ ) and 1150 to 1550 ( $\text{kg}/\text{m}^3$ ). The highest sample number ( $X_4$ ) corresponds to 680  $\text{kg}/\text{m}^3$  of fine aggregate (or sand content). The GGBFS content ( $X_5$ ) varies from 40 to 460  $\text{kg}/\text{m}^3$ , but the values are mostly in the range of 70 to 270 ( $\text{kg}/\text{m}^3$ ). The carboxylic-type hyperplasticizing content ( $X_6$ ) ranges from 2 to 14  $\text{kg}/\text{m}^3$ . However, hyperplasticizing is not used in almost all cases, accounting for about 330 samples (on a total of 453). Besides, almost all samples have zero superplasticizer content ( $X_7$ ) except for six samples, representing only a proportion of 1%. With the age of samples, there are ten values; the minimum age of the sample is one day, and the maximum age of the sample is 365 days.

The correlations between the inputs and compressive strength are plotted in Figure 2. The correlation values are shown in different colors. As clearly shown, some of the variables are slightly correlated, such as  $X_4$  and  $X_6$  for aggregate content and carboxylic-type hyperplasticizing content, respectively. Overall, the correlation between the inputs and compressive strength is relatively low. Therefore, all variables are included to increase the accuracy of the final model developed.

## 4. Simulation Using Random Forest

**4.1. Random Forest.** Random forest (RF) [35] is an ensemble for classification and regression developed by Leo Breiman at the University of California, Berkeley. Breiman is also a

TABLE 1: Detail of database collection.

No.	Reference	Data number	Shape of sample	Percentage (%)
1	Oner and Akyuz [5]	168	Cubic	37.09
2	Shariq et al. [6]	63	Cubic	13.91
3	Chidiac and Panesar [34]	36	Cylindrical	7.95
4	Boga et al. [31]	6	Cubic	1.32
5	Bilim et al. [33]	180	Cubic	39.73
Total		453		100%

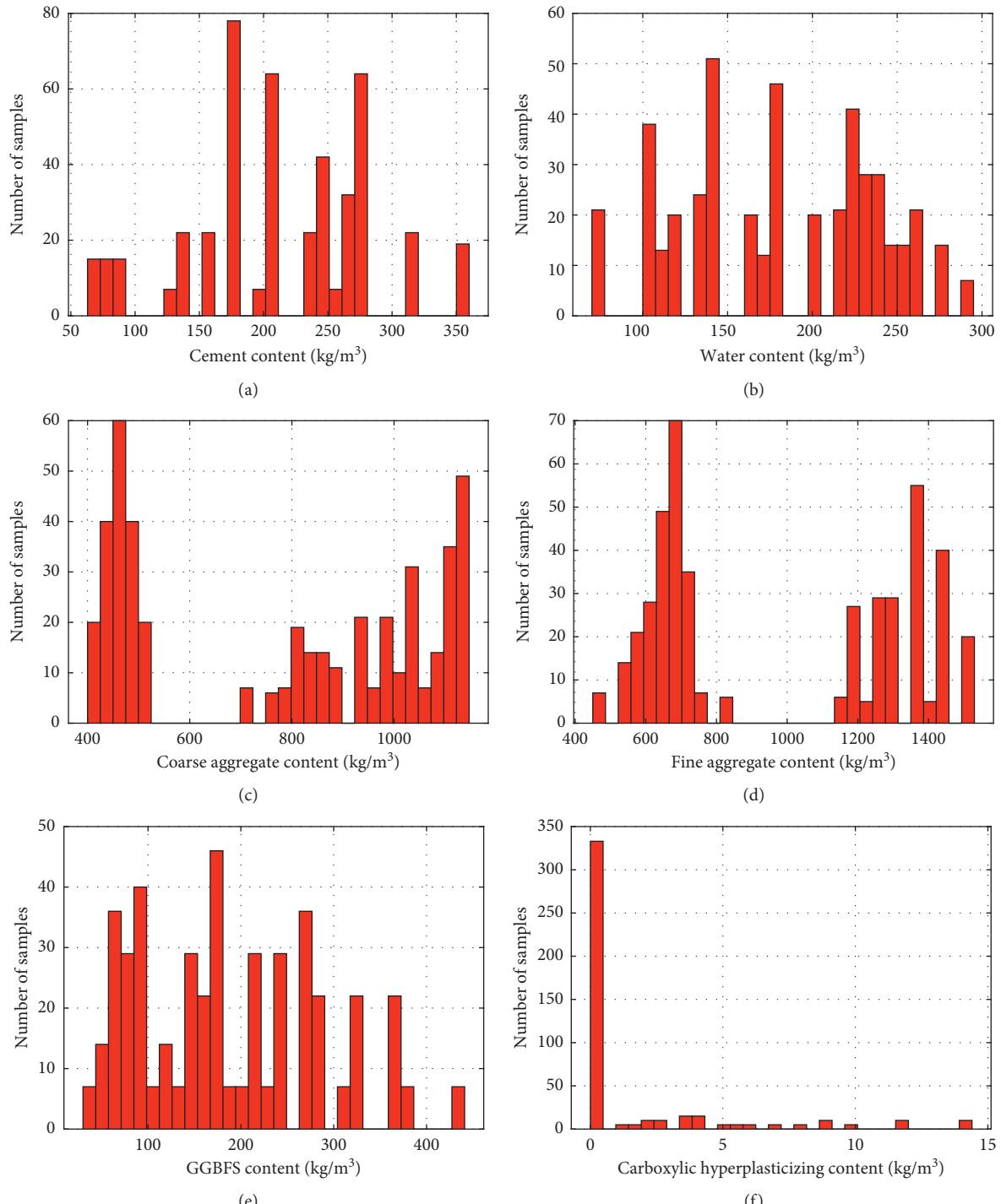


FIGURE 1: Continued.

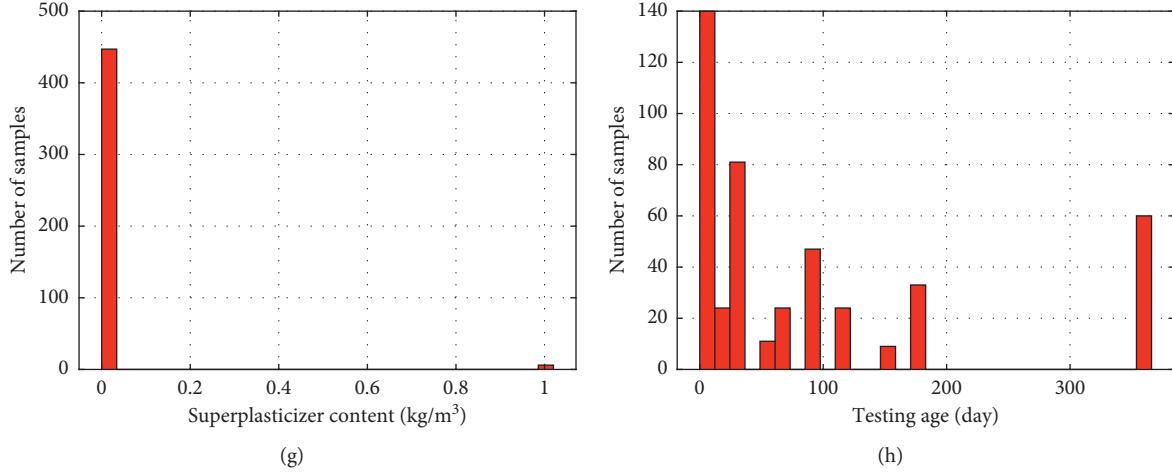


FIGURE 1: Histograms of the input variables used in this study: (a) cement content; (b) water content; (c) coarse aggregate content; (d) fine aggregate (or sand) content; (e) ground granulated blast furnace slag content; (f) carboxylic-type hyperplasticizing content; (g) superplasticizer content; and (h) testing age of samples.

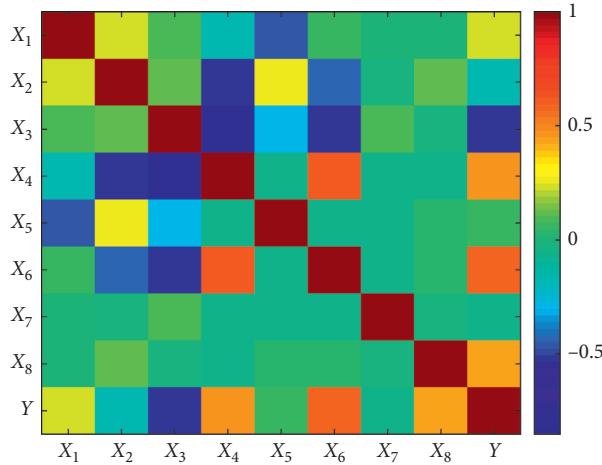


FIGURE 2: Multicorrelation graph of input and output variables used in this study.

coauthor of the classification and regression tree (CART) method [36]. Random forest (RF) is an innovative method of bootstrap aggregating (bagging). RF uses 2 random steps, in which the first one is sample randomization using the bootstrap method with a return and the second one is to randomly extract a number of attributes from the original attribute set. These 2 randomized subdatasets are highly diverse and unrelated, helping to reduce variance. CARTs constructed from this subset of data form the forest. When aggregating results, RF uses a voting method for the classification problem and takes the average value for the regression problem.

In recent years, RF is used quite commonly because of its superiority compared with other algorithms; it can handle data with a large number of properties and able to estimate the importance of the attributes, often with high accuracy in classification (or regression) and fast learning process. In RF,

each tree selects only a small set of attributes during construction (2nd random step); this mechanism makes the RF execute with the dataset with a large number of attributes in an acceptable time when calculating. The user can default to the number of properties to construct trees in the forest; normally the optimal default is  $\sqrt{p}$  for the classification problem and  $p/3$  for regression problems ( $p$  is the number of all properties of the original dataset). The number of trees in the forest should be set large enough to ensure that all attributes are used a number of times usually 500 trees for the classification problem and 1000 trees for the regression problem. Due to the use of the bootstrap method of random return sampling, the subdatasets have about 2/3 of the samples that do not overlap for tree construction. About 1/3 of the remaining samples are called out-of-bag because they do not participate in the construction of trees, so RF uses these out-of-bag samples to test and calculate the CART's

attribute importance in the forest. To summarize, the random forest algorithm is built according to the following steps (Figure 3):

*Step 1.* From dataset  $D$ , we generate random data (bootstrap sample)

*Step 2.* Using random sampling data subsets  $D_1, D_2, \dots, D_k$  to build trees  $T_1, T_2, \dots, T_k$

*Step 3.* Combine trees: use the majority voting strategy with the classification problem or average the predicted values from trees with the regression problem

Overall, the RF model is selected in this study because of many advantages, such as the prediction accuracy, fast simulation speed, robustness to noise, and overfitting [37], and easily parallelized, especially useful for Monte Carlo simulations and useful for error estimation and determination of variable importance.

**4.2. Validation of Models.** In this study, three statistical criteria are used to evaluate the error between the actual value and the predicted value of the compressive strength of concrete, namely, correlation coefficient ( $R$ ), root mean square error (RMSE), and mean absolute error (MAE). The  $R$ -value is used to examine the linear correlation between the actual value and the predicted value in the range  $[-1; 1]$ . Both the RMSE and MAE measure the average error between the actual and predicted outputs, in which RMSE is used to evaluate the difference in value between actual and predicted, and MAE displays the average error between actual and prediction values. For these two indicators, the smaller value denotes the better performance of the model. The closer the absolute value of  $R$  is to 1, the more accurate the RF model is in predicting the compressive strength of concrete.  $R$ , RMSE, and MAE are determined by the following equations:

$$\begin{aligned} R &= \frac{\sum_{j=1}^N (Q_{0,j} - \bar{Q}_0)(Q_{t,j} - \bar{Q}_t)}{\sum_{j=1}^N (Q_{0,j} - \bar{Q}_0)^2 \sum_{j=1}^N (Q_{t,j} - \bar{Q}_t)^2}, \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{j=1}^N (Q_{0,j} - Q_{t,j})^2}, \\ \text{MAE} &= \frac{1}{N} \sum_{j=1}^N |Q_{0,j} - Q_{t,j}|, \end{aligned} \quad (1)$$

where  $N$  is the number of database,  $Q_0$  and  $\bar{Q}_0$  are the actual experimental value and the average real experimental value, and  $Q_t$  and  $\bar{Q}_t$  are the predicted value and the average predicted value, calculated according to the model output.

## 5. Methodology Flow Chart

The methodology of constructing the RF model to predict the compressive strength of concrete containing GGBFS is described in Figure 4 including the primary steps as follows:

*Step 1.* Preparation of the database: In this step, the database including 453 experimental results determining compressive strength of concrete is collected to build RF models. The basic parameters to predict the compressive strength of concrete containing GGBFS include 8 input variables. The dataset is randomly divided into two parts, where 70% of the data are used to train the RF model and the remaining 30% is used to validate the built model.

*Step 2.* Determination of the optimal of the RF architecture model: In this second step, the number of trees and the leaf size of the RF model are determined based on the results obtained by the dataset. The criterion used to validate an optimal RF model architecture is mean square error (MSE).

*Step 3.* Training the optimal model: In this step, the training dataset is used to train the RF model with the optimal architecture.

*Step 4.* Validating the model: In this final step, the testing dataset is used to test and validate the RF model. The performance of the RF model is evaluated by statistical criteria, including  $R$ , RMSE, and MAE.

## 6. Results and Discussion

**6.1. Investigation on Random Forest Architecture.** In this section, the RF architecture is determined through the mean square error (MSE), as shown in Figure 5. The MSE value is strongly affected by the grown trees and leaf size number. The number of grown trees ranges from 0 to 2000, and seven leaf size values are introduced in this investigation, including 1, 3, 5, 10, 20, 50, and 100. In all leaf size numbers, the RF model is stable after about 500 grown trees. It is observed that the higher the leaf size number, the lower the MSE value. Overall, the MSE value of RF is the lowest in the case of simulation with 1 leaf size and about 500 grown trees [500–1]. Therefore, the best architecture of the RF model constitutes one leaf size and 500 grown trees. This architecture is used for the subsequent investigation.

**6.2. Prediction Performance.** In this section, the RF model performance is assessed by three criteria such as  $R$ , RMSE, and MAE. However, due to the sampling technique in which the training database and testing database are randomly constructed, the results of the RF model need to be evaluated after a sufficient number of simulations. In this paper, 500 simulations seem to be appropriate for obtaining reliable results. After 500 simulations, RMSE, MAE, and  $R$  values of the training and testing parts are shown in Figures 6(a)–6(c), respectively, including the average and Std values. For the training part, the average value of RMSE is about 5.22, and the average value of MAE is 3.84, and the average value of  $R$  is 0.97. For testing, the average value of RMSE is about 7.26, and the average value of MAE is 5.36, and the average value of  $R$  is 0.95. It is observed that the Std value is relatively small, which indicates the RF architecture is relatively stable. This result shows that RF architecture is suitable for determining the compressive strength of concrete containing

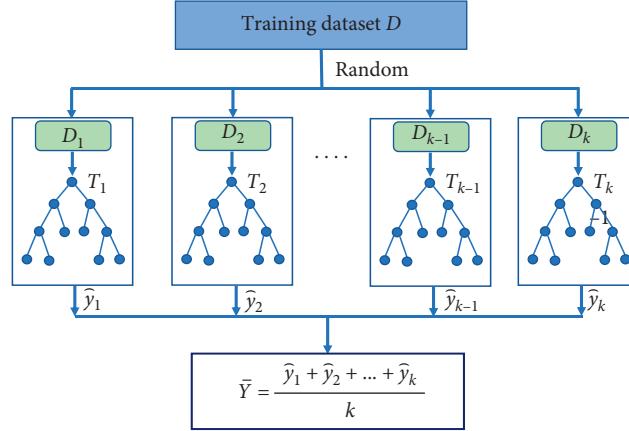


FIGURE 3: The architecture of the random forest algorithm.

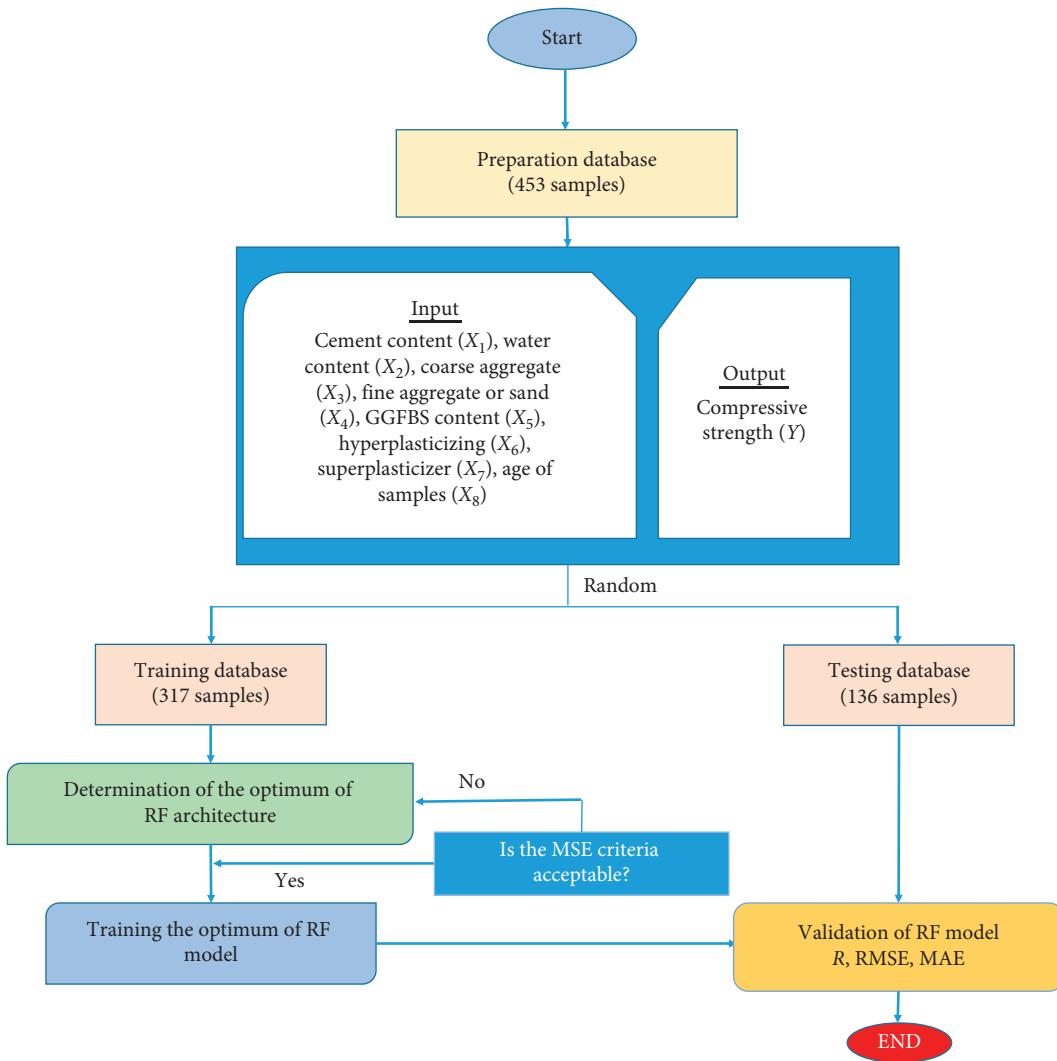


FIGURE 4: Methodology flow chart.

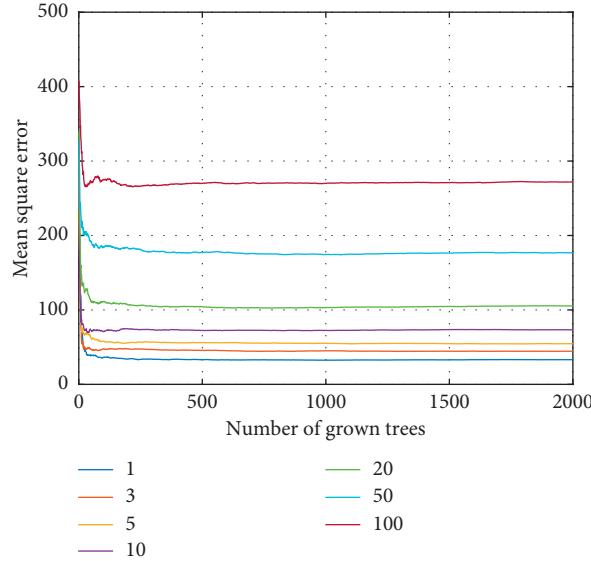


FIGURE 5: Values of MSE in function of grown tree number and leaf size number.

GGBFS. Besides, the Std value for training is smaller than Std value for the testing part, which means the model's performance for training is higher than that for testing. This could be useful to prevent overfitting problems. The exact values are given in Table 2.

**6.3. Prediction Accuracy.** Once the best architecture is found, this section is dedicated to the presentation of the best simulation using the RF algorithm. Figures 7(a) and 7(b) show the correlation between the experimental and the predicted values for the training and testing datasets under regression graphs, respectively. A linear fit is plotted in each case, in black color. It is worth noting that the correlation lines are very close to a perfect linear fit, which confirms an excellent agreement between the actual and predicted compressive strength of concrete using GGBFS.

The comparison shows that the predicted value is very close to the experimental value. The model error is plotted between the predicted value and the experimental value for the training database (Figure 8(a)) and the testing database (Figure 8(b)). The error values corresponding to the training and testing databases are small. Based on the cumulative distribution (black line), the percentage error of samples within a range can be easily determined. For example, with the training database, the percentage of samples with errors in the range  $[-5; 5]$  kN is about 70%. Similarly, the 60% error between the experimental value and the RF simulation of the testing database is about  $[-5; 5]$  kN. The results of the performance criteria show that the RF model with 500 grown trees and 1 leaf size architecture can quite accurately predict the compressive strength of concrete containing GGBFS.

Table 3 presents the different quality assessment criteria for the best RF predictor. The best value of  $R$  is 0.9759 for the

training part and for the testing part is 0.9729. The values of RMSE, MAE, Err. mean, and Err. Std for the training dataset are 5.4480, 4.1365, -0.0563, and 5.4563 and for the testing dataset are 4.9585, 3.9423, 0.6252, and 4.9647, respectively. Overall, these results show that RF could accurately predict the compressive strength of concrete containing GGBFS.

Table 4 shows the comparison of different machine learning models proposed in the literature with the proposed RF model. The comparisons are presented in the form of the machine learning algorithm, input number, number of data, and performance measure. The results show that the RF model of this investigation, containing 500 growth trees and 1 leaf size, could predict the compressive strength of concrete with higher reliability, higher accuracy, and low time consuming than that of almost all investigations. Based on a large number of simulations performed, this paper shows the simple architecture of the RF model for the higher prediction of concrete compressive strength. Overall, these results indicate that if the architecture of an ANN model is carefully selected, it could be used as an alternative prediction tool for material engineers.

Finally, Figure 9 shows the feature importance analysis of compressive strength with respect to each input variable using the RF model. There are 8 input databases, including cement content ( $X_1$ ), water content ( $X_2$ ), coarse aggregate ( $X_3$ ), fine aggregate or sand ( $X_4$ ), GGFBS content ( $X_5$ ), hyperplasticizing ( $X_6$ ), superplasticizer ( $X_7$ ), and age of samples ( $X_8$ ). After 500 simulations, the average value of  $X_7$  is the smallest, whereas the average value of  $X_8$  is the highest. The results show that the superplasticizer content exhibits the most negligible effect on the compressive strength of GGBFS concrete, which mainly depends on the testing age of samples. More importantly, different from most of the previously published results, the analysis shown in Figure 9 is conducted by taking the average feature importance re-

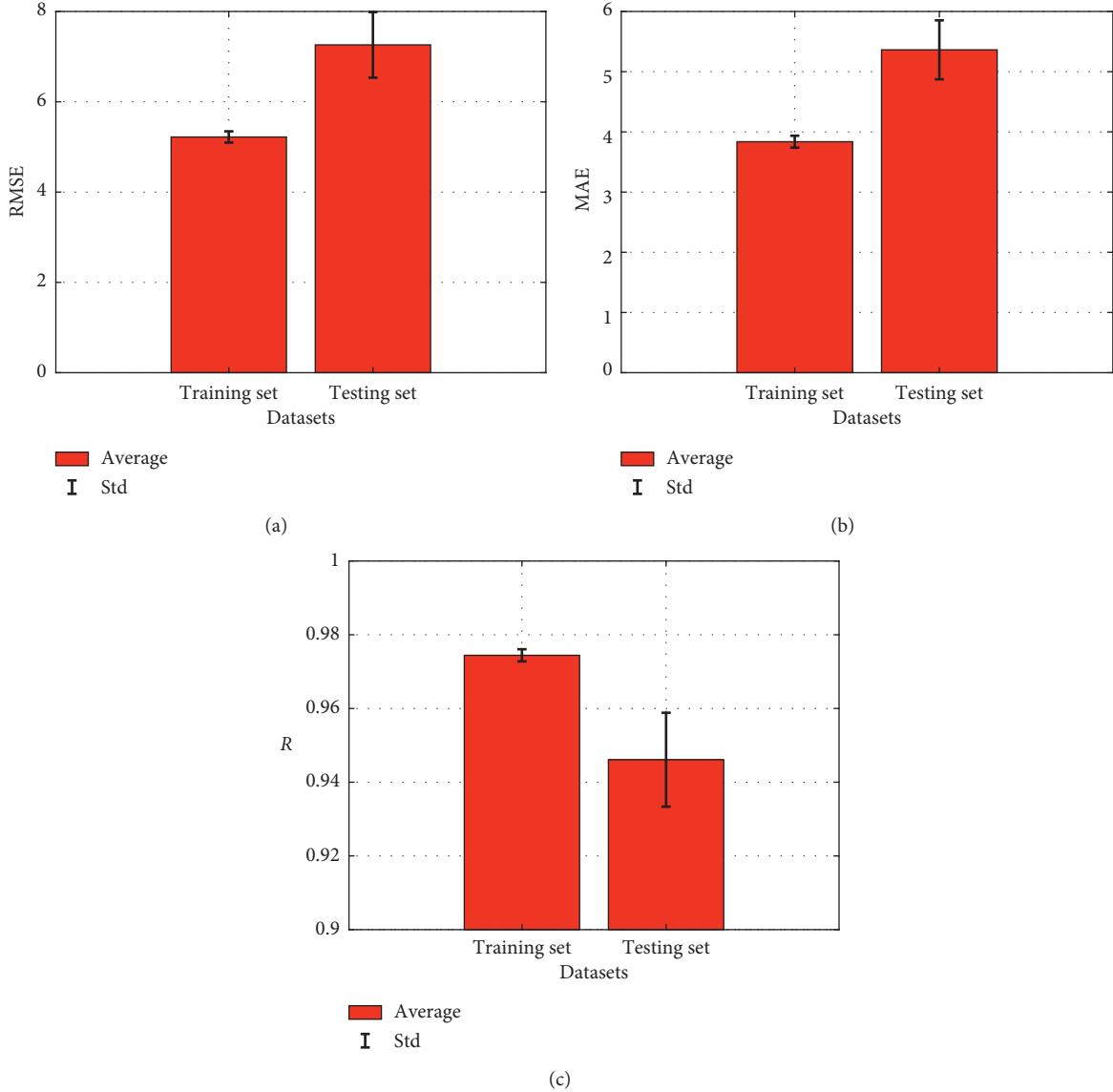


FIGURE 6: Analysis of the results over 500 simulations (presented in average values with standard deviation) using different RF architectures: (a)  $R$ ; (b) RMSE; (c) MAE.

TABLE 2: Summary of different quality assessment criteria over 500 simulations with the best RF architecture.

Criteria	R		RMSE		MAE	
	Training set	Testing set	Training set	Testing set	Training set	Testing set
Min	0.9700	0.9054	4.6951	4.9858	3.5261	3.9423
Average	0.9744	0.9461	5.2203	7.2602	3.8367	5.3628
Max	0.9805	0.9729	5.5260	9.6108	4.1365	6.8476
Std	0.0016	0.0127	0.1251	0.7264	0.0992	0.4900

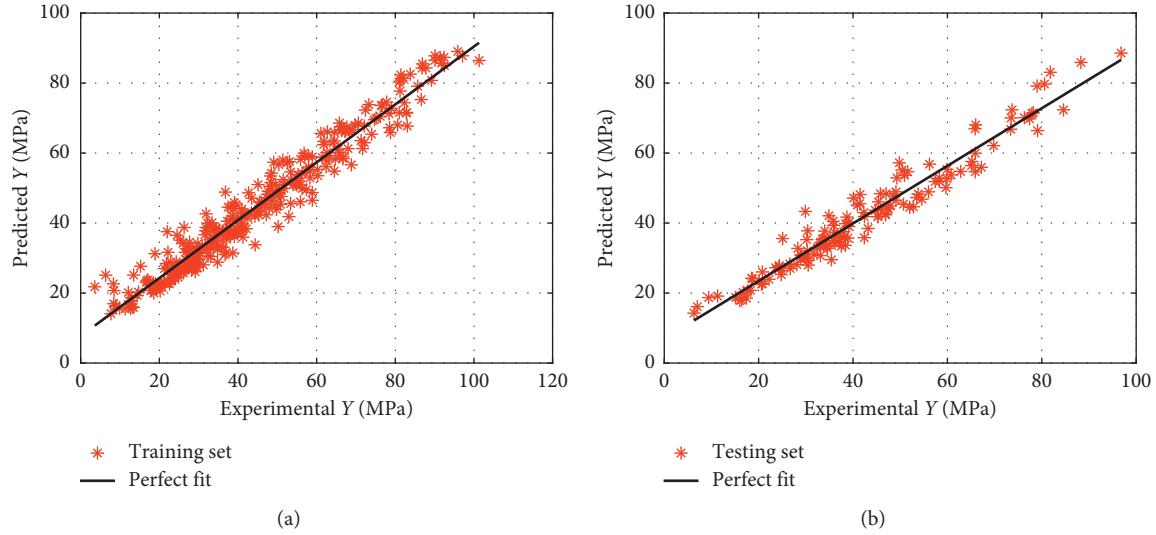


FIGURE 7: Regression graphs of the best predictor RF between experimental and predicted compressive strength: (a) training dataset and (b) testing dataset.

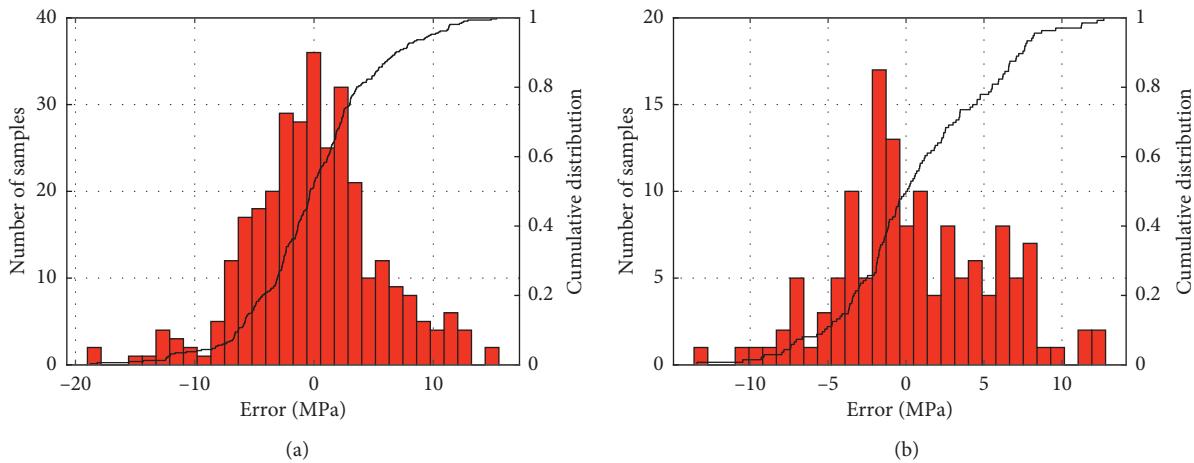


FIGURE 8: Error between target and output value plots for the case of the best RF architecture: (a) training dataset and (b) testing dataset.

TABLE 3: Summary of different quality assessment criteria for the best RF predictor.

	RMSE	MAE	Err. mean	Err. std	$R$
Training set	5.4480	4.1365	-0.0563	5.4563	0.9759
Testing set	4.9585	3.9423	0.6252	4.9647	0.9729

TABLE 4: Comparison of different machine learning models for predicting compressive strength of concrete.

Reference	Machine learning algorithm	Input	Number of data	Performance measure
Saridemir et al. [38]	ANN and fuzzy logic models ANFIS	5 inputs: TA, C, GGBFS, W, and Agg.	284	$R = 0.9904$
Bilim et al. [33]	ANN model	6 inputs: C, GGBFS, W, SP, Agg., and TA	225	$R = 0.9798$
Kandiri et al. [30]	Hybridized multiobjective ANN and a multiobjective slap swarm algorithm (MOSSA)/the M5P model tree algorithm	7 inputs: C, GGBFS, GGBFS grade (SG), W, fine Agg., coarse Agg., and TA	624	$R = 0.9700$
Han et al. [28]	ANN model	7 inputs: curing temperature, W/binder, GGBFS/total binder, W, fine Agg., coarse Agg., SP	269	$R = 0.9803$
Boukhatem et al. [29]	ANN model	5 inputs: C, W/C, GGBFS, temperature, TA	726	$R = 0.9600$
Boğa et al. [31]	ANN model and the adaptive neuro-fuzzy inference system (ANFIS)	4 inputs: cure type, curing period, BFS ratio, CNI ratio	162	$R = 0.9854$
This work	RF model	8 inputs: C, W, coarse Agg. or gravel, fine Agg. or sand, GGBFS, FA, SP, TA	453	$R = 0.9729$ , $MSE = 4.9585$ , $MAE = 3.9423$

C: cement; GGBFS: ground granulated blast furnace slag; W: water; SP: superplasticizer; TA: age of samples; FA: fly ash; Agg.: aggregate.

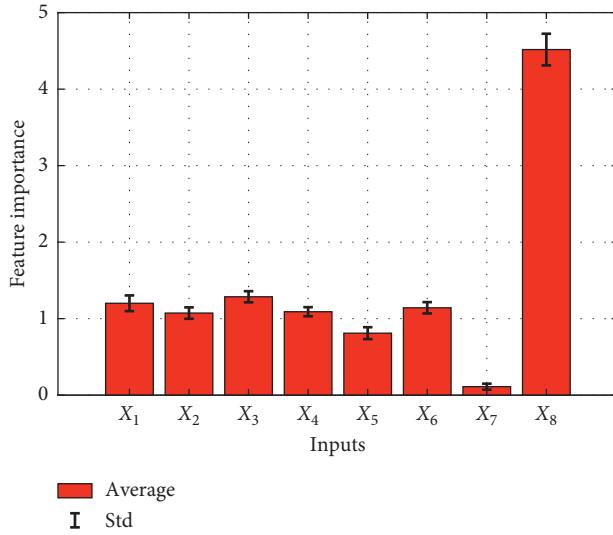


FIGURE 9: Feature importance over 500 simulations.

sults over 500 simulations. By doing so, the variability of the input space while constructing the RF model is minimized, and the reliability of the analysis is clearly shown (i.e., low Std values).

## 7. Conclusion

In this investigation, the RF algorithm is presented to predict the compressive strength of concrete containing GGBFS. A number of 453 experimental samples are gathered to develop the RF model. The database is randomly divided into two parts 70% of training data and 30% of testing data for the validation phase of the constructed RF model. To fully assess the RF model performance, a number of 500 simulations are performed using random

sampling technique. The results show that the RF architecture containing 500 growth trees and 1 leaf size is an excellent architecture to predict the compressive strength of concrete using GGBFS, in which the mean values of  $R$ , RMSE, and MAE are, respectively, 0.9461, 7.2602, and 5.3628 for the testing part. The highest accurate RF model shows an excellent prediction performance with  $R = 0.9729$ . The results can be used in developing a reliable model to predict accurately and quickly the compressive strength of concrete using GGBFS. Once the model is built, the prediction process would take a short time to estimate the compressive strength of a mix design.

Several short-term research directions of the present work could be mentioned. First, although the effectiveness of the RF model is clearly shown in this study, the

model's applicability can be improved by collecting more data samples with a broader range of input and output variables. This could be conducted based on the investigations of Golafshani and Behnood [39], as well as Behnood et al. [32]. The updated database is expected to be the most significant number of samples with the broadest range of values. Moreover, the accuracy of the RF model could be further improved by using optimization algorithms such as particle swarm optimization, genetic algorithm, and artificial bee colony algorithm in determining the best RF hyperparameters. Last but not least, the specimen type of compressive strength could also be considered as an input parameter. Overall, within the range of input and output variables of the present study, a numerical tool is developed in Matlab, which could be an alternative prediction tool for engineers to quickly estimate the compressive strength of concrete containing GGBFS.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] A. Behnood and E. M. Golafshani, "Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves," *Journal of Cleaner Production*, vol. 202, pp. 54–64, 2018.
- [2] A. Cheng, R. Huang, J.-K. Wu, and C.-H. Chen, "Influence of GGBS on durability and corrosion behavior of reinforced concrete," *Materials Chemistry and Physics*, vol. 93, no. 2-3, pp. 404–411, 2005.
- [3] E. Özbay, M. Erdemir, and H. İ. Durmuş, "Utilization and efficiency of ground granulated blast furnace slag on concrete properties-a review," *Construction and Building Materials*, vol. 105, pp. 423–434, 2016.
- [4] H. Song and V. Saraswathy, "Studies on the corrosion resistance of reinforced steel in concrete with ground granulated blast-furnace slag-An overview," *Journal of Hazardous Materials*, vol. 138, no. 2, pp. 226–233, 2006.
- [5] A. Oner and S. Akyuz, "An experimental study on optimum usage of GGBS for the compressive strength of concrete," *Cement and Concrete Composites*, vol. 29, no. 6, pp. 505–514, 2007.
- [6] M. Shariq, J. Prasad, and A. Masood, "Effect of GGBFS on time dependent compressive strength of concrete," *Construction and Building Materials*, vol. 24, no. 8, pp. 1469–1478, 2010.
- [7] R. Siddique and D. Kaur, "Properties of concrete containing ground granulated blast furnace slag (GGBFS) at elevated temperatures," *Journal of Advanced Research*, vol. 3, no. 1, pp. 45–51, 2012.
- [8] M. M. Tüfekçi and Ö. Çakır, "An investigation on mechanical and physical properties of recycled coarse aggregate (RCA) concrete with GGBFS," *International Journal of Civil Engineering*, vol. 15, no. 4, pp. 549–563, 2017.
- [9] R. K. Majhi, A. N. Nayak, and B. B. Mukharjee, "Development of sustainable concrete using recycled coarse aggregate and ground granulated blast furnace slag," *Construction and Building Materials*, vol. 159, pp. 417–430, 2018.
- [10] D. V. Dao, H.-B. Ly, H.-L. T. Vu, T.-T. Le, and B. T. Pham, "Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete," *Materials*, vol. 13, no. 5, p. 1072, 2020.
- [11] H. Q. Nguyen, H.-B. Ly, V. Q. Tran, T.-A. Nguyen, T.-T. Le, and B. T. Pham, "Optimization of artificial intelligence system by evolutionary algorithm for prediction of axial capacity of rectangular concrete filled steel tubes under compression," *Materials*, vol. 13, no. 5, p. 1205, 2020.
- [12] H.-B. Ly, P. G. Asteris, and B. T. Pham, "Accuracy assessment of extreme learning machine in predicting soil compression so efficient," *Vietnam Journal of Earth Sciences*, vol. 42, no. 3, 2020.
- [13] B. T. Pham, T. Nguyen-Thoi, H.-B. Ly et al., "Extreme learning machine based prediction of soil shear strength: a sensitivity analysis using Monte Carlo simulations and feature backward elimination," *Sustainability*, vol. 12, no. 6, p. 2339, 2020.
- [14] I. Reis, D. Baron, and S. Shahaf, "Probabilistic random forest: a machine learning algorithm for noisy data sets," *The Astronomical Journal*, vol. 157, no. 1, p. 16, Dec.
- [15] A. Shaqadan, "Prediction of concrete mix strength using random forest model," *International Journal of Applied Engineering Research*, vol. 11, pp. 11024–11029, 2016.
- [16] L. Pengcheng, W. Xianguo, C. Hongyu, and Z. Tiemei, "Prediction of compressive strength of high-performance concrete by random forest algorithm," *IOP Conference Series: Earth and Environmental Science*, vol. 552, p. 012020, 2020.
- [17] M. H. Mohana, "The determination of ground granulated concrete compressive strength based machine learning models," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 8, no. 2, 2020.
- [18] T. A. Pham, H.-B. Ly, V. Q. Tran, L. V. Giap, H.-L. T. Vu, and H.-A. T. Duong, "Prediction of pile axial bearing capacity using artificial neural network and random forest," *Applied Sciences*, vol. 10, no. 5, p. 1871, 2020.
- [19] W. Ma, K. Tan, and P. Du, "Predicting soil heavy metal based on Random Forest model," in *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4331–4334, Beijing, China, July 2016.
- [20] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, 2021.
- [21] H.-B. Ly and B. Thai Pham, "Soil unconfined compressive strength prediction using random forest (RF) machine learning model," *The Open Construction & Building Technology Journal*, vol. 14, no. 1, pp. 278–285, 2020.
- [22] M. Wang, W. Wan, and Y. Zhao, "Prediction of the uniaxial compressive strength of rocks from simple index tests using a random forest predictive model," *Comptes Rendus Mécanique*, vol. 348, no. 1, pp. 3–32, 2020.
- [23] L. Pham, L. Luo, and A. Finley, "Evaluation of random forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds," *Hydrology and Earth System Sciences*, vol. 305, 2020.
- [24] C. Valencia-Payan and J. C. Corrales, "A rainfall prediction tool for sustainable agriculture using random forest," *Advances in Soft Computing*, vol. 11288, pp. 315–326, 2019.

- [25] A. J. Hill, G. R. Herman, and R. S. Schumacher, "Forecasting severe weather with random forests," *Monthly Weather Review*, vol. 148, no. 5, pp. 2135–2161, 2020.
- [26] R. Yu, Y. Yang, L. Yang, and G. Han, "RAQ-A random forest approach for predicting air quality in urban sensing systems," *Sensors*, vol. 16, no. 1, p. 86, 2016.
- [27] M. V. and D. Kumar, "Evolving Differential evolution method with random forest for prediction of Air Pollution," *Procedia Computer Science*, vol. 132, pp. 824–833, 2018.
- [28] I.-J. Han, T.-F. Yuan, J.-Y. Lee, Y.-S. Yoon, and J.-H. Kim, "Learned prediction of compressive strength of GGBFS concrete using hybrid artificial neural network models," *Materials*, vol. 12, no. 22, pp. 3708–22, 2019.
- [29] B. Boukhatem, M. Ghrici, S. Kenai, and A. Tagnit-Hamou, "Prediction of efficiency factor of ground-granulated blast-furnace slag of concrete using artificial neural network," *Materials Journal*, vol. 108, no. 1, pp. 55–63, 2011.
- [30] A. Kandiri, E. Mohammadi Golafshani, and A. Behnood, "Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm," *Construction and Building Materials*, vol. 248, p. 118676, 2020.
- [31] A. R. Boğa, M. Öztürk, and İ. B. Topçu, "Using ANN and ANFIS to predict the mechanical and chloride permeability properties of concrete containing GGBFS and CNI," *Composites Part B: Engineering*, vol. 45, no. 1, pp. 688–696, 2013.
- [32] A. Behnood, V. Behnood, M. Modiri Gharehveran, and K. E. Alyamac, "Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm," *Construction and Building Materials*, vol. 142, pp. 199–207, 2017.
- [33] C. Bilim, C. D. Atış, H. Tanyıldızı, and O. Karahan, "Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network," *Advances in Engineering Software*, vol. 40, no. 5, pp. 334–340, 2009.
- [34] S. E. Chidiac and D. K. Panesar, "Evolution of mechanical properties of concrete containing ground granulated blast furnace slag and effects on the scaling resistance test at 28 days," *Cement and Concrete Composites*, vol. 30, no. 2, pp. 63–71, 2008.
- [35] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [36] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, CRC Press, Boca Raton, FL, USA, 1984.
- [37] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 602–609, 2014.
- [38] M. Sarıdemir, İ. B. Topçu, F. Özcan, and M. H. Sevencan, "Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic," *Construction and Building Materials*, vol. 23, no. 3, pp. 1279–1286, 2009.
- [39] E. M. Golafshani and A. Behnood, "Estimating the optimal mix design of silica fume concrete using biogeography-based programming," *Cement and Concrete Composites*, vol. 96, pp. 95–105, 2019.